SECOND LANGUAGE READING IN A CALL ENVIRONMENT：THE EFFECT OF ANNOTATION USE AND THE ROLE OF WORKING MEMORY

Jiancheng Qian

ABSTRACT

The widespread application of computer technologies to language learning has led to increasingly academic reading taking place in a CALL environment. This provides opportunities for multimedia annotations to assist the reading process. However, despite the increased interests and emerging literature in multimedia annotations, research has revealed wide variation in their effectiveness for reading comprehension and insufficient investigation into the mediating role of individual difference factors in annotation use. This study focuses on the effect of annotation use on L2 reading comprehension. In particular, it explores whether L2 reading is influenced by different types of annotations, and whether working memory (WM) mediates annotation use in influencing reading performance.

One hundred and sixty-two ESL learners from a Chinese university participated in the study. They were identified to be low, medium or high in WM capacity based on the composite z-scores of three WM measures. The study required all participants to read four texts of the same difficulty, each text under a different condition: no annotations, verbal annotations, visual annotations, or verbal plus visual annotations. These conditions were counterbalanced across the readings. Participants completed a reading comprehension test for each reading, and a follow-up questionnaire on their perceptions of annotations. Fifteen participants were further interviewed to provide supplementary qualitative data concerning their reading strategies, interactions with annotations, and views of annotation design. The findings were mainly discussed with reference to the interactive model of reading, the theory of WM, the cognitive theory of multimedia learning, and the qualitative data from questionnaire and interview surveys.

Results revealed that (a) annotation use led to improved reading comprehension, (b)

the combined annotation treatment was most beneficial to reading comprehension, (c) annotation use benefited the low WM learners the most and the high WM learners the least, and (d) annotation types did not make a difference to reading comprehension for learners of similar WM capacity. The results thus suggest that annotations can enhance reading comprehension to varying degrees, and support the crucial role of WM in reading comprehension.

ACKNOWLEDGEMENTS

The completion of this dissertation owes a great deal to many people, and I would like to take this opportunity to express my sincere gratitude to all of them.

My utmost gratitude goes to Dr. Jenefer Philp, my main supervisor. To me, she is one of the most caring and inspiring mentors. It was her countless enlightening face-to-face discussions, email exchanges and Skype talks that deepened my insights into this research and guided me through the process. She devoted numerous hours to reading what I wrote for the dissertation, giving critical comments and insightful suggestions. Her great support, understanding and patience have made my academic experience at the University of Auckland enjoyable and fruitful. Without her contributions, I would not have been able to complete this research and learn so much in the process. More importantly, the high-level scholarship, wisdom and integrity that she brought to the process have taught me how to become a good researcher and will benefit me forever.

I am deeply indebted to Associate Professor Michael Barlow, my co-supervisor. His expertise in the field of CALL, valuable comments and thought-provoking questions have facilitated my research and benefited me both academically and professionally. I am very grateful to Dr. Christine Biebricher, my advisor. I was fortunate enough to have an opportunity to share her expertise in the field of reading. Her invaluable feedback and unfailing encouragement contributed much to the completion of this dissertation.

My sincere thanks also go to other faculty members in the School of Cultures, Languages and Linguistics, particularly Associate Professor Gary Barkhuizen, Professor Yan Huang, Associate Professor John Read, Dr. Shaofeng Li, and Dr. Rob Batstone. I have appreciated and benefited from them in one way or another. I am also grateful to a number

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CALL   *Computer assisted language learning*

L1   *First language*

L2   *Second language*

ESL   *English as second language*

EFL   *English as foreign language*

WM   *Working memory*

WMC   *Working memory capacity*

RST   *Reading span task*

OST   *Operation span task*

HWM   *High working memory*

MWM   *Medium working memory*

LWM   *Low working memory*

VBA   *Verbal annotation*

VSA   *Visual annotation*

VVA   *Verbal plus visual annotation*

NA   *No annotation*

1   INTRODUCTION

This study aims to investigate the effectiveness of annotation use for second language (L2) reading in a computer-assisted language learning (CALL) environment.

1.1 Research background

CALL is defined as the search for and study of applications of the computer in language teaching and learning (Levy, 1997, p. 97). It is now perceived as interplay between new concepts of language learning and new technology that in many cases generates new forms of teaching and learning (Schwienhorst, 2000). A great number of studies have demonstrated the effectiveness of CALL in the language learning process (e.g., Levy, 1997; Pennington, 1996; Warschauer, 1996; Warschauer & Healey, 1998; Zhao, 2003). In particular, with the new developments in learning technology such as hypermedia, hypertext, and multimedia, CALL has displayed a great potential for enhancing language learning. This is because learning materials can be presented in different modes or a combination of different modes, and thus learners can choose the modes and control the pace of learning. This helps motivate learners, cater for their diverse learning styles, and accommodate their varied psychological and cognitive needs (Zhao, 2003). Recently, however, there has been concern that technology-based language instruction is not necessarily superior. Some researchers argue that the flexibility and control that the technologies afford learners over their learning may come at a cost because they can increase complexity (Ellis & Kurniawan, 2000). As a result, CALL learners may not progress as efficiently as anticipated (Carnevale, 2004; Ward, 2004). These claims on the effectiveness of technologies for language learning are apparently far from being

conclusive. Beaty (2005) points out that CALL is still a largely unresearched area in need of exploration. Even where much is known, details have not been made clear. The use of multimedia is an example in point. In this area, a number of questions have not been explored in sufficient detail. These involve the target audience of multimedia assistance for reading, the response of the reader to different types of multimedia assistance, and means of assessing the use and potential benefits of this assistance (Chun & Plass, 1997, p. 71).

Multimedia reading, or CALL reading, has become a prominent area of recent CALL research. In a CALL environment, a reading text is mainly in electronic form. The most prominent feature is its nonlinear organization in presenting information in multiple forms of media (Akyel & Ercetin, 2009). Following current reading theories that emphasize the integration of both bottom-up and top-down processing, CALL reading programs have been developed to aid readers in performing interactive processes of reading comprehension. To date, the main technologies that have been used for L2 reading are electronic dictionaries and software that provides multimedia annotations (Chun, 2006). Annotations, also known as "glosses", are explanatory aids that highlight or clarify important points, or provide lexical or syntactic information for a reading text where these are perceived to be deficient in the reader's procedural and factual knowledge (Roby, 1999). Multimedia annotations refer to the explanatory aids that are provided in multiple forms of media, such as text, picture, graphic, audio, video, and animation. They are intended to facilitate the reading process as well as enhance reading comprehension (Marshall, 1997; Overstreet, 2006). There is abundant research on multimedia annotations documented in literature (e.g., Akyel & Ercetin, 2009; Chun & Payne, 2004; Chun & Plass, 1996a; Davis & Lyman-Hager, 1997; Lomicka, 1998; Yanguas, 2009). According to Chun and Payne (2004), learners' annotation look-up behaviors, different types of annotations

and their impact on learning, and individual differences among learners are important factors to consider in studying the effectiveness of CALL for reading comprehension.

The present study intends to explore the impact of multimedia annotations on L2 reading comprehension, and the mediating role of individual differences in working memory (WM) in using multimedia annotations to help L2 reading comprehension. The motivation for the exploration firstly comes from the inconclusive findings in extant literature regarding the relationship between annotation use and reading comprehension. Some researchers found that multimedia annotations positively affect reading comprehension and claimed that with annotations readers were actively engaged in the reading process, which led to deeper processing of information (e.g., Abuseileek, 2008; Akbulut, 2008; Bowles, 2004; Lomicka, 1998; Yanguas, 2009). In contrast, others found that multimedia annotations negatively affected reading comprehension, and asserted that annotations compounded the complexity of reading process and produced detrimental effects on reading (e.g., Ariew & Ercetin, 2004; Kirby, 1993; Sakar & Ercetin, 2005). Still others did not find any meaningful relationship between annotation use and reading comprehension (e.g., Abraham, 2008; Davis & Lyman-Hager, 1997; De Ridder, 2002; Roby, 1999). These studies have approached the issue from a variety of theoretical backgrounds and adopted different research designs (detailed in the next chapter). They have contributed to our understanding of the fundamental process of reading with annotations. However, the mixed findings suggest that more research, preferably with a larger sample size, is needed to address the relationship between annotation use and reading comprehension.

Secondly, the motivation comes from a research gap in extant literature that has not been investigated: whether and how WM mediates the effects of annotations on reading

comprehension, given the strong association between reading comprehension and WM found in the literature from both cognitive psychology and applied linguistics (e.g., Abu-Rabia, 2003; Chun & Payne, 2004; Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Tomitch, 1996). From a theoretical perspective, WM is strongly associated with reading and, in CALL contexts, with multimedia reading. WM is assumed to be a temporary storage system under attentional control that underpins our capacity for complex thought (Baddeley, 2007). It deals with the cognitive processes that maintain information in the mind during active processing of information, enabling the coding, processing and recording of information here and now (Numminen, 2002). Reading is a cognitive process, where readers store and process lexical, semantic, syntactic, and pragmatic information and then use this information to analyze and clarify the following parts of the text. Therefore, WM is involved in any of these processes. It is believed that WM has only a limited capacity, which differs among individuals, and this difference affects a wide range of cognitive tasks such as problem solving, reasoning, vocabulary acquisition, and reading comprehension (Cantor & Engle, 1993; Oberauer, Weidendeld, & Hornig, 2006). Previous studies have indicated that WM plays a crucial role in traditional L1 first language reading comprehension (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992; Tomitch, 1996) and L2 reading comprehension (e.g., Alptekin & Ercetin, 2009; Harrington, 1992; Harrington & Sawyer, 1992). According to Baddeley (1986), WM consists of two subsystems that store and manipulate visual images or verbal information, as well as a central executive that coordinates the subsystems (detailed in the next chapter). This multi-component model of WM can help explain the multimedia model of learning (Grimley & Banner, 2008; Mayer, 2001). Partially on the basis of Sweller's cognitive load theory (1988), Mayer (2001) proposes the cognitive theory of multimedia learning, which

provides a rationale for how learning in multiple symbol systems occurs (Ariew, 2006).

According to the theory, visual and verbal information are processed in separate systems;

when they enter WM, they will produce cognitive load. Cognitive load refers to the amount

of mental resources required to process a given amount of information (Feinberg &

Murphy, 2000; Sweller, 1994). When the information is presented in such a way that it

does not overload WM, learning in multimedia environments is facilitated. In a similar

vein, when the cognitive load exceeds the capacity and limitations of WM, learning will be

inhibited. Therefore, as an important aspect of multimedia learning, WM may be crucial to

understanding learner behavior in using multimedia annotations, and has significant

implications for the study of the effects of annotations on reading comprehension. Plass,

Chun, Mayer and Leutner (2003) suggest that "one probable cause for the detrimental

effects that were found in some studies is the cognitive load which is imposed on the

learner when using multimedia information for learning and the limited processing

capacity of the human working memory (p. 222).

Recently, some studies related to WM have scraped the surface of multimedia

learning. For example, a study by Plass, Chun, Mayer and Leutner (2003) explored the

cognitive load in reading an L2 text with multimedia aids and the influence of verbal and

spatial abilities. The results indicated that multimedia learning processes were executed

under the constraints of limited WM, but the study did not include the WM capacity

(WMC) of the participants as a variable. Chun and Payne (2004) conducted a study

involving WM, annotation use, and reading comprehension, which revealed that the low

WM learners used multimedia annotations to compensate for their WM constraints while

reading and learning new vocabulary. The focus of the study, however, was on the

relationship between WM and annotation look-up frequency and how the frequency was

related to reading outcomes. In addition, there was only one annotation type developed for an annotated word (either in the form of text, or image, or video-clip). Although these studies suggested that WM is closely related to reading comprehension in CALL environments, in fact to my knowledge no empirical study to date has attempted to explore whether WM affects the efficacy of multimedia annotation on reading comprehension, or the relationship between WM and the effects of different types of annotations on reading comprehension. The paucity in literature points to a need to empirically examine the effectiveness of annotations mediated by WM for reading comprehension.

1.2 Research purpose and significance

Motivated by previous studies on multimedia annotations and the existing research gap in the literature, the present study explores the effectiveness of annotations for reading comprehension, and whether individual differences in WM may mediate the potential benefits of annotations in a CALL reading context. It is conducted within the frameworks of the interactive model of reading (Rumelhart, 1977) and the cognitive theory of multimedia learning (Mayer, 2001). The learning environment for this study is an L2 multimedia reading program that requires university students of English to read advanced L2 texts. The reading task is the comprehension of the texts, which is assessed through a reading comprehension test. The annotations are provided to the texts at both word and topic level, based on the assumption that CALL reading is an interactive process, where bottom-up and top-down operations interact with the help of the annotations to assist global and local processing of information. Three annotation types are developed: verbal (text), visual (picture), and verbal plus visual, following the multimedia model of learning. Clicking of the annotations is required due to the exploratory nature of the study. This will

help unveil the relationship between a particular annotation type and reading comprehension. Learners' WM is measured by a battery of WM tasks, so that the effects of annotations on reading comprehension according to WMC can be obtained.

This study has significance for theory and practice. On the theoretical level, it aims at contributing to our understanding of reading, multimedia learning and WM. It applies theories of these three areas to a CALL reading context, and in particular, to creation and use of multimedia annotations in this context. In doing this it adds to the growing body of literature investigating the effects of multimedia annotations on reading comprehension. The findings may in turn provide important support for these theories. On a practical level, the study is expected to be informative and constructive for the design of CALL reading programs, and in particular for the development of multimedia annotations for a reading text. In addition, of special interest in the study is individual difference in WM. Individual differences are characteristics or traits in respect of which individuals may be shown to differ from each other (Dornyei, 2005). The study of the relationship between learners' individual differences and the effective design of multimedia environments to facilitate L2 text comprehension and vocabulary learning has become a central concern for CALL reading research (Abraham, 2008). However, individual difference factors have not been adequately researched in current literature on multimedia annotations. Therefore, although this study only considers one factor of individual differences, it is hoped that it will inspire more interest in and draw more attention to other factors in the research on CALL reading.

The context of this study was CALL at a university in China, where modern information technology is being introduced to the teaching of English as a foreign language (EFL) to the students. The previous research on CALL by the present researcher (e.g., Qian, 2007, 2011), combined with his growing teaching experience in CALL in the

university, serves as the primary motivation for his interest in this area. To the best of the researcher's knowledge, there has been no study in China on the relationship between annotation use and reading comprehension. Therefore, this study is likely to contribute to extending the research on annotations to a new educational context.

1.3 Organization of this dissertation

The remainder of this dissertation unfolds in five chapters. Chapter 2 presents a review of major theoretical issues and empirical studies relevant to this study, through which to provide a rationale for this study. Chapter 3 introduces the methodology of this study, with a detailed description of the materials used and instruments for data collection. The results of a pilot study on the validity and reliability of the instruments are also described in the last section of this chapter. Chapter 4 first reports the results of the WM tests, based on which the participants are categorized, and then presents the statistical results based on each of the research questions with some basic interpretations. This is followed by analyses of the data generated from questionnaire and interview surveys. Chapter 5 is devoted to an in-depth discussion of the findings for each of the research questions, providing explanations in reference to relevant theories. The last chapter summarizes the major findings of this study, considers its possible contributions and pedagogical implications, then discusses its limitations and makes recommendations for future research.

2 LITERATURE REVIEW

This chapter begins with a review of current theories of reading, explaining that the study of annotations in L2 texts should be grounded in what we know about the L2 reading process. The second section discusses how computer technologies have changed the reading process and elaborates a theory related to a multimedia model of learning. The third section is dedicated to the summation of studies associated with annotations, vocabulary learning and reading comprehension. The last section first reviews theories and measures of working memory (WM), and then examines relevant practice in relation to WM, annotations and reading comprehension. Thus, this chapter reveals the interdisciplinary nature of this study --- L2 reading, CALL, and educational psychology.

2.1 Current theories of reading

2.1.1 Theoretical models of reading

Researchers from a variety of disciplines have made abundant studies of reading, among which the reading process has been one of the key topics. Three major models that explain the reading process have been identified: bottom-up, top-down, and interactive.

Bottom-up model of reading

The bottom-up model considers reading as a strictly serial process from the lowest unit, the letter, to the highest unit, the passage. According to Gough (1972), who is believed to have developed this reading model, a reader begins with letters, converting them into phonemes through decoding, then s/he recognizes the phoneme as a word; s/he then goes on with the next word; this continues until s/he recognizes all the words in a sentence, at which point s/he applies the syntactic and semantic rules to give a meaning to

the sentence. Evidently, this model views reading as a hierarchically organized process, from most specific at the bottom to most general at the top, where the reader uses the information received from the text to process comprehension. Since it places emphasis on textual decoding, it is sometimes called "text-driven model" or "data-driven model" (Carrell, 1983; Chun & Plass, 1997).

The bottom-up model of reading suggests that the reader has to process each of its individual words and sentences in a text before arriving at full understanding of the text, thus attaching importance to vocabulary knowledge in reading comprehension. However, this view was criticized on the grounds that it ignores the role of the reader and what s/he may bring to the process of meaning construction (Goodman, 1987). For example, there are occasions where the reader can skip over a new word and arrive at a correct comprehension of the sentence by using contextual clues, or occasions where comprehension fails even if literal meaning is fully recognized. The likelihood that readers can utilize their language knowledge, prior knowledge, and reading skills to interact intelligently with the text to achieve comprehension leads to the negation of the idea that reading proceeds in a linear manner, as suggested by the model (Eskey, 1988).

Top-down model of reading

In contrast to the view of the bottom-up model, Goodman (1967) regards reading as a psycholinguistic guessing game. He explains that the reader is an active participant in the reading process, making guesses about the message of the text and checking the text to confirm or refute them according to his or her prior knowledge. Grabe (1991) concurred with Goodman's arguments that reading was an "imprecise, hypothesis-driven process" (p. 377), where s/he reads by predicting information, sampling the text, and confirming the prediction. In this sense, reading follows a top-down model, or a hypothesis-testing process

(Zakaluk, 1998). In explaining this model, Smith (2004) points out that prediction is the core of reading, bringing potential meaning to texts, reducing ambiguity and eliminating in advance irrelevant alternatives. He cautioned that prediction is not reckless guessing, nor does it involve taking chances by betting everything on the most likely outcome (2004, p. 25). This model suggests that reading for meaning is the primary objective of reading rather than mastery of words, and that meaning is brought to the text, not derived from the text.

Clearly, the top-down model of reading emphasizes the role of prior knowledge that a reader brings to the reading process. This prior knowledge is known as "schema" (Nicholson, 1984; Rumelhart, David, & Ortony, 1977). According to Rumelhart (1980), a schema can be thought of as a knowledge structure, or framework, which interrelates all of one's knowledge about a given topic, and within a schema framework one's preexisting knowledge and life experiences join with new knowledge. Anderson (1985) refers to schema as our previously acquired knowledge of the world, called background knowledge, the structures of which are called schemata. The theory of schema (Rumelhart, 1977) posits that learning something new involves building bridges between the new information and the existing world knowledge. In regard to reading, researchers have identified three types of schemata: content schema, which refers to a reader's background or world knowledge; textual schema, which refers to the organizational forms and structures of written texts, and linguistic schema, which includes the decoding features needed to recognize words and how they fit together in a sentence (Singhl, 1998). Readers may use their schemata to predict the meaning of the text, and reading comprehension occurs when these schemata match the information in the text. Previous studies addressing L2 text comprehension found that comprehension was enhanced when learners possessed and used the appropriate

background knowledge, whether it was associated with subject matter or culture (e.g., Alderson & Urquhart, 1988; Carrell, 1988; Chen & Donin, 1997).

However, the top-down model is not without limitations. Some scholars have criticized it as essentially a model of fluent readers and one that does not account for all the needs of learners acquiring reading skills. For example, Eskey (1988) argues that the model tends to emphasize higher level skills like the prediction of meaning by means of context clues or background knowledge, and to deemphasize the perceptual and decoding dimensions of identifying lexical and grammatical forms. He insists that language decoding has a major role to play in the reading process, and good reading is a more language-structured affair than the guessing-game metaphor seems to imply (1988, p. 94). Clark (1988) also contends that this model is good for the skillful, fluent reader for whom perception and decoding have become automatic, not for the less proficient reader.

Interactive model of reading

As discussed above, both the top-down and bottom-up models of reading have some limitations, as neither can totally account for what occurs in the reading process. However, they both relate to each other one way or another (Bensoussan, 1998; Ji, 2003). The interactive model, developed by Rumelhart (1977), combines the valid insights of bottom-up and top-down models. According to this model, reading is neither a bottom-up nor a top-down process, but instead an integration of the two patterns, where the reader, the text, and many component skills interact simultaneously. Grabe (1991) elaborates that an interactive model refers to two different conceptions: first, it refers to the interaction between the reader and the text, where the reader constructs meaning based partly on the information extracted from the text and partly on the existing background knowledge that the reader has; second, it refers to the interactivity between the many component skills that

result in reading comprehension. Therefore, from an interactive approach, the reading process involves both low-level automatic identification skills and higher-level comprehension skills (Seng, 2006). As such, the reader is seen as a cognitively active learner, who both decodes and interprets the text.

The interactive reading model implies that top-down and bottom-up processes work in tandem to achieve a comprehension goal. In reading, we do not proceed from letter to words to meaning in step-by-step fashion, nor do we proceed from the whole to the part. Rather, we engage in parallel processing so that we use content schema, contextual schema and linguistic schema in combination, and comprehension results from these interactive variables operating simultaneously rather than sequentially. The simultaneous interaction is advantageous in that it can compensate for deficiency at any level. For example, a reader with limited vocabulary may rely on the topic knowledge that s/he has in reading a text successfully. As McCormick (1988) puts it, the interactive model attempts to take into account the strong points of the top-down and bottom-up models, and tries to avoid the criticisms leveled against each other, making it one of the most promising approaches to the theory of reading today.

The implication of the model for L2 reading is that lower-level processing skills are fundamental to reading, and higher-level processing abilities play a significant role in reading. To achieve reading comprehension, readers should conjointly develop their bottom-up recognition skills and their top-down interpretation strategies. Bernhardt (1991) suggests that L2 reading comprehension results from the interaction of such factors as word recognition, syntactic parsing, intra-textual perception, prior knowledge, and meta-cognition. Each of them may make a difference to the reading performance. This demonstrates the complexity of L2 reading comprehension. The interactive model of

reading is held in the present study. In CALL reading, readers have the resources of annotations that serve to assist both bottom-up and top-down reading processes. Specifically, lexical annotations can assist where there are gaps in receptive vocabulary knowledge. Conversely, topic annotations can address gaps due to a lack of background knowledge. Therefore, accessing annotations at word and topic levels will supposedly lead to better reading performance. Following the interactive model of reading, the present study develops word and topic annotations for the reading texts, and examines their potential benefits for reading comprehension in a CALL context. Although it concerns readers who are highly proficient and fluent in L1, in L2 learning contexts they may lack the proficiency of native speakers, or experience in certain domains, or have some gaps in cultural knowledge.

2.1.2 Hypotheses concerning the relationship between L1 and L2 reading

L2 refers to any language acquired after the native, or first language (L1) is fully established (Stern, 1983). Much of our knowledge of L2 reading depends on theories drawn from the research on L1 reading (Butler & Hakuta, 2006). However, L2 reading is a very complex activity, and there are a number of variables making it different from L1 reading, such as language proficiency, background knowledge, and textual schema. First, L1 readers already have a strong vocabulary base and implicit grammatical knowledge of the language, enabling them to process linguistic information of a text. L2 readers, on the other hand, do not have this advantage. Second, even with L2 linguistic skills, they may not possess the knowledge necessary to perceive L2 texts in a culturally authentic, culturally specific way (Singhl, 1998). Prior cultural knowledge or experience can prepare readers to comprehend the familiar information in the text. L2 readers, however, do not

possess the same degree of content schema as L1 readers, and this can result in reading difficulties. For example, Chinese readers who are unfamiliar with the specific cultural norms and rituals pertaining to a western wedding may be less successful in interpreting a text about the wedding, for they tend to construct meaning in light of their relevant native norms and rituals. Third, sometimes, L2 readers may know all the vocabulary and understand the main concepts of an L2 text, but they may not able to follow the specific development of the text, or they may not be completely familiar with genre expectation of certain types of L2 texts (Yamashita, 2004). Differences in text structures do exist among language groups (Deanel, Sheehan, Sabatini, Futagi, & Kostin, 2006) and some researchers suggest that a text violating readers' expectation about its development can have disruptive effects on reading comprehension (Carrell, 1984; Nambiar, 2009; Singhl, 1998). According to Grabe and Stoller (2002), L2 readers must learn to use L2 linguistic knowledge, deal with L1-L2 transfer effects, and learn to use L2-specific resources, for example, bilingual dictionaries, glosses, and translation. This suggests that L2 reading can be more demanding than L1 reading.

Considering why learners who are proficient in L1 find it difficult to read in L2, Alderson (1984) posed the question as to whether L2 reading difficulty was a reading problem or a language problem. Two hypotheses address this issue: linguistic threshold hypothesis and linguistic interdependence hypothesis. The linguistic threshold hypothesis was first developed by Clarke (1979), originally known as "short-circuit hypothesis". According to the hypothesis, L2 learners must first gain a certain amount of control over L2, or in other words, cross a critical linguistic threshold, before applying their L1 reading skills to L2 reading. A lack of L2 linguistic knowledge will ultimately "short-circuit" the L1 reading knowledge; that is, linguistic proficiency interferes with reading performance.

This "certain amount" is referred to as a "language ceiling" (Clarke, 1979), or a "threshold level of linguistic competence" (Cummins, 1979). No matter how proficient one is in L1 reading, below this level of linguistic competence, it is unlikely for L1 reading skills to be transferred to L2 reading. As Clarke (1988) indicates, limited L2 proficiency prevents the transfer of the top-down approach in L1 reading to L2 reading and confines the reader to the bottom-up approach using mostly word-attack strategies. Contained within this hypothesis is the belief that language is the key factor in reading activities. Alderson (2000) suggests that "in L2 reading, knowledge of the L2 is a more important factor than L1 reading abilities" (p. 23). Grabe and Stoller (2002) point out that if L2 readers devote most of their cognitive resources to figuring out the language of the L2 text, there are few cognitive resources left over for the fluent comprehension processes that would normally support the L1 reader. An empirical study by Taillefer (1996) investigating the short-circuit hypothesis found that the more difficult the reading task, the more important L2 knowledge became.

In contrast, the linguistic interdependence hypothesis (Cummins, 1979) posits that L1 linguistic knowledge and skills play an instrumental role in the development of an L2. According to this hypothesis, language operations are transferable. When a language operation such as reading has been acquired in one language, the same operation is not reacquired in another, but is simply available upon need. Therefore, L1 and L2 reading are only superficially distinct and are actually interdependent. Bernhardt and Kamil (1995) contend that reading performance in an L2 is largely shared with reading ability in an L1. Based on the claims of the hypothesis, if the reading process is taken as an interactive process, one's knowledge in L1 reading would be an asset in L2 reading, rather than irrelevant (Kong, 2006). Evidence for this can be obtained from recent studies with

Chinese university students, which found that Chinese reading ability is a significant predictor of English reading ability (Chen & Fan, 2007; Wang, 2006).

These two hypotheses have often been considered conflicting, and researchers have attempted to determine which of the two gives the better explanation of the relationship between L1 and L2 reading. Alderson (1984) reviewed several empirical studies with a view to gathering evidence in support of either of the two hypotheses. He found stronger evidence to show that L2 reading was a language problem for L2 readers with low foreign language proficiency levels. Yamashita (2002) also maintains that in the case of weak L2 linguistic proficiency, the strategies used in L1 are not always fully successful in helping readers construct an appropriate meaning. According to Eskey (2002), L2 proficiency is a variable that best correlates with effectiveness in L2 reading. It seems that linguistic threshold hypothesis should still be given priority. This is because L2 proficiency is a prerequisite for reading activity with that language as the vehicle of information presentation, and moreover has a strong influence on the transfer of L1 reading ability to L2 reading. However, there is no denying that L2 reading is both a language and a reading problem, and the two hypotheses should be viewed as being mutually complementary rather than exclusive. They each contribute to our understanding of the L2 reading process by touching on its different aspects, and neither fully explains success in L2 reading.

The present study is based on an L2 context and investigates the effectiveness of annotations for L2 reading comprehension in a CALL environment. The two hypotheses have important implications in relation to annotations. The use of lexical annotations may assist the reader in comprehending new words in a text to cross the linguistic threshold for L2 reading, and the reference to topic annotations like background knowledge adds to the reader's strategies and skills for effective reading. Based on these hypotheses, L2

annotation design should take account of what we know about the L2 reading process.

2.2 CALL reading

2.2.1 Computer technology and language learning

There are many terms used to describe the roles of the computer in language learning: CALL (computer-assisted language learning), CALI (computer-assisted language instruction), CELL (computer-enhanced language learning), TELL (technology-enhanced language learning) and so forth (Levy, 1997). These broad terms reflect the different perspectives on the concept of using computers in language learning. However, for its universal acceptance, "CALL" is often used as an umbrella term to include multimedia-based, computer-based and web-based language learning.

CALL has undergone several transformations, which have corresponded to the change of the dominant linguistic theories and the availability of computer technologies of the time (Fotos & Browne, 2004). According to Warschauer & Healey (1998), the history of CALL can be divided into three stages: behavioristic CALL, communicative CALL, and integrative CALL. The first stage of CALL appeared in the 1960s and 1970s. Computers functioned mainly as a mechanical tutor, providing learners with opportunity for repeated exposure to learning materials. CALL software was developed to allow for language drills and practices. The second phase of CALL was based on the communicative approach to teaching, which gained prominence in the late 1970s and 1980s. The communicative approach emphasizes meaning-focused language use rather than formal instruction, and it takes communicative competence as the goal of language teaching (Richards & Rodgers, 2001). Following this cognitive model of language learning, CALL software was designed aiming at stimulating learners' motivation, critical thinking, and analytical skills, and

facilitating understanding and manipulating of the target language (Fotos & Browne, 2004; Warschauer, 1996) for communicative use of the language. Integrative CALL emerged in the mid 1990s as a prospective way to integrate various technologies more fully into the language learning process (Warschauer & Healey, 1998). It is based on two important technological developments: multimedia computers and the Internet. Multimedia, which combines a wide range of communication elements such as text, sound, graphic, picture, photograph and animation, creates a platform for interactive programs. A typical multimedia language program may allow learners, for example, to view a video clip, take a comprehension test on its content, and receive immediate feedback. As a further development of multimedia, hypermedia, which links multiple forms of media, shows greater advantages for language learning, including a more genuine and natural learning environment, more control by learners over their own learning while learning at their pace, and facilitating focus on content (Warschauer, 1996). According to Fotos and Browne (2004), much of the theory underlying integrative CALL is derived from the socio-cultural model of language learning, in which interaction is regarded as essential for the creation of meaning.

The development of CALL indicates that language teaching and learning are influenced by technological innovations. This is because that technology can inspire new philosophies and approaches for language education, opening up new fields of language research and promoting language teaching and learning. Warschauer (1996) offers a summary of the roles of the computer in language teaching and learning: it can be a tutor which offers language drills or skill practice, a stimulus for discussion and interaction, a tool for writing and research, a medium of global communication, and a source of limitless authentic materials. Literature shows that CALL has been viewed as a valuable teaching

approach, and many teachers embrace the idea that CALL should be an integral component of effective teaching (Bell, 2005; Braul, 2006; Debski & Gruba, 1999). A large number of studies suggest that learners hold positive attitudes towards computers in their classes and they acknowledge that CALL has enhanced their ability to learn, and facilitated their access to up-to-date information (e.g., Rickma & Grudzinski, 2000; Schrodt & Turman, 2005; Taylor & Gitsaki, 2004).

However, not all teachers are convinced of the benefits of teaching with technology and some fail to see how technology can improve instruction. In particular, traditional teachers tend to think that technology does not fit with their beliefs about effective teaching, and that the use of computers waste class time which could have been more profitably devoted to teaching content (Arnold & Ducate, 2006). They also worry that using computers in class can remove the social aspect of teaching (Spodark, 2005), lower students' research skills, and encourages laziness (Russell, Bebell, O'Dwyer, & O'Connor, 2003). Ellis and Kurniawan (2000) argue that the freedom and control that the technologies afford learners can increase the complexity of the learning process. On the students' side, some complain that computers do not facilitate their learning or discussion, and that they waste time troubleshooting problems (Young, 2004).

Computer technology has become an important aspect of language teaching and learning. These conflicting views on the potentials of computer technology for language teaching and learning indicate that the field of CALL is still full of issues and opportunities to explore. There is a general consensus that technology alone can never ensure the success of language learning, and the use of technology does not automatically result in improved outcomes. Researchers (e.g., Arnold & Ducate, 2006; Zhao, 2003) have often reflected on to what extent CALL can contribute to learning achievement, and how CALL programs

can be better designed to facilitate learning and cater for the needs of diverse learners. These questions have guided much of the research in recent years.

2.2.2 Call reading vs. paper reading

In a CALL environment, learning takes place through computers. Computer reading, synonymously referred to as CALL reading in the present study, is different from traditional paper reading. In paper reading, the sentences, paragraphs of a text or a novel are linear, in the sense that they are read from beginning to end (Ashworth, 1996). When we come across new words or grammatical difficulties, we can turn to a dictionary or a reference book. With the fingers pointing at words, or turning over pages, we have the feeling of literally being in touch with the text. In CALL reading, however, the text is mainly in electronic form, known as hypertext. The most prominent feature of hypertext is its nonlinear organization in presenting information (Akyel & Ercetin, 2009), which can benefit reading in several ways. First, readers can link one piece of a text to another and display the latter as a separate page or as a pop-up window on the screen. Because of this flexibility, the reader can approach the text more globally, rather than linearly. Second, learners can select information based on their own needs and interests, and interact with the text at their own pace. It is argued that the interactivity of this kind between the reader and the text better motivates the learner in the reading process (Roblyer, Edwards, & Havriluk, 1997), and leads to deeper processing of information (Patterson, 2000).

Apart from hypertext, information can also be presented through multimedia and hypermedia. The former usually involves a combination of slides, movies and sound, while the latter incorporates multimedia features into linked texts and presents information in multiple forms of media such as text, pictures, graphics, audio, video, and animation

(Akyel & Ercetin, 2009; Jobst & Dillon, 2005). However, since hypermedia can be defined as hypertext with multiple forms of media, these terms are often used interchangeably (Jonassen, 1996). In this study, hypertext, multimedia and hypermedia will be used as synonyms just for the sake of simplicity, and as a result, hypertext reading, multimedia reading and hypermedia reading are subsumed under CALL reading. The linkages that these media provide can be beneficial to L2 reading. For instance, words, phrases, and sentences of a text can be annotated in pop-up windows, the text can be pronounced in digitized voice, and meaning can be illustrated with graphics, animations, and video clips (Ashworth, 1996). Therefore, multimedia reading is distinguished by the richness and depth of the information it provides through chunks of information that are linked together (Preece, 1993). Sakar and Ercetin (2005) point out that when readers have a widespread access to the multiple media used to aid reading in a CALL environment, the traditional models of L2 reading that emphasize interaction between bottom-up and top-down processes for successful reading comprehension fall short in explaining text processing in a hypermedia environment, because readers are engaged in processing not only the text but also the verbal and visual input provided through annotations (p. 28). According to De Ridder (2000), the presentation of the text on screen and the additional features that trigger the learner's attention while reading and/or clicking must determine the reading process. In a word, CALL reading requires additional reading skills to cope with the demands of the new environment.

2.2.3 Cognitive Theory of Multimedia Learning

The development of computer technology in language learning, as discussion above, has suggested that effective use of a new educational technology should be guided by a

research-based theory. Advances in cognitive psychology provide a basis for such theories. One of the most important areas of cognitive psychology is the understanding of how multimedia technology can facilitate learning. It is believed that the cognitive theory of multimedia learning (Mayer, 2001) provides rationales about how learning in multiple symbol systems takes place (Ariew, 2006).

The theory, which draws on Paivio's (1990) dual coding theory, Sweller's (1988) cognitive load theory, Baddeley's (1986) model of working memory, and Wittrock's (1989) generative theory, has three assumptions. The first is channel assumption. According to the dual coding theory, there are two cognitive subsystems, one specialized for the representation and processing of nonverbal objects, and the other specialized for dealing with language input. The two subsystems function independently, simultaneously and inter-connectedly (Paivio, 1990). In other words, visual and verbal information are processed separately and along different channels. When simultaneously activated, one type of information may be transferred to the other through interaction between the channels. The second assumption is limited capacity, which suggests that each channel has a limited cognitive capacity (Baddeley, 1986). That is, learners have limited cognitive resources to hold and process information in each channel at any one time. For example, if learners are faced with a series of pictures on the computer screen, they can bring in and hold only a few of them in the WM. The third assumption is the active processing assumption, which suggests that humans are actively involved in the construction of knowledge using their cognitive processes (Wittrock, 1989). Three important cognitive processes are involved: selecting, organizing, and integrating. Selecting is applied to incoming verbal information to produce a text base and is applied to incoming visual information to yield an image base; organizing is applied to the text base to create a

verbally-based model of the to-be-explained system and is applied to the image base to create a visually-based model of the to-be-explained system; integrating occurs when the learner builds connections between corresponding representations in the verbally-based model and the visually-based model, with the help of the his or her prior knowledge, to form a coherent mental structure (Moreno & Mayer, 1999). Therefore, the two models are believed to complement each other: pictures provide holistic and nonlinear information, while words provide discrete information in a linear way (Ariew, 2006; Mayer, 2001). Mayer (2005) asserts that learning in multimedia environments is facilitated when the information is presented in such ways as accompanying words and with pictures instead of only in words, placing words and pictures near rather than far from each other, presenting them simultaneously instead of successively, and in such a way that it does not overload the WM.

The cognitive theory of multimedia learning describes learning process in a multimedia context from multiple presentation modalities, and is graphically illustrated in Figure 2.1. It can be seen that three memory systems are represented: sensory memory, working memory and long-term memory. When learning begins, materials presented in verbal and visual forms enter the sensory memory through the eyes and ears. This process is referred to as "selecting". Within the context of the dual coding view, there can be two selecting processes: selecting relevant words for a verbal representation, and selecting relevant images for a visual representation (Mayer, 1997, p. 5). In the next step, only selected sounds and images enter the WM (due to its limited capacity), where they are temporarily stored and organized in a more coherent way. This process is called "organizing". In models of human information processing, organizing involves a transformation of verbal knowledge within the verbal WM and a transformation of visual

knowledge within the visual WM (Mayer, 1997, p. 5). Finally, a referential connection is

built between the verbal and visual representations. This process is called "integrating",

where prior knowledge already in long-term memory is activated to relate to the newly

acquired information. Mayer (1997, 2001) points out that for the integration to take place

the verbal information must be held in verbal WM at the same time that the corresponding

visual information is held in visual WM. However, because of the limited holding capacity

of WM, integration is constrained by memory load (Baddeley, 1992; Chandler & Sweller,

1991).

| Multimedia Presentation | Sensory Memory | Working Memory | Long-term Memory |
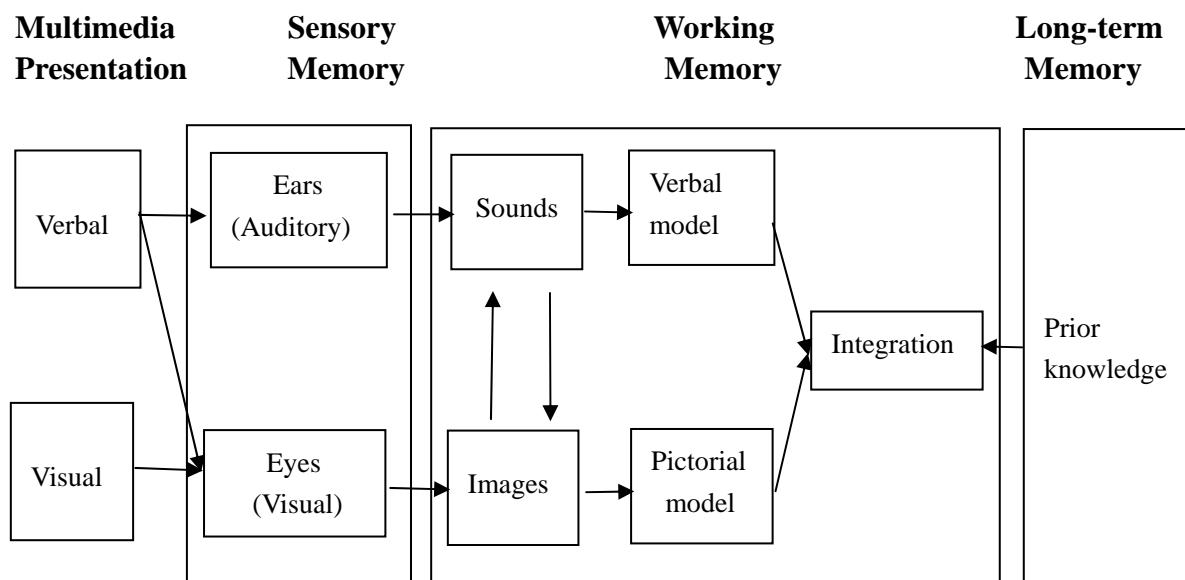
Figure 2.1 *Cognitive Theory of Multimedia Learning (adapted from Mayer, 2001)*

According to Xu (2010), the theory is "probably the most influential theory for L2

learning via multimedia in the past 10 years" (p. 312). Because the present study focuses

on the effects of annotations on L2 reading comprehension, this theory is chosen as one of

its theoretical frameworks. On the basis of this model and its application in empirical

studies, Mayer and Moreno have developed and tested some major principles of how to use

multimedia to enhance learning performance. These principles include:

(a) Multiple Representation Principle: It is better to present an explanation in words and pictures than solely in words;

(b) Contiguity Principle: It is better to present corresponding words and pictures contiguously rather than separately when giving a multimedia explanation;

(c) Coherence principle: Multimedia explanations are better understood when they include few rather than many extraneous words and sounds;

(d) Modality principle: It is better to present words as auditory narration than as visual on-screen text;

(e) Redundancy principle: It is better to present animation and narration than to present animation, narration, and on-screen text (where the last two are redundant).

(Adapted from Kirschner, 2002; Moreno & Mayer, 1999)

2.3 Annotation and reading

2.3.1 Definition and classification of annotation

Of research interest in the realm of CALL reading is how learners use different types of multimedia and whether these media aid both the lower-level process of understanding individual words and the higher-level process of text comprehension. Researchers have focused their investigation particularly on the efficacy of presenting annotations or glosses to reading texts in the different modalities such as text, picture, audio and video. "Annotation" and "gloss" are two terms frequently used in CALL literature. Nielsen (1995) refers to annotation as a special link type to a small additional amount of information which allows the reader to depart from the primary material temporarily and then return. Roby (1999) defines gloss as a kind of attempt to supply what is perceived to be deficient

in a reader's procedural or declarative knowledge about a certain word or expression. The two terms have been used interchangeably in L2 research and pedagogy. For example, Abuseileek (2008) states that "an annotation, or a gloss, is defined as adding comments or notes about difficult words, phrases, or ideas in order to provide their definitions or meanings in a particular context." (p. 260). In the present study, they will be taken as synonyms. "Annotation" will be used in most cases, and "gloss" is used instead in the case of quotations, to be consistent with the authors.

Traditional annotations were in the form of marginal notes either on the side or bottom of the page, written by students as study aids or by teachers as teaching aids. The advance of multimedia application in second language teaching and learning has led to the appearance of annotations. Multimedia annotations are sometimes used interchangeably with electronic glosses, hypertext annotations, or hypermedia annotations in the literature (Yao, 2006). Different from traditional annotations which are restricted to textual information, multimedia annotations use multiple modalities to present notes, such as text, sound, picture and video (Al-Seghayer, 2001).

Drawing upon theoretical discussions in the literature, Roby (1999) provided a comprehensive taxonomy of annotations, which is graphically displayed in Figure 2.2. According to the taxonomy, annotations can be textual and contextual in terms of focus, verbal and visual in terms of form. Textual annotations refer to those at word or sentence level, and contextual annotations to those at topic level (Marshall, 1997; Overstreet, 2006). A word annotation is defined as an explanation of the meaning of an unfamiliar word with a brief definition or synonym (Nation, 1990), and a topic annotation is one that provides background knowledge about the text topic. In terms of form, these annotations thus
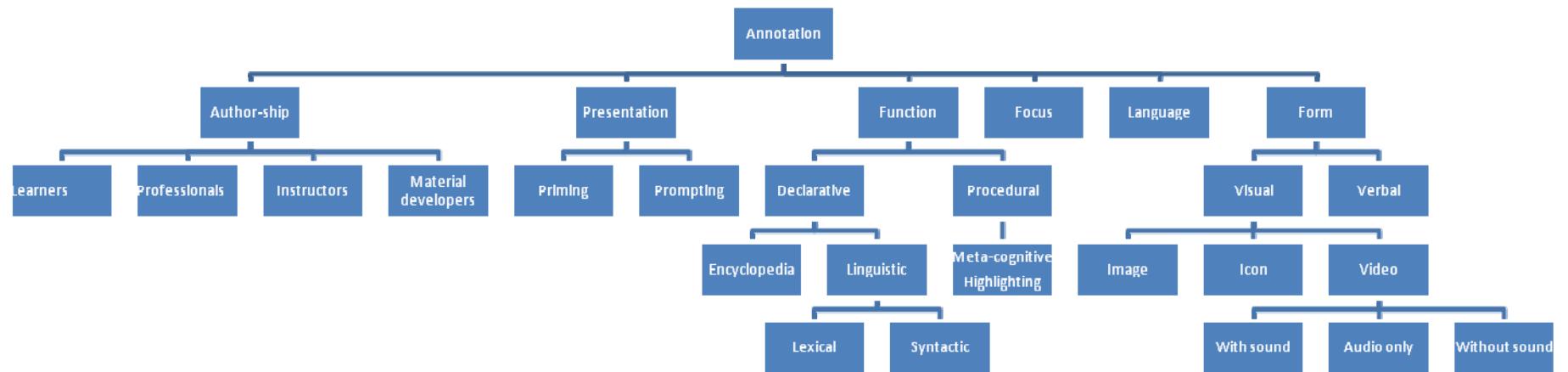
Figure 2.2 *Taxonomy of annotations (based on Roby, 1999, p.3)*

defined are known as verbal annotations since they are in written language, despite the fact that the written language is also visual to readers. However, if they employ pictures or video clips, without visible or audible language, they are called visual annotations. The present study involves textual and contextual annotations, both in verbal and visual forms. Three types of annotations are involved: verbal, visual, and verbal plus visual.

Multimedia annotations have some distinctive features in contrast with traditional annotations: they can be embedded invisibly in the text and therefore do not interrupt the reading process (Davis, 1989; Martinez-Lage, 1997); they offer learners the freedom to obtain as much or as little information as desired (Davis, 1989; Nation, 1990); they can be consulted immediately (Martinez-Lage, 1997). Among the concerns often raised about the potentials of annotations is the impact of glossing individual vocabulary via annotations embodied by different modes and media on vocabulary learning and reading comprehension. These concerns are reviewed in the following two sections respectively.

2.3.2 Annotation and vocabulary acquisition

Vocabulary knowledge is viewed as a key component of reading success (Grabe, 2004; Qian, 2002). As is suggested by the interactive model of the reading and the linguistic threshold hypothesis, word recognition plays a principal role in L2 reading. The rationale of providing word annotations for a reading text is that readers are supported in performing low-level (bottom-up) processing of information by being able to recognize and understand individual lexical items, which in turn frees up WM resources for higher level (top-down) processing (Chun, 2006; Chun & Plass, 1996b; Lomicka, 1998; Overstreet, 2006). There are a myriad of empirical studies investigating the effects of annotations on L2 vocabulary acquisition. Four studies in particular stand out as representative of multimedia annotations.

Al-Seghayer (2001) compared the effectiveness of different types of annotations for vocabulary acquisition. The participants, 30 intermediate-level ESL learners, were asked to read a narrative English text with three types of annotations developed for unknown words: printed text definition only, printed text definition coupled with still pictures, and printed text definition coupled with video clips. Two vocabulary tests were administered to assess the effect of each annotation type on vocabulary learning: a recognition test requiring the participants to answer multiple-choice items, and a production test requiring them to briefly define the annotated words in English. The time for the task (including reading and vocabulary tests) was controlled. The results of the two tests showed that all the words that were dually annotated with picture plus text definition and those with video plus text definition were better learned than those with text definitions only, and that words with text and video annotations were better remembered than those with text plus picture annotations. In addition, responses from questionnaires indicated that video was rated as a very helpful mode (86.6%), followed by pictures (70%), and by contrast, only 10% of the participants viewed textual definitions as being very helpful. Al-Seghayer (2001) suggests that one of the factors accounting for the text-plus-video condition being more effective than text-plus-picture condition was that video helps build a mental image, inspire curiosity, and promote a combination of image, sound, and printed text.

The superiority of dual annotation conditions over single annotation condition was also reported in a study by Chun and Plass (1996b). They conducted an experiment with 103 second-year university students, who were native speakers of English studying German as a foreign language, as subjects to examine how incidental vocabulary learning could be influenced by annotation use. They used a multimedia program which provides text definitions for 12 words, pictures and text definitions for another 12 words, and video

and text definitions for another 12 words. It was found that the participants received higher scores and correct answer percentages on the immediate and delayed vocabulary posttests than on the pretest, and that there was a pattern of increased mean scores and correct answer percentages on vocabulary tests for text-only, text-video, and text-picture conditions. The superior performance associated with text-picture annotations was later confirmed in another study by Plass, Chun, Mayer, and Leutner (1998), which revealed that the participants ($n = 103$) remembered words better when they selected both visual and verbal annotations than only one type of annotation. A possible explanation of the results can be drawn from the cognitive theory of multimedia learning (Mayer, 2001). However, Chun and Plass (1996b) found that words with still pictures and text definitions were recalled significantly better than words with video and text definitions, a different finding from Al-Seghayer's (2001) study. They postulated that pictures can be viewed for as long as the learner wishes, allowing for the development of a mental model of the information, as opposed to video presentations which are usually too short for the information to be established in long-term memory. In a similar study on vocabulary annotations, Kost, Foss, and Lenzini (1999) randomly assigned their participants (56 university L2 students) to one of three annotation condition groups: textual annotation alone, pictorial annotation alone, and text plus picture annotation. Three tests (production, picture recognition, and word recognition) were used to measure the students' vocabulary learning and retention two weeks later. Overall, it was found that the combination condition produced the best results in both immediate posttests and delayed posttests.

The four studies yielded findings consistent with the cognitive theory of multimedia learning (Mayer, 2001), which argues that learners can construct referential connections between verbal and visual forms of mental representations and thus learn more effectively.

There are other studies addressing the effects of annotations on vocabulary learning within the theoretical framework of attention. For example, Bowels (2004) and Yanguas (2009) investigated, from the perspective of Schmidt's (1990) noticing hypothesis, whether noticing of words with different types of annotations led to better learning of the words. Bowels (2004) compared the effects of three annotation conditions on vocabulary acquisition and reading comprehension. Fifty participants, who were college students enrolled in Spanish-language courses, were randomly assigned to one of three groups (no annotation, marginal annotation, and hyperlinked annotation), where they read the same text. Vocabulary learning was assessed through a recognition test (multiple-choice) and a production test (write equivalents in Spanish). Results indicated that the two annotation groups performed significantly better than the control group on the two measures. Yanguas's (2009) study is an extension of Bowels's study in that multimedia annotations were involved. Ninety-four college-level L1 English participants read a Spanish text under one of four annotation conditions (no annotation, text annotation, picture annotation, and text-picture annotation). Similar results emerged from this study that all the annotation groups noticed and recognized significantly more of the target words than the control group. These results support the noticing hypothesis (Schmidt, 1990). According to the hypothesis, noticing is the essential starting point for language acquisition, and in order to learn new linguistic forms, learners must first notice the linguistic features or forms from the language input (Schmidt, 1990). Annotations, by giving salience to the unknown words, render the annotated words more noticeable, and thus draw learner attention toward unknown words. However, both studies failed to find a significant difference between annotation groups in performance on recognition or production tasks. Cognitively, the annotations presented both visually and textually could be expected to lead to better

performance than those presented singly. One possible explanation is that the participant did not attempt deeper processing of the annotated words when their main concern was to comprehend the text, that is, they noticed targeted words at a low level of awareness.

2.3.3 Annotation and reading comprehension

In the context of CALL reading, vocabulary annotations are meant to aid lower-level reading processes, which will in turn lead the reader to decode the text as a whole. Much research has been done in this area. However, in contrast to general consensuses on the potentials of annotations for vocabulary learning, previous studies that explored the effects of annotations on reading comprehension have yielded inconsistent findings.

Some studies have found positive effects of annotations on reading comprehension. For example, Lomicka (1998) conducted a study that explored how annotations influenced the level of comprehension. Twelve American college students (French being L2) were randomly but evenly assigned to one of the three annotation groups: no annotation, L1 and L2 word annotation, and full annotations (including L1 and L2 word definitions, images, references, questions, and pronunciation). They were asked to think aloud when reading an excerpt of a poem on the computer screen, and a tracker recorded their interaction with the annotations. An analysis of the data from the think-aloud protocol showed that the students gained more understanding when they had access to full annotations. Therefore, the study suggested that reading with different types of annotations promoted a deeper level of text comprehension. But the results were not significant due to the small number of participants in each group ($n = 4$). Testing with larger numbers of participants is required to explore this further. The study by Chun and Plass (1996a), which is actually the same study as Chun & Plass (1996b) reviewed above but with a different research focus, found that when both

visual and verbal vocabulary annotations were looked up, the comprehension, measured by recall protocols, was better than when only verbal annotations or no annotations were accessed. This finding supports the dual coding theory (Paivio, 1990) and its extension to the multimedia model of learning. Two other findings of the study also deserve mentioning. One is that exposure to video preview of the major ideas was linked to a greater extent in the recall protocols than for no exposure, suggesting that visual advance organizers (providing background knowledge) are facilitators of comprehension. One possible reason that might account for this relation was that visual forms of information were more easily remembered and better at helping learners to build mental connections between images and words, without which "information is harder to organize in WM and consequently harder to link to structures in long-term memory (Chun & Plass, 1996a, p. 515). The other is that no significant correlation was found between look-up behavior of vocabulary words and overall comprehension. This might be because readers with larger vocabularies looked up fewer words and they could draw more on inference strategies in determining the meaning of unknown words than on access to annotations.

The issue of the relationship between annotation use and reading comprehension has been addressed from different perspectives. Yanguas' (2009) study (described in the previous section), which was based on the noticing hypothesis, also investigated the impact of different types of annotation on reading comprehension. The comprehension task administered consists of 11 multiple choice questions. Results revealed that the combination annotation group significantly outperformed all other groups in reading comprehension, another piece of evidence showing the positive efficacy of annotations for reading comprehension from a different theoretical perspective. Unlike this study, Akbulut (2008) investigated factors influencing L2 reading comprehension in a hypermedia

environment within the theoretical framework of dual coding and cognitive load theories. The participants ($n = 54$), who were senior students at a university in Turkey, read an English (L2) hypertext with word and topic annotations in the forms of text, graphics, audio, and video, to provide information both at the word level and topic level. Then, they were required to write down everything they could remember. It was found that text lookup explained approximately 20% of the variation in the recall score on its own and topic interest explained an additional 7% of the variation. This finding supports the cognitive load theory (detailed in Section 2.4.3), because the provision of annotations reduces the intrinsic load of the materials. It was also found that videos and textual information were the most useful types of annotations, which is consistent with the dual coding theory.

With a different research focus, Abuseileek (2011) explored whether hypermedia annotation locations impact vocabulary acquisition and reading comprehension. Seventy-eight undergraduate ESL learners were randomly and roughly evenly assigned to one of the five groups: (a) traditional glosses at the end of the text, (b) computer-based glosses in the margin, (c) the computer-based glosses at the bottom of the screen, (d) computer-based glosses in pop-up windows, and (e) no glosses. The reading comprehension test followed a recall protocol format, where the learners were required to summarize in English the reading text. Results showed that the annotation groups outperformed the no-gloss group in terms of the performance on text memory. This finding supports the positive relationship between annotations and reading comprehension.

Some researchers alert us to the disadvantages of reading with information presented in multiple modes. Mayer and Anderson (1991) cautioned that a visual presentation mode (e.g., animation) accompanied by a verbal presentation mode (e.g., text) may lead learners

to split attention between both types of information, which may have deleterious effects on learning. Kirby (1993) echoed this view, saying that learners exposed to both visual and verbal information may direct their attention to the type of information perceived as more important or more interesting, away from the other mode which may in fact contain more important information. According to Rouet and Levonen (1996), hypertexts may impose a cognitive load on readers, who must know where they are, decide where to go next, and keep track of where they have been in a network. There is evidence from empirical research that shows annotations affect reading comprehension negatively.

Ariew and Ercetin (2004) explored the relationship between the use of different types of hypermedia annotations and reading comprehension. Their participants ($n = 84$), who were intermediate and advanced level adult ESL learners, read an L2 text provided with annotations in the form of text, picture, graphics, and video. A tracking tool was used to record the amount of time they spent on a given annotation. The reading comprehension was measured through multiple-choice questions, short-answer questions and open-ended questions. Results indicated that annotation use did not contribute to reading comprehension for the advanced group, and a negative relationship was found between the time spent on video and graphic annotations and reading comprehension for the intermediate group. However, one factor that may challenge the results is that the reading text was not available to the participants during the test, for it has been argued that assessments that do not allow look-backs may actually underestimate a student's level of comprehension (Leslie & Caldwell, 2009). This study was replicated later by Sakar and Ercetin (2005), with the participants (44 intermediate-level learners) being the only major difference. They also found a negative relationship between reading comprehension and either the frequency of annotations access or the amount of time spent on annotations. In

particular, videos affected reading comprehension negatively, corroborating the findings in the study by Ariew and Ercetin (2004). Since the learners in these studies were exposed to multiple types of annotations, it is speculated that the negative relationship was due to the split-attention effect that usually occurs with multimedia learning (DeStefano & LeFevre, 2007; Yeung, Jin, & Sweller, 1997).

The split-attention effects were observed only in multimedia annotations, but also in multiple annotations that do not use multimedia. Wallen, Plass and Brunken (2005) designed three types of annotations for a reading text to support one of the three cognitive processes described in the cognitive theory of multimedia learning (Mayer, 2001): selection-level annotations (word or concept annotations), organization-level annotations (phrase or idea annotations), and integration-level annotations (paragraph annotations by showing the links of the ideas in the paragraph). The participants ($n = 98$) were randomly and roughly evenly assigned to one of the seven groups: one control group, three single-annotation groups, and three double-annotation groups. The participants were required to click all the annotations available and participants who consulted fewer than 75% of the annotations were excluded from the analyses. Three tests were administered to the participants, each to measure factual knowledge (recognition test), conceptual understanding (comprehension test) and higher-level understanding (transfer test). Results revealed that generally single types of annotations led to better learning outcomes; in contrast, the double types of annotations resulted in lower performance, in particular in tests of higher-level processing. One reason that might account for the findings is that multiple annotations caused extrinsic cognitive load, especially when they did not complement each other, as suggested by the cognitive theory of multimedia learning. However, the small sample size in each group (about 10 on average) is a caveat of the

conclusion of the study.

In addition, pictorial annotations have been reported to be negatively associated with reading comprehension. For example, Plass, Chun, Mayer, and Leutner (2003) compared the effects of different annotation conditions on reading. This study was based on the cognitive theory of multimedia learning (Mayer, 2001). The participants ($n = 152$), who were L2 college students, were randomly assigned to each of the four annotation groups: no annotation, verbal annotation, visual annotation, and verbal plus visual annotations. The verbal annotation of a word was an L1 translation, and the visual annotation was a corresponding image of the word. In the verbal annotation condition, students could find an English translation. One of the tasks performed by the participants was writing a summary of the text in L1. Results indicated that the performance was equally good in the no-annotation condition and verbal plus visual annotation condition, but worst in the visual annotation condition. One reason that might explain the results is that visual annotations imposed a high cognitive load, because the participants had to select the relevant information from the images to understand the words, and the decision making in the process consumed additional mental resources (DeStefano & LeFevre, 2007).

Finally, some researchers claimed that annotations did not contribute to reading comprehension. For example, Davis and Lyman-Hager (1997) examined the relationship between computerized annotations and reading comprehension. The participants ($n = 42$) were undergraduate English (L1) learners of French (L2). They read a text with various types of annotations: they could click on words or expressions to obtain L1 and L2 definitions, use a pronunciation feature to hear words, access grammatical notes and explanations to enhance textual understanding, click on a question control button to enhance global comprehension, and click a cultural reference control button to get cultural

knowledge. An unobtrusive query log was included in the reading program to provide

information on all the lookup behaviors. Their reading comprehension was measured

through a multiple-choice test and a recall protocol in English. It was found that as the

course grade (considered as a rough measure of language level of the students) increased,

the scores on the comprehension measures also increased, whereas the frequency of access

to the annotations decreased. The results suggested that course grade significantly

predicted performance on reading comprehension tests, and the information from the

annotations had no effect on comprehension. However, there are two limitations in the

research design of the study which might lower its internal validity. First, there was no

pre-test of the annotated words before the exposure. Second, the time on task was not

controlled in the study, as indicated by the author. In addition, the log data revealed that

the students tended to consult almost exclusively word definitions in English although

there were other types of annotations available. This might provide an explanation of the

findings, for L2 readers require much more informational support than merely word

definitions, as suggested by the interactive model of reading. The participants'

comprehension might have been improved if the students had accessed other types of

annotations as well. Therefore, it seems necessary to instruct learners on how to effectively

use annotations available.

Conclusive evidence has not been found in favor of any type of annotation for reading

comprehension. As reviewed above, any conclusion about the correlation should be closely

considered. These studies approached the issue from different perspectives and utilized

different research designs such as reading measures, sample size, target language,

annotation types, and annotation lookup requirement, which together make it difficult to

interpret the results as a whole. However, they have revealed the complexity of L2 reading

with annotations, and suggested a need for more research.

2.3.4 Learner perception of annotations

Previous studies on CALL reading have not only addressed the effectiveness of annotations on vocabulary acquisition and reading comprehension, but also involved learners' perception of annotations. First, there is evidence that learners tend to use annotations which are easily accessible. For example, Aust, Kelley, and Roby (1993) found that when electronic annotations were readily accessible, readers used them more frequently to look up even familiar information. This is confirmed by Ariew and Ercetin' (2004) study, where the participants, when interviewed, indicated that they tended to use almost all of the electronic annotations because of their accessibility. Similarly, Roby (1999) found that learners who had access to electronic dictionaries looked up substantially more words than those who had access to paper dictionaries.

Second, learners show preference among different types or forms of annotations. For example, in Davis and Lyman-Hager's (1997) study, the participants' log data showed that 85% of the information accessed was L1 definitions of words. This is consistent with Lomicka's (1998) finding that the learners preferred definitional annotations to other types of annotations. Cooledge (2004) found that when both L1 and L2 word annotations were available, readers rarely stopped at L2 annotation, but opted for L1 annotations in most cases. Therefore, it seemed that learners viewed L1 definitional annotations as being more useful. Some researchers analyzed learners' preferences in terms of their language proficiency. For example, Ercetin (2003) explored what types of annotations learners prefer to use when they are engaged in reading a hypermedia text, and whether there are any differences between intermediate and advanced learners with regard to their use of

hypermedia annotations. The data obtained from tracking records, questionnaires, and interviews indicated all the learners ($n = 84$) favored definitions over graphics to understand words, and videos and graphics to get extra information about the topic; however, the intermediate learners consistently made use of both textual and contextual annotations more frequently than the advanced learners, which suggests that they were trying to compensate for their lower level of language proficiency by using annotations. The advanced learners' less frequent access to annotations confirmed Knight's (Knight, 1994) and Ercetin's (2003) findings that lookup frequency and time spent on annotations decreased as proficiency level increased. Sakar and Ercetin (2005) found that intermediate-level learners preferred visual annotations (i.e., graphics and videos), and regarded them as being highly helpful. As to the location of annotations, Abuseileek (2011) found that 95% of the participants preferred hypermedia annotations. Specifically, 78.8% preferred the marginal glosses, 15% the bottom-of screen glosses, 5% the pop-up window glosses, and only 1.3% the end-of-text glosses. This is because marginal glosses are the closest to the glossed word, permitting an easier and faster access.

Some researchers have explored the reason why learners prefer annotations. First, with multiple forms of media providing textual and contextual annotations, reading has become more enjoyable and interesting (Ercetin, 2003; Tozcu & Coady, 2004), which positively affects learners' motivation for and interest in reading. Second, learners find access to annotations highly useful in helping them cope with an L2 text, because it lessens the disruption of the reading process usually caused by conventional dictionary look-ups (Hwang, Wang, & Sharples, 2007) Third, annotations help them recognize words quickly and easily, so reading becomes easier and more comprehensible (Sakar & Ercetin, 2005; Tozcu & Coady, 2004). In addition, learners find the reading progress to be more

interactive in that they can choose annotations according to their needs or interests (Ercetin, 2003; Milim & Shen, 2006), and they can get immediate feedback about their errors (Nagata, 1999). These positive views towards annotations are likely to increase the chances of reading success.

2.4 Working memory and CALL reading

2.4.1 Theories of working memory

WM is assumed to be a temporary storage system under attentional control that underpins our capacity for complex thought (Baddeley, 2007). Traditionally, researchers differentiated between short term memory and long term memory, a distinction based on the assumption that either form of memory is solely a storage space which retains information until it is extracted (Abu-Rabia, 2003). Nowadays what was formerly known as short term memory is generally viewed as WM (Baddeley, 1986; Carpenter & Just, 1989; Daneman & Carpenter, 1980). The change in terminology stems from the notion that WM is responsible not only for the storing of information but also for the processing of information with previously acquired knowledge, so the term "working memory" has been developed as a way to refer to a more active part of the human processing system (Newell, 1973, cited in Daneman & Carpenter, 1980).

There are many approaches to WM, and one of the most salient that has contributed to our understanding of L2 learning is Baddeley's multi-component model (Chun & Payne, 2004), originally developed by Baddeley and Hitch (1974). According to the model, WM has four components (see Figure 2.3). The central executive component is the most important, which functions to regulate information flow within WM, to retrieve information from other memory systems such as long-term memory, and to process and

store information. The processing resources used by the central executive to perform these

functions are, however, limited in capacity (Baddeley, 1986). The efficiency with which

the central executive fulfills a particular function therefore depends on whether there are

other demands simultaneously placed on it. The greater the competition for the limited

resources of the executive, the more its efficiency at fulfilling particular functions will be

reduced (Gathercole & Baddeley, 1993). The central executive has two "slave systems" ---

phonological loop and visual-spatial sketchpad. The phonological loop is specialized for

the storage of speech-based and possibly purely acoustic information, and the storage is

assumed to be "dependent on a memory trace that would fade within seconds unless

refreshed by rehearsal" (Baddeley, 2007, p. 7). The visual-spatial sketchpad is involved in

short-term processing and maintenance of material which has a strong visual or spatial

component (Baddeley, 1986). The episodic buffer is a limited-capacity store holding
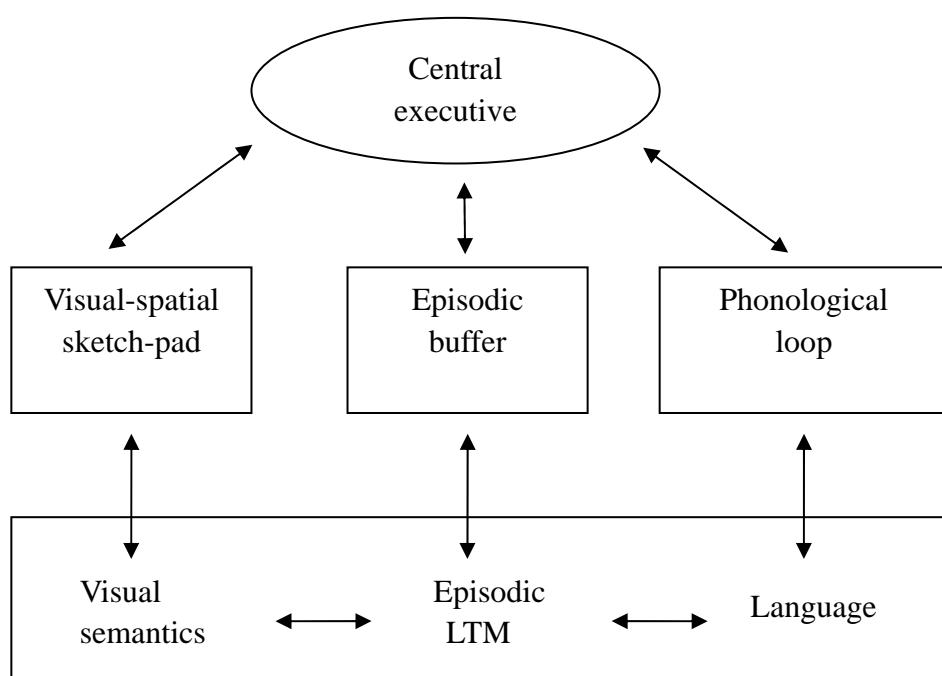
Figure 2.3 *The Multi-Component Model of Working Memory*
(adapted from Baddeley, 2003)

information that is multimodal, enabling information from other stores and from long-term memory to be synthesized and stored as an episodic representation (Baddeley, 2000). When all the components are taken together, WM is a system which involves the concurrent processing and storage of information. The study of WM is thus concerned with the way in which the temporary manipulation and storage of information is carried out. Mayer (2001) argues that Baddeley's multi-component model of WM sits well with the multimedia model of learning, and according to Grimley and Banner (2008), the model is very good at explaining how multimodal (multimedia) information is integrated into real-world learning.

The inception of this model has provided a useful theoretical framework for investigating a wide range of human activities, at the same time generating both controversy and progress in the task of understanding the role of memory in our capacity to think (Baddeley, 2010, p. 140). More important, it initiates tremendous enthusiasm towards conceptualizing WM constructs among cognitive psychologists, who with varied research perspectives have formulated a dozen other models (Miyake & Shah, 1999). For example, in the domain of complex reading, Just and Carpenter (1992) proposed a linguistic approach to WM, such that linguistic WM directly constrains the operation of language comprehension processes, a theory called "Capacity Constrained Comprehension". According to Just and Carpenter (1992), both storage and processing are fueled by the same commodity: activation. Individual differences in WM are then explained in terms of "total capacity" and also in terms of "processing efficiency". The total capacity explanation posits that individuals vary in the amount of activation they have available in WM for storage and processing. The processing efficiency explanation posits that some individuals have more efficient mental processes than others, which demands relatively little from the

total WM resources, thus leaving more capacity for the storage of products of the task being performed. Engle and his colleagues (Conway et al., 2005; Engle, Cantor, & Carullo, 1992; Turner & Engle, 1989) hold that WM is a general and domain-free system independent of any one processing task, a hypothesis referred to as the General Capacity Theory. According to the theory, individual differences in WM correspond to differences in general attentional resources used to perform effortful processing.

The theoretical arguments over WM are a long way from being conclusive. However, Wen (2012) points out that most current WM models have a lot more in common concerning the nature of the WM construct, and research from different perspectives is complementary in nature. According to Miyake and Shah (1999) the differences between these models arise from their respective foci with a specific component of the same construct of WM, rather than reflecting fundamental differences in their theoretical stance. In addition, in spite of the debates over its conceptualizations, WM has generally been recognized as a limited capacity, which differs among individuals, and the difference affects a wide range of complex cognitive tasks such as problem-solving, reasoning, and reading comprehension (Alptekin & Ercetin, 2009; Baddeley, 2007; Cantor & Engle, 1993; Daneman & Carpenter, 1980).

2.4.2 Measurement of working memory

Harrington defines WMC as "the relative capacity to intake and integrate information in immediate, on-line processing" (1992, p. 123). WMC should be construed as operational capacity that can be measured. There are mainly two categories of WM measures adopted in WM research: simple span task and complex span task. The variants of simple span task typically include word span task (La Pointe & Engle, 1990), digit span task (Turner &

Engle, 1989), and non-word repetition task (Gathercole, 2006). The first two involve the recall of sets of unrelated words or numbers, and the third involves the repetition of a string of letters that do not exist in the given language but still conform to its phonotactic rules (Wen, 2012), and thus measures the phonological loop WMC. Because of the nature of these tasks, they primarily tap storage of information (Juffs & Harrington, 2011). In view of the research findings that simple span tasks did not predict L1 reading comprehension (Turner & Engle, 1989) and that L2 simple span measures have little relationship with L2 reading skill (Harrington, 1992), the present study does not use these simple span tasks, and thus will not review them.

Complex span tasks involve both the storage and processing functions of WM. A widely used complex span task is the reading span test (RST) developed by Daneman and Carpenter (1980), which is also known as WM span test (Baddeley, 2007), or sentence span test (Abu-Rabia, 2003). This task is believed to be a suitable measure for the central executive of WM in that it presumably taxes the dual functions of storage and processing that are subsumed by the executive component of WM (Daneman & Merikle, 1996; Friedman & Miyake, 2005). In the RST, the subjects are required to read aloud three sets each of two, three, four, five, and six sentences while simultaneously trying to remember the final word of each sentence in the order in which they occur. The task is terminated once a subject fails all three sets at a particular level. The subject's reading span is determined by the level at which a subject is correct on two out of three sets. This task has become the foundation of a large literature in the research into the psychology of reading and comprehension for adults (Juffs, 2006).

Following Daneman-Carpenter RST, several other WM span tasks have been devised. One of them is the operation span task (OST), first proposed by Turner and Engle (1989).

In the task, the subjects perform simple arithmetic operations followed by a

to-be-remembered word. Each string consists of two arithmetic operations and a stated

final answer. The first operation is a simple multiplication or division problem in

parentheses, while the second operation is a simple addition or subtraction of a single-digit

integer. The number of operation strings in a trial gradually increases, from 2 to 5, with

three trials at each set size. Subjects are required to verify whether each of the stated

answers is correct or incorrect, and then write the to-be-remembered words in order or in

any order. The span is determined by the maximum size of the set of to-be-remembered

words in which the subject can correctly recall the memory words two out of the three

times. According to Guan (2007), OST is a measure more of the domain-general capacity,

rather than the content-specific capacity aspect of WM. OST replaces the sentences in

Daneman-Carpenter RST, but the task demands are unchanged. Therefore, it is believed to

be a good predictor of comprehension (Daneman & Merikle, 1996).

Literature has documented many variants of the RST and OST, which are collectively

known as "verbal and math process plus storage measures" (Daneman & Merikle, 1996).

In the case of the verbal process plus storage measures, the commonly used variants add a

comprehension check to the processing component of the task: subjects are required to

make a true-or-false judgment, or a sensibility judgment, or grammaticality judgment

about the sentence, or to recall another word following each sentence, or to recall a

different word in the sentence. In the case of the math process plus storage measures, the

commonly used variants substitute a solution for the stated answer, or replace the

to-be-remember word with a letter. Another variant of the RST is the use of L2 (e.g.,

Harrington & Sawyer, 1990) instead of L1.

All these variants of complex WM tasks have been argued by some researchers to

have good reliability and validity, and they likely reflect a common construct (Conway et al., 2005; Unsworth, Heitz, Schrock, & Engle, 2005; Waters & Caplan, 1996). That is, although these tasks differ in the types of materials used to assess WMC, performance on these tasks have been shown to be reliable indicators of a broader WM construct. For example, in Turner and Engle's (1989) study, four span tasks (a sentence-word span task, a sentence-digit span task, an operation-word span task, and an operation-digit span) were found to be correlated with one another. Through a meta-analysis of the data from 6,179 participants in 77 studies, Daneman and Merkle (1996) concluded that measures that tap the combined processing and storage capacity of WM are better predictors of comprehension than those that tap only the storage capacity. After reviewing the popular span tasks, Conway et al. (2005) drew a conclusion that irrespective of what WM span tasks are supposed to measure, evidence suggests that they measure with reasonably accuracy, whatever it is that they actually measure (p. 776), and they show considerable construct validity insofar as they predict performance on a wide array of tasks for which control of attention and thought are important (p. 778). This conclusion is consistent with the total capacity explanation of WM by Turner and Engle (1989) that the total capacity of WM, independent of the efficiency of the specific tasks being performed, drives the significant relationship between WM measures and reading comprehension measures.

However, different tasks are also arguably suitable for different cognitive behaviors. For example, Guan (2007) contends that the RST measures the WM involved while reading and the OST is somewhat inferior to the RST in measuring WM in relation to reading comprehension. Furthermore, an important issue has arisen concerning L1 and L2 RSTs. On the one hand, researchers observed reliable correlations between them (e.g., Berquist, 1997; Harrington & Sawyer, 1992; Miyake & Friedman, 1998; Osaka & Osaka,

1992). According to Osaka and Osaka (1992), WMC is a domain-general ability rather than a language-specific one. On the other hand, however, Osamu (2006) suggests that L2 WMC (not L1 WMC), rather than a general language-based WMC, is related to L2 reading (p. 119), and this relationship is mediated by differences in the efficiency with which individuals process a second language, although this processing efficiency cannot alone explain a large proportion of individual differences in working memory capacity (p. 120).

To date, there has been no consensus on the measurement of WM (Alptekin & Ercetin, 2009; Juffs & Harrington, 2011), for no one measure is recognized to be perfect. The RST has received some criticisms for negligence of the semantic and syntactic acceptability and the accuracy of sentence verification (Turner & Engle, 1989; Waters & Caplan, 1996), and for the small score range (between 2 and 6), which greatly limits the sensitivity of the measure (Oberauer & Sub, 2000). Researchers have attempted to improve the measure in one way or another. Some studies (e.g., Hong, 2007; Turner & Engle, 1989; Unsworth et al., 2005) excluded participants from analyses who scored below 80% or 85% in sentence verification. Others (e.g., Alptekin & Ercetin, 2009; N. P. Friedman & Miyake, 2005) used composite scoring: the total number of words recalled across all trials (representing the measure of storage capacity) and the total number of sentences accurately judged (representing the measure of processing capacity) were converted to z-scores and their average indicates the WM composite score. The composite scoring approach, however, is not without limitations, especially in studies on adults. According to Conway et al. (2005), when the subjects are instructed to attend to the verification task, the processing accuracy is typically close to ceiling in light of the simplicity of the sentences.

Considering the limitations in the current achievement of WM study, it seems that an optimal research strategy is to administer more than one WM span task and then use the

average or composite scores of all the tasks as the measure of WM. Just as Conway et al.

(2005) suggest, "when a construct is measured with imperfect tools, it is best, from a

measurement standpoint, to use multiple, reliable measures that do not replicate one

another" (p. 780). The present study will follow their suggestion in assessing the WM of

the participants by using an L1 RST, an L2 RST and an OST.

2.4.3 Working memory and cognitive load

The term "cognitive load" refers to the amount of mental energy required to process a

given amount of information (Feinberg & Murphy, 2000). It is used in cognitive

psychology to illustrate the load related to the executive control of WM. Sweller (1988)

proposed the cognitive load theory, which has developed into one of the most important

theories in educational psychology. The theory emphasizes the inherent limitations of

concurrent WM load on learning. Therefore, it primarily deals with the amount of

information that individuals are able to store and process in their WM before transferring it

into the long-term memory (Sweller, 1994). According to the theory, in complex learning

activities the amount of information that must be processed simultaneously can either

under-load, or overload the limited amount of WM that one possesses. Thus, WM theories

provide a useful way of operationalizing the construct of cognitive load because a common

assumption of WM models is that a limited amount of information can be simultaneously

processed (Baddeley & Logie, 1999).

Researchers (e.g., Chandler & Sweller, 1991; Sweller, 1988) have conceptualized

three types of cognitive load that affect WM: intrinsic, extraneous and germane cognitive

load. Intrinsic load is an intrinsic part of the material to be learned; it is incurred by

complexity of the material relative to the learner's prior knowledge, or the amount of

information a learner needs to hold in WM. Extraneous load is caused by the manner in which information is presented to the learner and the activities in which s/he is required to engage. When the load becomes unnecessary, it does not contribute to, but rather interferes with learning. Germane cognitive load refers to the effort that the learner applies when trying to process or/and understand the material. It occurs when free WMC is used for deeper construction and automation of schemata (Penkl & Atkinson, 2003).

As regards the relationship among the three types of cognitive load, Paas, Renkl, and Sweller (2003) argue that they are additive in that the total load together cannot exceed the WM resources available if learning is to occur. Intrinsic cognitive load is irreducible other than by alleviating the difficulty of the material or by constructing new schemas. Any WMC available after mental resources have been allocated to deal with intrinsic cognitive load can be allocated to deal with extraneous and germane load (Brunken, Plass, & Leutner, 2003). Thus, if learning is improved by an instructional design that reduces extraneous cognitive load, the improvement may have occurred because the additional WMC freed by the reduction in extraneous cognitive load has now been allocated to germane cognitive load (Paas, Renkel, & Sweller, 2004, p. 2). Germane cognitive load is associated with processes that are directly linked to learning. WM is limited, and thus can be overloaded. When the cognitive load exceeds the capacity and limitations of WM, learning will be inhibited. For this reason, it can be claimed that WMC and its constraints are determinants of instructional design effectiveness. This has important implications for the design of annotations in a CALL reading context. It is a principle to follow that the design of annotations as learning aids should manipulate intrinsic load, decrease extraneous cognitive load and increase germane cognitive load.

The cognitive load theory has become a major theory providing a framework for

investigation into the instructional design for learning. To take reading for example, if a learner must divide his or her attention between different sources of information that need to be integrated before comprehension can occur, a split of this kind is likely to induce an extra cognitive load, and thus results in deterioration in reading performance. This split-attention effect is more likely to emerge in a multimedia environment where reading materials are presented in multiple forms of media such as text, picture, and video. In a similar vein, multimedia materials which integrate different sources of information that could be understood in isolation are likely to increase cognitive load and create a redundancy effect, thereby interfering with learning. Therefore, the cognitive load theory is a highly effective guide for the design of multimedia learning, and is included as one of the bases of the cognitive theory of multimedia learning (Mayer, 2001).

There are a host of studies on instructional design in CALL environments in relation to cognitive load on WM (Al-shehri & Gitsaki, 2010; DeStefano & LeFevre, 2007; Kalyuga, Chandler, & Sweller, 1999; Kirschner, 2002; Plass et al., 2003; van Bruggen, Kirschner, & Jochems, 2002). For example, Al-shehri and Gitsak (2010) investigated the influence of split-attention and integrated instructional formats of an online text on students' cognitive load and how they might facilitate L2 online reading. They found that the students received higher reading scores when each of the comprehension questions was inserted in the part of a text where there was information related to the question than when the text was shown on one webpage and the comprehension questions on another. This study supports the use of integrated reading formats in reading comprehension, which is likely to reduce the extraneous cognitive load caused by navigating between web pages. Researchers (Feinberg & Murphy, 2000; Rowland, 2008) have warned that when learners must split their attention between the materials, a higher extrinsic load is generated. The

present study involves multimedia annotations in a CALL environment, and their impact

on reading comprehension will be discussed with reference to the theories of cognitive

load and WM.

2.4.4 Working memory and reading comprehension

WM plays a crucial role in a wide range of complex cognitive activities. However, its

function in language comprehension is especially evident because language comprehension

entails processing a sequence of symbols that is produced and perceived over time (Just &

Carpenter, 1992, p. 122). According to Lee and Tedder (2003), the crucial ability in

reading comprehension is the construction of integrated mental presentations, and this

process makes heavy demands on both the processing and storage functions of WM.

However, as WMC is limited, its two functions, processing and storage, compete for the

limited capacity available to WM. Individuals differ in the ability to coordinate the

processing and storage functions, and this may be a major source of individual difference

in reading comprehension (1992). Abu-Rabia (2003) argues that individuals with low WM

are less able to maintain necessary information in an active state than those with high WM,

and thus have poorer text integration ability. Similarly, Daneman and Carpenter (1980)

claimed that low WM learners are less able to integrate successively encountered words,

phrases, and sentences into a coherent representation, which leads to deficits in

comprehension. There is a substantial body of evidence in the literature supporting a

straightforward relationship between WM and comprehension ability, with higher WM

readers outperforming lower WM readers in a variety of reading components. These

include: retrieving facts and computing pronominal references (King & Just, 1991);

processing syntactically complex sentences (Daneman & Carpenter, 1980); drawing

inferences (Whitney, Ritchie, & Clark, 1991); reasoning ability (Kyllonen & Christal, 1990); maintaining multiple, tentative interpretations and use text elements to test those interpretations (Lee-Sammons & Whitney, 1991); carrying relevant information over from one sentence to the next in a passage (Daneman & Carpenter, 1983); and integrating text (Yuill, Oakhill, & Parkin, 1989).

In the past two decades, researchers have shown great interest in the relationship between WM and L2 reading comprehension. For example, with a focus on the relationship between L2 WM and L2 reading comprehension, Harrington and Sawyer (1992) administered a battery of WM tests (including digit spans, word spans, and reading spans in both L1 Japanese and L2 English), and a reading comprehension test (consisting of the Grammar and Reading Sections of the TOEFL and a cloze passage), to advanced (graduates) Japanese learners ($n = 34$). The RST tests were based on the Daneman-Carpenter RST, but the 42 sentences are shorter and syntactically simpler in order to avoid possible floor effects in performance. A grammatical judgment task was incorporated to ensure that the subjects processed the sentences for meaning while reading. The number of words correctly recalled was taken as the WMC in the RST measures. Results showed that the L2 reading span measure had a strong correlation with the TOEFL Grammar ($r = .57$) and TOEFL Reading ($r = .54$) and a weaker correlation with the cloze ($r = .33$). These findings were later confirmed in a similar study by Osamu (2006). They are similar in terms of the RST (L1 Japanese and L2 English) and the reading comprehension test (TOFEL exam) employed. Osamu (2006) observed a reliable L2 RST-TOFEL correlation ($r = .329$), and attributed the lower magnitude of the correlation than that ($r = .54$) found in Harrington and Sawyer (1992) to the lower proficiency level of his participants (undergraduates). Results of these studies indicate that L2 WMC correlates

with L2 reading comprehension.

Further evidence can be found in a study by Tomitch (1999), which investigated the relationship between reading ability, WMC and readers' use and recall of the mechanism of prediction. The participants ($n = 12$), who were Brazilian speakers of Portuguese, read texts containing the mechanism of Prediction. Their reading ability was measured through free recall and comprehension questions, and their WM was assessed following the Daneman-Carpenter RST. Results showed that high span readers recalled both the predictive signal and all the predicted elements correctly, and more propositions from the text; low span readers, by contrast, simply disregarded the predictive signal and recalled only some of the predicted elements and less information. These finding are consistent with the literature indicating WM is a good predictor of L2 reading comprehension. However, one limitation of the study that might pose a challenge to its findings is the small number of participants.

Recently, Alptekin and Ercetin (2009) approached the relationship between WM, operationalised as composite scores, and L2 reading comprehension, compartmentalized into literal and inferential dimensions of understanding. The L2 RST, modeled after the Daneman-Carpenter RST has two versions. One involved a recall task where the participants (30 undergraduates whose TOFEL scores were all above 550) were required to remember the sentence-final words, while the other involved a recognition task where they were required to choose the sentence-final words from a list of options. To obtain composite scores, the storage scores (the total number of words recalled or recognized) and the processing scores (total number of sentences correctly judged) were converted into z–scores, and then averaged. Comprehension of the text was measured with a multiple-choice test, consisting of 10 textually explicit questions and 10 textually implicit

questions. Results indicated that WM correlates with inferential understanding when recall-based tasks are employed to measure storage. One justification is that literal reading is relatively less demanding in terms of intrinsic cognitive load; in contrast, inferential reading which involves performing a series of complex tasks associated with textually implicit information, requires more cognitive resources. This study points to a possible correlation between L2 proficiency level and the contribution of WM to reading comprehension in the L2, which confirms Osamu's (2006) claim that that L2 WM is related to L2 reading. In this light, L2 literal reading could be associated with WM only for low L2 proficiency readers. Therefore, this study has important implications for the assessment of reading comprehension.

2.4.5 Working memory, annotation and reading comprehension

One prominent feature of CALL reading is the rich access to different types of annotations provided to facilitate the reading process. As reviewed in Section 2.3.3, the literature is inconclusive as to the effect of annotation use on reading comprehension. For a deeper understanding of the issue, researchers have begun to examine factors in individual differences that may mediate the effect. Individual differences are enduring personal characteristics that are assumed to apply to everybody and on which people differ by degree (Dornyei, 2005). In the field of language learning and acquisition, individual differences represent an issue of both theoretical and pedagogical significance. According to Chun and Payne (2004), learners' annotation look-up behaviors, different types of annotations and their impact on learning, and individual differences among learners are important factors to consider in studying the effectiveness of CALL for reading comprehension. Through a meta-analysis of 11 studies of computer-mediated annotations

in L2 reading comprehension, Abraham (2008) asserts that the study of the relationship between learners' individual differences and the effective design of multimedia environments to facilitate L2 text comprehension and vocabulary learning has become a central concern for CALL reading research. There have been some important findings from empirical studies along this line. For example, Plass at al. (1998) found that the participants' reading performance was better when they had the opportunity to receive their preferred mode of annotation. This suggests that individual differences in cognitive style may play a role in determining the effectiveness of annotation use for reading comprehension.

As discussed in the previous sections, WM, as a variable in individual differences, play a significant role in reading comprehension. The literature contains one study on the role of WM in mediating the efficacy of annotation use for reading comprehension. This study (Chun & Payne, 2004) investigated relationship between WM, annotation look-up behavior, vocabulary learning and text comprehension. The participants ($n = 13$) were English-speaking undergraduates enrolled in a second-year German language course. They were asked to read a story with 27 difficult words annotated (9 in English definitions, 9 in still images, and 9 in video-clips), and then complete a vocabulary test and a reading comprehension test on the first day, and write a recall protocol on the next day. The reading comprehension test consisted of 8 multiple-choice questions and 4 drag-and-drop questions, all being discrete-point questions in nature. Their WMC was measured on the third day through a non-word repetition test, which measured the phonological loop WMC (PWMC). In addition, a modified version of the Daneman-Carpenter RST was used. Results indicated that the participants with low PWMC looked up annotations more frequently, suggesting that these students use the multimedia look-up features to

compensate for WM constraints while reading and learning new vocabulary. However, individual differences in WM, as measured by the RST, did not account for multimedia look-up behavior. This could be explained by the reading span being generally associated with comprehension but not necessarily with vocabulary learning. No significant differences were found on any of the three performance measures between high WM (HWM) and low WM (LWM) groups, identified by either the no-word repetition test or the RST. These results seem to suggest that the WM was not related to reading comprehension, vocabulary learning, or the recall of text. However, they should be considered tentative in light of its apparent limitations. First, the sample size ($n = 13$) was very small. Second, the time spent on each task was not controlled. Third, some students only partially completed the reading comprehension test, resulting in several missing data. Furthermore, the participants did not seem to be taking the tests very seriously, as confirmed by the instructor administering the treatment. Accordingly, the conclusion from the study is far from convincing. Future studies that use a larger sample and complete data on the reading comprehension measures might generate different sets of results.

No other studies have been located in literature on WM as an independent variable in examining the efficacy of annotation use for reading comprehension. Given the fact that in Chun and Payne's (2004) study, the words were only annotated in a single way, it could be argued that there is no study that has explored the relationship between WM, multimedia annotations and reading comprehension. Therefore, the paucity in literature points to a need to further investigate the mediating role of WM in influencing the effects of annotation use on reading comprehension.

As a matter of fact, some relevant studies concerning the multimedia model of learning have touched on the issue of WM. The study by Plass et al. (2003), mentioned in

Section 2.2.3, found that the reading performance did not vary under no-annotation condition and verbal plus visual annotation condition but was worst under visual annotation condition. They postulated that the cognitive load induced by the visual annotations could help explain the results. However, the study was not followed up by looking at the WMC of the participants. WM was considered in Lee and Tedder's (2003) study, which explored the effects of different computer text formats on reading comprehension. Three text formats were involved: a traditional text, a structured hypertext (consisting of each sub-topic on a different screen, linking hierarchically by highlighted key terms in the text)，and a networked hypertext (containing a link of a term to an elaborating excerpt). Ninety-six undergraduates were randomly assigned to one of three groups, corresponding to the three text formats. Reading comprehension was assessed through 16 factual recall questions. The participants' WM was measured through a modified version of the Daneman-Carpenter RST three times, classifying them as being HWM, MWM (medium WM) and LWM learners. Results showed that higher WM participants received higher recall scores than the lower WM participants, but their recall scores did not vary by the text format; lower WM participants received the best recall scores when reading the traditional text format. The findings suggest that the structure and presentation of text influence how well information is recalled based on WM. They are, however, inconclusive due to several limitations inherent in the study. First, since the WM span was divided into three groups and compared in three different conditions, there were not enough participants to compare in certain categories. Another limitation concerned the reliability of the recall test, for the Cronbach score (0.61) was relatively low.

Similar to this study, Fontanini and Tomitch (2009) examined the relationship between WM and L2 reading comprehension under different text formats. Two groups of

university ESL students, one consisting of 21 Brazilians, and the other 21 Chinese, read

two texts of similar contents. The two texts were designed to be presented as a linear text

and as a hypertext. In the first format, the texts comprised an introductory paragraph and

some related subtitles; in the second format, the texts comprised the introductory paragraph

located in the first page of a computer screen, and nodes corresponding to the same

subtitles. The two groups and the two texts were arranged in a counterbalancing way.

Three different instruments were used to measure comprehension: free recall,

multiple-choice questions, and detection of contradictory information. A modified version

of the Daneman-Carpenter RST was used as a measure of WM. It was found that in

relation to the comprehension, no significant correlation existed between WM scores and

the linear text, and a moderate positive correlation between WM scores and the hypertext.

These results seem to suggest that reading performance in the hypertexts was related to the

amount of WM resources for processing information. This finding is consistent with

previous findings in the literature which show that the more complex the task, the more

WM resources are needed for performing it (Tomitch, 1999). The two studies reviewed

above indicate that individual differences in WM interact with the different computer texts.

Although they focused on text format, they suggest that WM is a crucial factor for CALL

reading, and thus has important implications for the future research on multimedia

annotations to consider the role of WM. In this, they will extend the investigation of the

effect of multimedia annotations on reading comprehension to the content area of

educational psychology.

2.5 Summary

CALL reading is different from paper reading in that information can be presented in

the form of text, picture, graphic, audio, and video. Readers have the flexibility to choose their preferred presentation modes and the freedom to control their reading process. The rationale for using annotations is that they are supposed to facilitate the reading process and enhance reading comprehension. Readers generally have positive attitudes towards annotations. Although annotating texts may be useful in helping learners comprehend a text, it may also make text processing more complicated. Previous studies have yielded insufficient, inconsistent and inconclusive results about the relationship between annotation use and reading comprehension. The mixed results may be due to many factors, such as the different theoretical foundations of the studies, sample sizes, types of annotations used, annotation look-up requirements, and measures of reading comprehension. Further, most previous studies applied annotations only to words (e.g., Abuseileek, 2008, 2011; Bowles, 2004; Chun & Plass, 1996a; Cooledge, 2004; Davis & Lyman-Hager, 1997; Lomicka, 1998; Marzban, 2011; Plass et al., 1998; Plass et al., 2003; Yanguas, 2009), and only a few studies by Ercetin and his colleagues involved topic annotations (Ariew & Ercetin, 2004; Ercetin, 2003; Sakar & Ercetin, 2005). The word-level annotations help the bottom-up processing, but hardly assist the top-down processing. Following an interactive model of reading, more studies that provide both word annotations and topic annotations are likely to generate a more comprehensive and deeper understanding of the effects of annotations on reading comprehension.

WM is a limited capacity system, and individuals differ in the WMC that is available for complex cognitive tasks. Previous studies have shown that WM is a major source of the differences in reading performance in a traditional reading context. In a multimedia reading environment, where reading materials can be presented nonlinearly, WM has also been shown to be a crucial factor for reading performance. With regard to the discrepancy of

previous research findings about the effectiveness of annotation use for reading

comprehension, it has been postulated that different multimedia modes of annotations

might have caused cognitive load and resulted in negative impact on comprehension.

However, there is little research that empirically probes this issue.

Accordingly, the previous inconclusive findings of the effect of annotation use on

reading comprehension and a research gap in the literature on the mediating role of WM in

influencing annotation use provide a rationale for the present study. Following an

interactive model of reading, the present study develops annotations at both word and topic

levels; following the cognitive theory of multimedia learning, annotations are designed in

the verbal, visual and verbal plus visual forms; WM is determined and categorized for an

understanding of its role in the interaction between annotation use and reading

performance. The methodology of this study is described in the next chapter.

3 METHODOLOGY

This chapter presents a description of the design and methods implemented in this study. It first states the research questions to be investigated, then describes the participants employed, reading texts selected, instruments used to collect data, and finally reports the results of a pilot study.

3.1 Research questions

Based on the review of the relevant literature in the previous chapter, this study seeks to answer the following four questions:

(a) Does annotation use make a difference to reading comprehension?

(b) Does annotation type yield differential effects on reading comprehension?

(c) Does annotation use make a difference to reading comprehension based on learners' working memory?

(d) Does annotation type yield differential effects on reading comprehension based on learners' working memory?

An experiment was conducted to empirically address these questions. In the experiment, annotations were provided for the reading texts in three types, that is, verbal annotation (VBA), visual annotation (VSA), and verbal plus visual annotation (VVA). The participants were identified to be high working memory (HWM), medium WM (MWM) and low WM (LWM) learners. Reading comprehension was quantified as scores on a reading comprehension test. The first question is addressed by comparing the reading comprehension scores following reading with and without annotations, so as to reveal whether each of the three annotation types leads to improved reading performance. It is hypothesized that reading with annotations will lead to better reading outcomes, and the

result is expected to corroborate the findings by Abuseileek (2011), Lomicka (1998), Yanguas (2009), and Bowels (2004).The second question is addressed by comparing the reading comprehension scores associated with the three annotation types, so as to find out whether one annotation type is more beneficial than another to reading comprehension. It is hypothesized that VVA will be superior to VBA or VSA in enhancing reading comprehension, as the combination treatment enables an inferential link to be built between verbal and visual representations (Mayer, 2001). The result is expected to be consistent with the findings by Akbulut (2008), and Chun and Plass (1996a). To address the third question, the study compares comprehension scores between reading with and without annotations for HWM, MWM and LWM learners, so as to unveil whether each of the three annotated reading conditions leads to better reading outcome than the non-annotated reading condition based on WM. It is hypothesized that all the annotated reading conditions will yield better reading comprehension regardless of WM capacity (WMC), based on the positive role of annotations in facilitating reading process, as reported by the studies mentioned above. To address the fourth question, the study compares the reading scores of the HWM, MWM and LWM learners when exposed to the three different annotation types, so as to ascertain whether one annotation type is more effective for reading comprehension than another according to WM. It is hypothesized that no one annotation type will be superior to another in improving reading comprehension under the constraints of WMC, as some studies (e.g., Fontanini & Tomitch, 2009; Lee & Tedder, 2003) have suggested that WM is a powerful predictor of CALL reading comprehension.

3.2 Participants

This study was carried out in a Chinese university, which ranks 36th among more than one thousand universities in China. In a typical Chinese university, usually the course "College English" is offered to non-English majors at 6 levels. The target population for this study comprised Level-5 students. The original pool included a total of 259 students from six classes. With the consent of their teachers, the researcher went to each of the classes to give a brief presentation to the students, explaining the purpose of this study, the tasks involved, the expected time length of each task, and asked for their voluntary participation. The students were assured that all their information would be kept confidential and their performance on any of the tasks would not affect their grade in this course. The researcher promised to give each participant a special bookmark as an appreciation of their participation. Those who volunteered to participate in the study filled out a consent form, indicating their willingness. To identify participants for this study, a WM test consisting of three WM measures (detailed in Section 3.4.1) were first administered to the students. The data of 162 students on the WM test were found to be complete and valid. These students (93 males and 69 females) came from the Nursing School, the Chemical Engineering Department, the School of Journalism and Communication, the Business School, the Medical School, and the School of Information Engineering at the university. Their ages ranged from 19 to 24 ($M = 21.26$, $SD = 1.43$), and on average they had been learning English for 7.70 years.

3.3 Reading texts and software

Reading texts

The reading texts used for this study were chosen from IELTS (academic module)

past papers. This choice was made out of several considerations. In the first place, the IELTS test is one of the most authoritative tests for ESL learners. As is stated in the official website of IELTS (www.ielts.org), it has been designed by world leading experts in language assessment, and its production involves an extensive and scrupulous process. The highest qualitative standards as well as quantitative statistical data are used to ensure that test materials are suitable. In the second place, the academic reading section contains authentic texts on academic topics of general interest. They are written for a non-specialist audience, and therefore do not discriminate between students of different academic backgrounds. Third, the test is claimed to be appropriate to candidates entering undergraduate or postgraduate courses (www.ielts.org). In view of the proficiency level of the participants, it is assumed to be a suitable source for the reading materials of the present study.

More important, the reading test in IELTS has good validity and reliability. In the first place, a wide range of reading skills is assessed by asking both literal and inferential questions (Leslie & Caldwell, 2009; Sweller, 1994), including reading for main ideas and details, understanding inferences and implied meaning, recognizing a writer's opinions, attitudes and purpose, and following the development of an argument (www.ielts.org.com). Alptekin and Ercetin (2009) point out that reading comprehension itself, as a dependent variable, should be operationalised and measured in terms of its principal components of literal and inferential reading (p. 635). In the second place, a variety of question types, or question formats (Tal, Siegel, & Maraun, 1994) is used, including multiple choice, information identification or matching, information completion, and short-answer questions. It is argued that different question types seem to measure different aspects of reading comprehension (Kobayashi, 1995, 2002). Therefore, the choice of IELTS reading

texts with different question types makes it possible to assess reading comprehension more accurately. In addition, the test materials are designed carefully so that every version of the test is of a comparable level of difficulty (www.ielts.org). This is especially helpful to the present study, as it used reading texts of equal difficulty from different versions of the IELTS test.

The study involved four annotation conditions: NA (no annotation), VBA, VSA, and VVA. The design of the study required each participant to experience each of the four conditions, each with a different reading text, for it makes no sense for a participant to read the same text four times. The four annotation conditions were counterbalanced across four reading groups, each of nearly even numbers of HWM, MWM and LWM learners assigned randomly, as shown in Table 3.1. The counterbalancing design requires the four texts to be equivalent in terms of difficulty, that is, no one is easier or harder than others. Some

Table 3.1

The Counterbalancing Design of Annotation Conditions

Reading Group	No. of participants	Text 1	Text 2	Text 3	Text 4
Group A	41	NA	VBA	VSA	VVA
Group B	41	VBA	NA	VVA	VSA
Group C	40	VSA	VVA	NA	VBA
Group D	40	VVA	VSA	VBA	NA

criteria were applied to the selection of appropriate texts: (a) they have similar readability, (b) they have a similar style, (c) they do not vary much in length, (d) they have more than one question type, (e) they have the same number of question items, and (f) their contents

are generally unfamiliar to the participants. Four texts were finally chosen from the *Examination Papers from University of Cambridge ESOL Examination* (Books 2-6, published by Cambridge University Press, 2009) (see Appendix 1). However, the selection of the four texts underwent a rigorous procedure.

To establish appropriate reading texts for this study, the researcher first chose 8 IELTS texts for a preliminary analysis. They were first scrutinized by the Range program with British National Corpus 14,000 (http://www.victoria.ac.nz/lals/about/staff/paul-nation), which is a computer software designed by Paul Nation and Alex Heatley (2002), Victoria University of Wellington, New Zealand, to analyze the vocabulary in a text and evaluate the vocabulary load of the text. The Range program is based on the 14 British National Corpus 1,000 word lists (Nation, 2004), which were created according to the frequency and range of occurrence of word families in the BNC. The Range Program produces a table showing the vocabulary data of a text, including the total words, the number and percentage of tokens, types and word families according to each 1000 word list. Each word in a text counts as one token and each different word in a text counts as one type. It is generally accepted that the type and token ratio are good indicators of the complexity of language in corpora (Nation & Heatley, 2002). The vocabulary loads of the 8 texts were computed by using the data of the first six 1000 word lists. This range was determined by reference to the estimated vocabulary size of the participants (According to the College English Teaching Syllabus, one indicator of Level 5 is the vocabulary size, which is about 5600 words). Based on the reports by the program, two texts were excluded because they were located on the two extreme ends across the 8 texts in terms of token, type and family. The remaining six texts were administered as a paper-and-pen reading comprehension test

to a group of 30 Level-5 students, who had also participated in piloting of the WM measures (detailed in Section 3.5). They did not participate in the main study. After taking the reading comprehension test, the students were asked to respond to a survey on their familiarity with the contents of the six texts before reading them. The survey used a Likert 5-point scale, where 1 to 5 respectively stands for *Not at all familiar*, *Slightly familiar*, *Moderately familiar*, *Very familiar*, and *Extremely familiar*.

Table 3.2

Descriptive Statistics for the Reading Texts Selected from IELTS Test Papers

Text	Total	Range statistics			Mean score	Topic familiarity	
		Tokens	Types	Families		*M*	*SD*
Text 1	836	778/93%	396/47%	340/41%	6.86	2.00	1.08
Text 2	831	735/88%	337/41%	275/33%	6.64	2.1	1.12
Text 3	840	749/89%	340/40%	294/35%	6.29	2.05	1.00
Text 4	788	693/88%	322/41%	279/35%	6.96	1.95	1.10

Table 3.3

ANOVA Results for Paper-and-Pen Reading Comprehension Test

	SS	*df*	*MS*	*F*	*Sig.*
Between groups	7.53	3	2.51	.86	.47
Within groups	316.54	111	2.93		

Twenty-eight students submitted complete and valid data, which helped further remove another two texts, and four texts were finalized for the main study. Table 3.2 shows the Range statistics, mean reading scores and survey results for the four reading texts. It

can be seen that the mean scores were close to each other, and generally the students reported low familiarity with their contents. The ANOVA results, displayed in Table 3.3, suggested that there was no significant difference in the test scores of the four texts, F (3, 111) = .86, p = .47 > .05.

The students, who participated in the paper-and-pen reading comprehension test, were also instructed to underline all the difficult words in the texts while reading. These words were collected and listed in descending order of marking frequency, and this was used as a basis for selection of words to be annotated. Selection also required that all words could be annotated both verbally and visually. Where a word was exceedingly difficult to describe visually without introducing potential confusions (e.g. *irresistible*), it was excluded for the sake of balance in the different annotation types. In addition, if a comprehension question was about the meaning of a word, this word was not annotated. The number of word annotations for each text was kept equal. The verbal annotation (in L2) of a word is its dictionary definition (in most cases, the *Cambridge Advanced Learner's Dictionary* was consulted). However, if a word has more than one definition in a dictionary, or the definition is too general, only its contextual meaning in the text was provided; if a word is part of a collocation, then this collocation was defined as a whole. A verbal topic annotation was developed for each text to provide background information about the text. Visual annotations for words and topics were all in the form of pictures, selected from the Internet. The annotations finalized for this study are provided in Appendix 2.

Reading software

The reading software for this study was designed by the researcher and developed by a software engineer. The four texts were digitized in the software in the form of a webpage. Each text had four versions corresponding to the four annotation conditions. Each version

itself had a file package, containing Data, JavaScript, Images, Style and a Webpage index icon (linking to the webpage of the text). Each page had a green page navigation bar at the top middle, which showed in black the current page that one was reading. The topic annotation of a text was embedded on the upper-left corner of the first page, indicated by "Background Information", while words with invisible annotations were highlighted in red in the text. When they were clicked, a small separate window popped up on the screen, showing the annotations. In a combination treatment condition, both the verbal and visual annotations appeared in the same window simultaneously. As to the position of the pop-up window, some researchers argue that it should not block the text section where the annotated word is embedded, in order not to interfere with the reading process (Chen, 2006). However, a pop-up window outside the text will cover part of the screen, so that less space is available for the text. To solve this problem, the pop-up windows in this study were designed to be movable --- they could be dragged to any place on the screen. The participants were instructed to click all the annotations available in a text, and this special requirement was followed up through a technical design which prevented the participants from going to the next page unless they had clicked all the annotations on the current page.

3.4 Instruments and data collection

The instruments used to collect quantitative data to address the research questions included a WM test and a reading comprehension test. In addition, a questionnaire survey and an interview survey were undertaken to obtain additional information for looking at the research questions and assist in interpreting the research findings through triangulation.

3.4.1 Working memory test

As reviewed in the previous chapter, many instruments for the measurement of WM have been documented in the literature. In this study, the tasks used to measure WM consisted of: an operation span task (OST), an L1 reading span task (RST), and an L2 RST (see Appendix 3). The choice of the three tasks was made based on the following considerations. First, as no single instrument is a perfect measure of WM (Conway et al., 2005), multiple instruments that do not replicate each other can be utilized to produce a more accurate measure of WMC, as recommended by Conway et al. (2005) in a methodological review of WM measurement. Second, different WM span tasks are argued to be suitable for different cognitive behaviors (Guan, 2007; Waters & Caplan, 1996). Since this study concerns reading comprehension, a RST can best serve the purpose, as contended by Guan (2007) that a RST involves sentence comprehension from the lower level process that decodes individual words, to the higher level process that computes the semantic, syntactic, and referential relations among the successive words. In view of the debate over the issue of WMC being language-general or language-specific, it was decided that both an L1 RST and an L2 RST were employed. At least, an L1 RST can avoid confounding with L2 proficiency. Third, OST is also a good predictor of comprehension (Daneman & Merikle, 1996), and it also lessens the demand on language proficiency. All three measures are similar in that they require subjects to switch between a processing component and a memory component, and are thus appropriate to the L2 reading comprehension research.

The L2 RST was obtained from Professor Meredyth Daneman of the Department of Psychology at the University of Toronto, Canada, in November, 2009. This is her recent version of the WM span task, and it has been documented in other studies (e.g., Daneman

& Hannon, 2001; Hannon & Daneman, 2001). The RST consists of 60 unrelated sentences (each having 10 - 14 words), half making sense, and the other half not. The L1 RST was taken from Li's (2005) doctorial thesis. This RST, following the Daneman-Carpenter RST, comprises 60 unrelated sentences (each having 12-14 words), half being correct, and the other half incorrect. Each sentence is followed by a high-frequency two-character noun phrase, which was randomly chosen from the *Chinese Frequency Dictionary*. The researcher generated the OST, following Turner and Engle (1989). It consists of 60 mathematical equations, each followed by a to-be-remembered English word. There are two arithmetic operations on one side of the equation: a simple multiplication or division operation, followed by a simple addition or subtraction operation, and a solution on the other side. Half of the solutions are correct and half incorrect. All the to-be-remembered words, 4 to 6 letters in length, are high-frequency nouns taken from Rebecca Sitton's Word Bank of 1200 High-Frequency Writing Words (http://www.suu.edu/faculty/lundd/ readingsite/ readingresources/SittonList.htm). In all three measures, the 60 sentences/equations were put into three groups evenly, and in each group the 20 sentences/ equations were arranged in 5 sets, from a 2-sentence/equation set up to a 6-sentence/equation set.

Following Engle, Cantor, and Carullo (1992), the set presentation order in the three tasks were randomized rather than in ascending order (items with fewer elements first), which prevented the participants from anticipating the number of to-be-remembered words. It has been argued that the randomization can effectively eliminate dependence on any tactics that come from knowing the size of memory set in advance, known as proactive interference in WM span tasks (Lusting, May, & Hasher, 2001; May, Hasher, & Kane, 1999). However, to avoid the risk that early presence of difficult items may discourage

some participants, all the tasks started with a 2-sentence/equation set. In order to maximize operational consistency and minimize operational difference, the sentences/equations were displayed automatically, with an interval of 6 seconds for the L2 RST and 5 seconds for the L1 RST and the OST. Researchers have cautioned that a longer interval between presentations may permit rehearsal of the last word, thus making the task more a measure of retention than of WMC (Conway et al., 2005; D. Friedman & Miyake, 2004). When a sentence/equation in a certain set appeared on the computer screen individually, the participants were required to read aloud the sentence/equation, decide its sensibility/ correctness, and try to remember the last word. They were told to try to perform the sentence/equation task accurately and to recall the last words in a set in order or, at least, not to write the final word first. At the end of each trial, the participants were given some time to write down the words.

The WM test was conducted in the computer labs of the Foreign Language Audio-Visual Center of the university. Each of the tasks took about 20 minutes. Before starting the tasks, the participants were given an opportunity for some practice. In marking the test, each correct recall was assigned one point, so the score range of each task was 0-60. Since the tasks were different, the raw scores in each task were converted into *z*-scores and their average was taken to be the WM scores of the participants. Those students who scored below 85% (Turner & Engle, 1989; Unsworth et al., 2005) in sensibility / correctness judgment were excluded from data analysis, for they might be focusing only on remembering while ignoring processing, or not taking the test seriously. Of the 251 students tested, 89 were excluded due to attrition, non-completion of all tasks, or failure to meet task criteria. Finally, the data of 162 students were used. A 4-percentiling method was employed to identify them to be high, medium and low WMC learners. The

results are reported in Section 4.1 of the next chapter.

3.4.2 Reading comprehension test

To administer the reading comprehension test, the four reading groups were assigned to each of four computer labs, where the reading software was uploaded to the computers according to the counterbalancing design of annotation conditions. That is, all the groups were exposed to the four texts, but each text varied in its annotation condition in different groups. For example, Group A read Text 1 with NA, Text 2 with VBA, Text 3 with VSA, and Text 4 with VVA, while Group B read Text 1 with VA, Text 2 with NA, text 3 with VVA, and Text 4 with VSA. The reading texts were available to the participants while answering questions, as text availability can cater to the search for or confirmation of answers. Previous research has suggested that text unavailability requires a significant dependence on the reader's memory skills (Andreassen & Braten, 2007), and assessments that do not allow look-backs may actually underestimate a student's level of comprehension (Leslie & Caldwell, 2009).

The reading comprehension test took place in one session of 100 minutes. Before the test, instructions were given to the participants on how to use the software. The four texts had the same number of question items (n=13). The scoring followed the guideline as set down in the IELTS reading test scoring rubric: each correct answer receives 1 point. Since the four texts were established to be equivalent, reading scores on each text were collected by annotation condition.

3.4.3 CALL reading questionnaire

The purpose of this questionnaire survey was to generate quantitative or numerical

description of how the participants look at annotations. The questionnaire was constructed

by the researcher with reference to those used in other studies (e.g., Ercetin, 2003; Hwang

et al., 2007; Sakar & Ercetin, 2005). In addition, two CALL teachers were asked to review

the questionnaire to ensure its construct validity. It consists of two parts (see Appendix 4).

The first part elicits the demographic data of the participants, including their ID number,

gender, age, major, and years of learning English. The second part seeks the participants'

views on CALL reading and their perceptions of different types of annotations, using

Likert 5-point format (*Strongly Disagree, Disagree, Undecided, Agree, Strongly Agree*).

There are 30 items in this part: 12 concerning their attitudes towards reading with

annotations, opinions on the necessity and role of annotations, and strategies in using

annotations, 6 concerning their perceptions of each of the three annotation types. The

questionnaire was written in Chinese to ensure that it could be clearly and fully understood

by the participants. It was administered immediately after the reading comprehension test.

In collecting the data, items asked in a negative way (indicated by an asterisk) were reverse

scored.

3.4.4 Interview

The goal of the interview was to obtain in-depth data on CALL reading. The themes

explored in the interviews included reading strategies, perceptions of annotations, and

views on the design of the reading software, based on which 7 semi-structured questions

and one open-ended question were developed (see Appendix 5). Additional questions were

brought up during the interview as a result of what the interviewees said. Sometimes,

follow-up questions were asked for clarification or specific information. Five HWM, five

MWM, and five LWM participants were randomly chosen to take part in a 15-minute

interview individually. The interviews took place in a computer lab where the reading texts were available, so that the interviewees could refer to the texts in answering questions, if necessary. The interviews were conducted in Chinese to ensure that the interviewees could clearly and accurately express their ideas. The researcher transcribed and translated the recorded interview data first, and then asked a teaching assistant (an English-major postgraduate) in the School of Foreign Languages at the university to cross check the translation. After that, the researcher and the teaching assistant independently coded the translated data and identified categories and corresponding themes from the data. They discussed points of disagreement until consensus was reached.

3.5 Pilot Study

The pilot study was conducted as a precursor to the main study to establish the validity and reliability of the instruments used to elicit data, and gain some insight into the research design of the study. After obtaining permission to conduct the research from the University of Auckland's Human Participants Ethics Committee (Ref. No. 2009/422), a series of tests were carried out to (a) pilot the WM measures, (b) trial the reading comprehension test, (c) obtain feedbacks on the design of the software and the annotations, and (d) evaluate the questionnaire and interview questions.

Working memory test

The original three WM measures were trialed with 30 Level-5 students, who voluntarily participated in the WM test. Twenty-one of them submitted valid answers. The descriptive statistical data for the three WM measures are shown in Table 3.4. A Pearson correlation analysis was performed, which revealed that they were not significantly correlated (see Table 3.5), for all the p values were above .05. However, these findings were not of

significance, likely due to the small sample size.

Table 3.4

Descriptive Statistics for WM Test in the Pilot Study

MW measure	N	MIN	MAX	M	SD
L1 RST	21	24.00	54.00	42.19	7.53
L2 RST	21	19.00	57.00	38.62	10.95
OST	21	29.00	52.00	42.10	7.60

Table 3.5

Pearson Correlations of the WM Measures in the Pilot Study

Pearson correlation	L1 RST	L2 RST
OST	.254 .267	.363 .106
L1 RST		.429 .053

The pilot test helped in troubleshooting and revising the design of the WM measures in several aspects. First, regarding the interval length between the sentences /equations presented automatically on slides, the literature has documented 5 seconds (Daneman & Carpenter, 1980), 6.5 seconds (van den Noort, Bosch, & Hugdahl, 2006), and 7 seconds (Chun & Payne, 2004; Strehler, 2008). The original L2 RST used 7 seconds, and L1 RST and OST used 6 seconds. As the means of the three measures were somewhat higher (L1 RST, $M = 42.19$; L2 RST, $M = 38.62$; OST, $M = 42.10$), one second was deducted from each of the three measures for the main study to improve their sensitivity. Second, the

instructions did not seem to be clearly stated or strictly followed. In the Answer Sheet for L1 WM test, the students were expected to write down two-character noun phrases, for one character does not make sense in the sentence. However, several students only wrote one word. One student wrote the English equivalents of the nouns. Several students failed to give sensibility / correctness judgment on all the sentences/equations in the tasks. Therefore, the instructions were reformulated to give more detailed explanation of the tasks, with important points capitalized for emphasis, and the final two-character noun in each sentence was put in quotation marks to indicate that it should be remembered as a whole. Third, based on the feedback on the design of the measures, the yellow words against the red background of the slides were replaced by black words against white background to alleviate eyestrain. Fourth, longer sentences which spread over two lines were reduced to one line by changing the margin or font size. This pilot study showed that the total time for completing the three WM test was approximately 60 minutes.

Reading comprehension test

Since the above-mentioned students had done the paper-and-pen reading comprehension test, a different group of students ($n = 24$) were recruited to voluntarily pilot the computer-version reading comprehension test. The pilot test revealed some points which were taken into consideration in administering the formal reading test. For example, the researcher found that two students looked up unknown words which were not annotated in the text through the Internet. This reminded the researcher that access should be denied to the Internet or dictionaries embedded in the computers when the participants were taking the test. Moreover, the reading test was conducted at 3:30 pm, immediately after the students had class between 1: 40 pm and 3: 10pm. Therefore, the students had been loaded with information from the class and might feel tired during the test. This

suggested that the actual reading comprehension test should preferably be scheduled at a time advantageous to the participants. Based on the feedback from interviews with three students on the design of the original reading software, two major revisions were made before it was used for the main study. First, two arrows indicating paper turning on the lower right corner of every page were changed to a blue page navigation bar located at the top middle, highlighting the current page in black. Second, when a picture was rather difficult to understand, an arrow was added pointing to a certain place in the picture to foreground its intended meaning.

Questionnaire and interview survey

The original questionnaire consisted of 31 items, including 30 statements and one open-ended question. A Cronbach's alpha internal consistency analysis based on complete data from 20 students revealed that the questionnaire had higher reliability ($a = .74$). Revisions were made according to suggestions from some students. First, the negative statements were marked with an asterisk as a caution in case that the students might respond to them as a positive statement. Second, some statements were reformulated to be more accurately expressed, like deleting the vague or fuzzy word "*quite*" in "*CALL reading is quite different from paper reading*". In addition, the open-ended question was removed because no student responded to it substantially. The original interview questions ($n = 7$) were emailed to 4 students, asking them to indicate whether these questions were clearly asked, and to provide suggestions that they might have on the coverage of the questions on the basis of their CALL reading experience. Their responses showed that they understood what each question was intended to elicit. Revisions made accordingly included adding "why" and/or "why not" to a general question, and one more question concerning the improvement of the reading software.

In summary, the pilot study was intended to validate the instruments and investigate the applicability of the research design to be used for this study. The results of the WM test showed that the three WM measures were not significantly correlated. Although the small sample size rendered the finding insignificant, they suggested that it might make more sense to use the three different measures, as a composite score can bring together measures that tap into slightly different aspects of WM. The questionnaire was proved to be a reliable instrument to obtain students' perceptions of CALL reading. Changes were made to the instruments as a result of this study to improve their validity and reliability. Finally, the pilot study provided useful references and experience to draw upon in the implementation of the actual study.

4   RESULTS

This chapter communicates the data analysis results and findings related to each research question. The first section reports the working memory (WM) test results, based on which the participants were assigned to different WM groups. The second section presents the results of inferential analyses of the data as relevant to the research questions. The third and fourth section describes the results of the questionnaire and interview surveys respectively. Basic interpretations of these results are given, and in-depth analyses of the results are provided in the next chapter.

4.1 WM test results and groupings

The descriptive statistical data of the three WM measures are given in Table 4.1. It can be seen from the table that the participants received similar mean scores on the operation span task (OST) and L1 reading span task (RST) (M = 40.08 and 37.90 respectively), but a relatively low mean score on the L2 RST (M = 26.72). The Skewness and Kurtosis values of each of the three tasks indicated a normal distribution of the scores, for none of the statistics were greater than 2.58 times their standard errors (Field, 2005, p. 139). Thus, a Pearson correlation analysis was conducted to examine the correlations between the three measures. The results, as displayed in Table 4.2, showed that they were each positively correlated, with the strongest correlation found between the L2 RST and the OST (p = .00, r = .45). This seems to suggest that the three WM measures tapped the same construct. This finding supports the claim by Conway et al. (2005) that OST and RST represent a similar construct and confirm the reliable relationships between L1 and L2 RST scores found in some previous research (e.g., Berquist, 1997; Chung, 2008; Juffs, 2004). In

order to compute a composite score for each participant, the z-score of each WM test was calculated. The descriptive statistics of the composite z-scores are recorded in Table 4.3.

Table 4.1

Descriptive Statistics for the WM Measures (n = 162)

WM measure	MIN	MAX	*M*	*SD*	Skewness		Kurtosis	
					Stats.	*SE*	Stats.	*SE*
OST	20.00	58.00	40.08	7.97	-.277	.191	-.427	.379
L1 RST	23.00	53.00	37.90	6.67	-.107	.191	-.580	.379
L2 RST	10.00	52.00	26.72	8.62	.293	.191	-.139	.379

Table 4.2

Pearson Correlations of the WM Measures (n = 162)

Correlation	L1 RST	L2 RST
OST Pearson Correlation	.17*	.45**
	.03	.00
L1 RST Pearson Correlation		.23**
		.003

Note. * $p<.05$, 2-tailed; ** $P< .01$, 2-tailed.

The composite z-scores were found to be normally distributed. Therefore, a 4-percentiling method was applied to identify WM groups based on the score distribution in descending order: the upper quartile of the distribution was labeled as high WM group ($n = 40$), the middle two quartiles as medium WM group ($n = 82$), and the lower quartile as low WM group ($n = 40$). The results for the groupings are presented in Table 4.4.

Table 4.3

Descriptive Statistics for WM Z-scores (n = 162)

WM measure	MIN	MAX	*M*	*SD*
OST	-2.52	2.25	.00	1.00
L1 RST	-2.24	2.27	.00	1.00
L2 RST	-1.94	2.93	.00	1.00
Composite	-6.18	6.30	.00	2.17

Table 4.4

Results for WM Groupings Based on Composite Z-score

WM group	*n*	Composite *z*-score
High WM	40	>1.31
Medium WM	82	1.31 – -1.27
Low WM	40	< -1.27

4.2 Results and findings by research question

This section reports the results and findings for each of the research questions based on the data from the reading comprehension test. The equivalence of the four reading texts had been established before they were employed for this study. Since the four texts were counterbalanced in the actual study, no one text associated with any one annotation condition, no significant difference in mean scores was assumed to be found across the four texts, and this assumption was checked statistically.

Table 4.5

Descriptive Statistics for the Reading Comprehension Scores by Text (n = 162)

Text	MIN	MAX	*M*	*SD*	95% CI	
					LB	*UB*
Text 1	3	12	7.77	2.02	7.46	8.08
Text 2	2	12	7.77	2.05	7.45	8.09
Text 3	2	12	7.64	2.25	7.29	7.99
Text 4	2	12	7.90	2.04	7.58	8.21

Note. *CI = confidence interval; LB = lower bound; UB = upper bound.*

Table 4.6

ANOVA Results for the Reading Comprehension Scores by Text (n = 162)

	SS	*df*	*MS*	*F*	*Sig.*
Between groups	5.19	3	1.73	.40	.76
Within groups	2809.55	647	4.36		

The descriptive data of the reading comprehension test scores by text are presented in Table 4.5. The mean scores of the four texts showed only slight variations, and their standard deviations were also close to each other. The results of a repeated measures ANOVA, as displayed in Table 4.6, indicated that there was no statistically significant difference among them, $F(3, 647) = .40$, $p = .76$. Thus, the assumption was confirmed. This suggests that the results were not confounded by the variable text; that is, no one text was found to be harder or easier than another. The means of the reading comprehension scores by text is shown graphically in Figure 4.1.
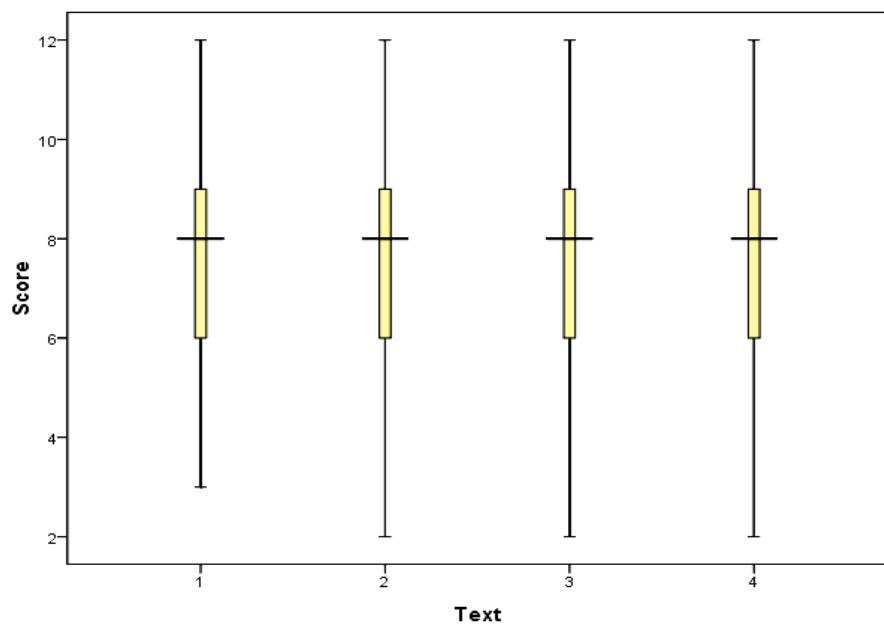
Figure 4.1 *Mean Reading Comprehension Scores by Text*

4.2.1 Research Question 1

The first research question explores whether annotation use makes a difference to reading comprehension. To address this question, the test scores were collected by annotation condition. A series of tests were carried out to statistically reveal whether there were any differences in the mean scores between no annotation (NA) condition and verbal annotation (VBA) condition, visual annotation (VSA) condition, and verbal plus visual annotation (VVA) condition. The descriptive statistics for the reading comprehension scores under different annotation conditions are given in Table 4.7. [1]

As reported in the table, the mean under each of the annotated reading conditions was higher than that under the non-annotated reading condition, with the largest difference

---

[1] The methods described in the following sections for analyzing the data were selected based on the nature of the research questions and the data. An alternative would be to use a factorial analysis (with all the variables included in that analysis) to address the research questions. The researcher tried the method and found that the results thus generated could lead to the same findings to the research questions as those reported in the thesis.

(8.49-6.65) occurring between VVA and NA. This seems to suggest that learners are likely

to score higher on texts with annotations than without annotations. That fact that the

standard deviation for NA (2.38) was the highest showed that there was more individual

variability in reading performance for this group than for the other groups. The values of

Skewness, Kurtosis and Standard errors of NA, VSA and VVA indicated that the data were

all normally distributed. However, this was not the case for VBA. In light of these results,

a paired samples *t*-test was run for NA vs. VSA and NA vs. VVA, while a non-parametric

test was applied to NA vs. VBA.

Table 4.7

Descriptive Statistics for the Reading Comprehension Scores by Annotation Condition

(n = 162)

Annotation condition	MIN	MAX	*M*	*SD*	Skewness		Kurtosis	
					Stats.	*SE*	Stats.	*SE*
NA	2	12	6.65	2.38	.08	.19	-.69	.38
VBA	4	12	7.88	1.99	.02	.19	-.99	.38
VSA	4	12	8.06	1.78	.05	.19	-.53	.38
VVA	4	12	8.49	1.69	.08	.19	-.34	.38

The paired samples *t*-test results, displayed in Table 4.8, showed statistically

significant differences between NA and VSA, *MD* = -1.41, *t(*161*)* = -6.68, *p* = .00, *d* = .68,

and between NA and VVA, *MD* = -1.84, *t*(161) = -9.17, *p* = .00, *d* = .89, with a 95%

confidence interval for the mean differences.

Table 4.8

Paired Samples T-test Results for NA vs. VSA and VVA

	MD	SD	t	df	Sig. (2-tailed)	95% CI	
						LB	UB
NA -VSA	-1.41	2.68	-6.68	161	.00	-1.82	-1.99
NA- VVA	-1.84	2.56	-9.17	161	.00	-2.24	-1.43

Note. *CI = confidence interval; LB = lower bound; UB = upper bound.*

A Wilcoxon signed rank test, a nonparametric equivalent of the *t*-test (Field, 2005, pp. 534-542) was conducted to compare the means of NA and VBA. The results, recorded in Table 4.9, showed that there was a statistically significant difference, $z = -5.49$, $p = .00$, with a medium effect size ($d = .56$).

Table 4.9

Wilcoxon Signed Rank Test Statistics Comparing VBA & NA

	MR	SR	z	Sig.
VBA –NA (*Negative rank*)	48.14	2022.00	-5.49	.00

Note. *MR= mean rank; SR = sum of ranks*.

In response to the first research questions, the results of both the parametric and non-parametric tests indicated that the participants when exposed to annotations performed better than when exposed to no annotations, which seems to suggest annotation use makes a difference to reading comprehension. Therefore, the hypothesis for this question is

confirmed. For Cohen's *d*, an effect size of 0.2 to 0.3 might be a small effect, around 0.5 a

medium effect, and 0.8 to infinity a large effect (Cohen, 1992). According to the criteria,

all the effect sizes were above medium. The difference is summarized graphically in Figure

4.2. It can be seen that while for all four conditions there were learners who scored at

higher levels, for the annotated conditions the lower ends dropped off compared to the

non-annotated condition, and the overall mean for the non-annotated condition was much

lower.
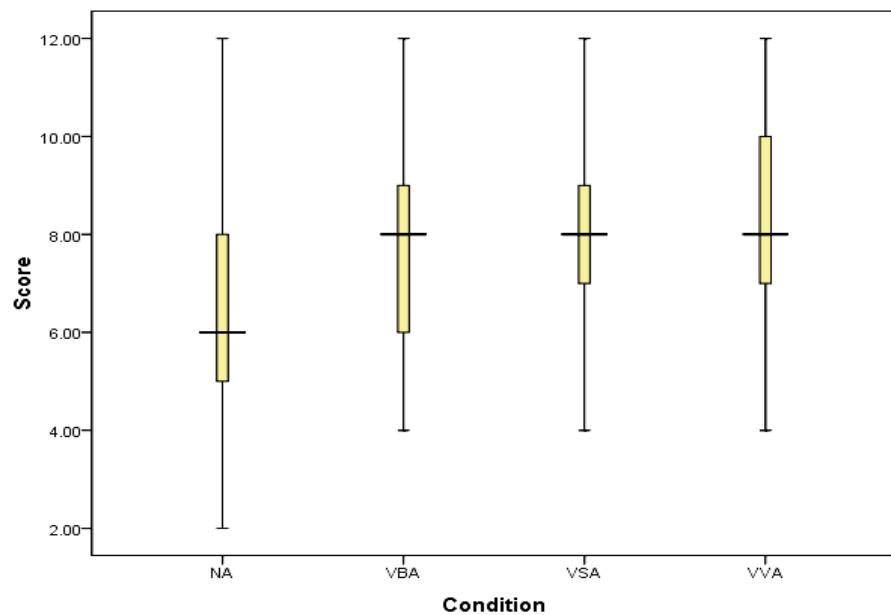
Figure 4.2 *Mean Reading Comprehension Scores by Annotation Condition*

4.2.2 Research Question 2

The second research question investigates whether the type of annotation yields

differential effects on reading comprehension. It focuses on annotated reading conditions

only. As mentioned in the previous section, the distribution of the reading scores across the

three annotation types did not all meet normality assumptions, and thus a Friedman test (a

non-parametric analog to repeated-measures ANOVA) was run to compare the mean

differences. The results, summarized in Table 4.10, revealed a significant difference

between the three annotation types, $X^2 = 17.11$, $df = 2$, $p = .00$. A post-hoc analysis with

Wilcoxon test (Bonferroni procedure was used to adjust the alpha level at .017) yielded the

results of multiple comparisons to help locate specific differences. As shown in Table 4.11,

no statistically significant difference existed between the means of VSA and VBA, $z = -.87$,

$p = .38$, $d = .10$, but a statistically significant difference occurred in the means between

VVA and VBA, $z = -3.1$, $p = .002$, $d = .33$, and between VVA and VSA, $z = -2.82$, p

$= .005$, $d = .25$.

Table 4.10

A Summary of Friedman Test Results for Annotation Types

Mean rank	X2	df	Sig.
VBA < VSA < VVA	17.11	2	.00

Table 4.11

Wilcoxon Test Results for the Reading Comprehension Scores by Annotation Type

	VSA - VBA	VVA - VBA	VVA - VSA
z	-.87[a]	-3.1[a]	-2.82[a]
p (2-tailed)	.38	.002	.005

Note. [a] = *based on negative ranks*.

The statistical results indicated that the type of annotations produced differential

effects on reading comprehension, depending on which type of annotation was introduced.

When the reading texts were provided with both verbal and visual annotations, the reading

performance appeared to be the best ($M = 8.49$); when annotated in a single manner, either verbally or visually ($M = 7.88$ and 8.06 respectively), the reading performance was significantly lower. These results suggest that the combination treatment of annotations is most effective for reading comprehension. Thus, the hypothesis for this question is confirmed.

4.2.3 Research Question 3

The third research question explores whether annotation use makes any difference to reading performance according to working memory capacity (WMC). It differs from the previous questions in that the participants' WMC is considered to see how WM mediates the effectiveness of annotation use for reading comprehension. Each WM group was examined individually. A series of paired samples t-tests were conducted to elicit statistical findings, all the data being normally distributed. Table 4.12 presents the statistics for the reading comprehension scores of the three WM groups by annotation condition, and Table 4.13 shows the paired samples t-tests results. The findings are reported according to WM group.

Low WM (LWM) Group

The descriptive statistics for the LWM group showed that the mean of NA ($M = 4.83$) was lower than that of VBA ($M = 7.20$), VSA ($M = 7.48$), and VVA ($M = 8.03$). The differences suggest that LWM learners are likely to score higher on reading tests when receiving annotations. In particular, the remarkably low mean and relatively high standard deviation of NA indicated that LWM learners performed rather poorly on the reading test and showed more individual variability in test scores when they did not access annotations. The results of a paired samples t-test (alpha level at .017), as displayed in Table 4.13,

pointed to a statistically significant difference in the means between NA and VBA, *SD* = 2.52, *t(39)* = -5.96, *p* = .00, *d* = 1.13, or between NA and VSA, *SD* = 2.57, *t(39)*= -6.53, *p* = .00, *d* = 1.43, or between NA and VVA, *SD* = 2.62, *t(39)* = -5.96, *p* = .00, *d* = 1.76, with a 95% confidence interval for the mean differences. These results suggested that it was not any single annotation type, but all, that could cause differences in reading performance.

Table 4.12

Descriptive Statistics for the Reading Comprehension Scores of the Three WM Groups by Annotation Condition

WM group	ANN	MIN	MAX	*M*	*SD*	Skewness		Kurtosis	
						Stats.	*SE*	Stats.	*SE*
LWM (N=40)	NA	2.00	10.00	4.83	2.10	.54	.37	-.29	.73
	VBA	5.00	12.00	7.20	2.09	.59	.37	-.83	.73
	VSA	4.00	11.00	7.48	1.57	.15	.37	-.15	.73
	VVA	6.00	11.00	8.03	1.48	.31	.37	-.85	.73
MWM (N=82)	NA	3.00	11.00	6.50	1.87	.38	.276	-.33	.53
	VBA	4.00	12.00	7.76	1.91	.09	.27	-.75	.53
	VSA	4.00	12.00	8.07	1.80	.02	.27	-.45	.53
	VVA	4.00	12.00	8.30	1.72	-.06	.27	-.18	.53
HWM (N=40)	NA	4.00	12.00	8.78	1.82	-.54	.38	.22	.73
	VBA	5.00	12.00	8.83	1.74	-.61	.37	-.01	.73
	VSA	5.00	12.00	8.60	1.79	-.17	.37	-.65	.73
	VVA	7.00	12.00	9.35	1.56	.27	.37	-.89	.73

Note. *LWM = low WM; MWM = medium WM; HWM = high WM; ANN=annotation.*

The improvement of the reading scores yielded by VBA, VSA and VVA represented a substantially large effect size. As reported, the mean differences demonstrated that for the LWM learners VVA facilitated reading comprehension the most.

Table 4.13

Paired Samples T-test Results for No Annotation vs. Annotations by WM Group

WM group	Comp.	*MD*	*SD*	*t*	*df*	*Sig. (2-tailed)*	95% CI	
							LB	*UB*
LWM (N=40)	NA- VBA	-2.37	2.52	-5.96	39	.00	-3.18	-1.57
	NA -VSA	-2.65	2.57	-6.53	39	.00	-3.47	-1.83
	NA - VVA	-3.20	2.62	-7.72	39	.00	-4.04	-2.36
MWM (N=82)	NA- VBA	-1.26	2.53	-4.50	81	.00	-1.81	-1.70
	NA -VSA	-1.57	2.60	-5.49	81	.00	-2.14	-1.00
	NA - VVA	-1.80	2.49	-6.57	81	.00	-2.35	-1.26
HWM (N=40)	NA- VBA	-.05	2.67	-.12	39	.91	-1.90	1.08
	NA -VSA	.18	2.22	.50	39	.62	-1.53	1.18
	NA - VVA	-.58	1.93	-1.88	39	.17	-1.19	1.04

Note. *Comp = comparison; CI = confidence interval; LB = lower bound; UB = upper bound.*

Medium WM (MWM) Group

Table 4.12 shows that for the MWM group the mean of each of the annotation types (*M*: VBA = 7.76; VSA = 8.07; VVA = 8.30) was higher than that of NA (*M* = 6.50), similar to the results observed in the LWM group. As shown in Table 4.13, a paired samples *t*-test

(alpha level set at .017) revealed a significant difference in each comparison, with a 95% confidence interval: for NA vs. VBA, $t(81)$ = -4.50, p = .00, and d = .67; for NA vs. VSA, $t(82)$ = -5.49, p = .00, and d = .86; for NA vs. VVA, $t(81)$ = -6.57, p = .00, and d = 1.00. These findings suggest that all the three annotation types are likely to enhance reading comprehension, leading to varying improvements in reading scores. In addition, VVA was found to be most beneficial to reading comprehension, similar to the case for the LWM group. However, the effect size was smaller in each case, compared to its counterpart in the LWM group. This seems to indicate that the effect of annotations on reading comprehension makes a greater difference for learners with lower WMC than for learners with higher WMC.

High WM (HWM) group

It can be seen from Table 4.12 that the HWM group outperformed the other two WM groups in the reading test under each of the four reading conditions (M: NA = 8.78; VBA = 8.83; VSA = 8.60; VVA = 9.35). However, the mean differences within the group were in fact very small. This seems to imply that HWM learners perform consistently regardless of annotation condition, a finding different from that in either the LWM or MWM group. The paired samples t-test results (alpha level set at .017) showed no statistically significant difference in the means between NA and VBA, $t(39)$ = -.12, p = .91, d = .03, or between NA and VSA, $t(39)$ = -.05, p = .62, d = .10, or between NA and VVA, $t(39)$ = -1.88, p =.07, d = .30, with a 95% confidence interval for the mean differences. In particular, the effect sizes, as reported, were low or negligible. The apparent advantage of VVA for reading comprehension over VBA and VSA as found in the other two WM groups was not observed, although it was still associated with the highest mean score. All the results suggest that annotation use does not produce differential effects on reading comprehension

for HWM learners.

Table 4.14

A summary of the Mean Difference and Effect Size of Annotation Use for Each WM Group

	LWM (*n* = 40)		MWM (*n* = 82)		HWM (*n* = 40)	
	MD	*d*	*MD*	*d*	*MD*	*d*
NA - VBA	-2.37*	1.13	-1.26*	.67	-.05	.03
NA - VSA	-2.65*	1.43	-1.57*	.86	.18	.10
NA - VVA	-3.20*	1.76	-1.80*	1.00	-.58	.30

Note: * = *significant difference*
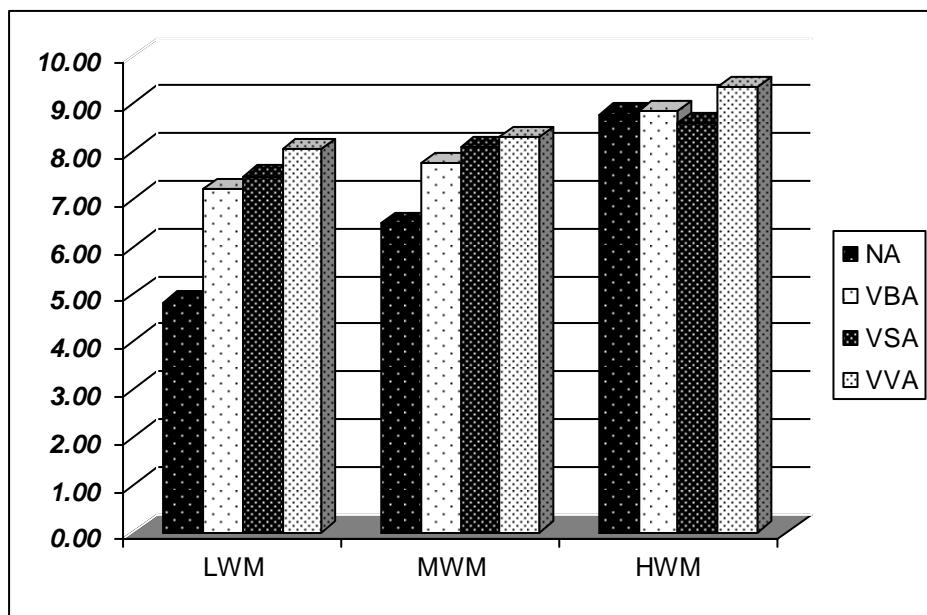
Figure 4.3 *Mean Reading Comprehension Scores by Annotation Condition for Each WM Group*

In response to the third research question, annotation use makes a difference to

reading comprehension for learners with low or medium WMC, but makes no difference

for learners with high WMC. The effect of annotations on reading comprehension seems to be greatest for LWM learners and smallest for HWM learners. The hypothesis for this question can only be partially accepted. The mean difference and the effect size of annotation use for each WM group are summarized in Table 4.14. The mean reading comprehension scores of each WM group under different annotation conditions are graphically shown in Figure 4.3.

4.2.4 Research Question 4

The last research question investigates whether the type of annotation yields differential effects on reading comprehension according to WMC. Focusing on annotated reading conditions only, it involves the following two sub-questions. (a) Is there any significant difference in reading comprehension for the three annotation types within each WM group? (b) Is there any significant difference in reading comprehension for each of the three annotation types across the three WM groups? The first sub-question was addressed through a within-group analysis, and the second through a between-group analysis. The statistical data for the analyses were based on Table 4.12, and the results are given below individually.

The first sub-question

To answer the first sub-question, a series of repeated measures ANOVA were run to statistically check whether there were significant differences in the mean between VSA, VSA, and VVA within each WM group. All the results are summarized in tables 4.15-4.16, and a detailed description is given for each WM group.

Table 4.15

ANOVA Results for Annotation Types within WM Groups

WM group	SS	df	MS	F	Sig.
LWM (n = 40)	14.11	2	7.06	2.35	.10
	351.35	117	3.00		
MWM (n = 82)	12.45	2	6.22	1.90	.15
	798.61	243	3.28		
HWM (n = 40)	11.85	2	5.93	2.05	.13
	338.48	117	2.89		

Table 4.16

Post Hoc Test Results for Annotation Types within WM Groups

WM group	Comparison	MD	SE	Sig.	95% CI	
					LB	UB
LWM (N=40)	VBA- VSA	-.28	.34	1.00	-1.12	.67
	VBA -VVA	-.83	.38	.11	-1.77	.12
	VSA - VVA	-.55	.29	.48	-1.49	.39
MWM (N=82)	VBA- VSA	-.32	.28	.79	-1.00	.37
	VBA -VVA	-.55	.26	.16	-1.23	.13
	VSA - VVA	-.23	.23	1.00	-.91	.45
HWM (N=40)	VBA- VSA	.23	.36	1.00	-.70	1.15
	VBA -VVA	-.53	.34	.51	-1.45	.70
	VSA - VVA	-.75	.32	.15	-1.67	.17

Note. *CI = confidence interval; LB = lower bound; UB = upper bound.*

LWM group

In this group, the mean score for VVA ($M = 8.03$) was higher than that for VSA ($M = 7.48$) and VBA ($M = 7.20$). It appears that LWM learners are more likely to receive higher reading scores when exposed to combined annotations. However, as shown in Table 4.15, an ANOVA with repeated measures revealed that the means for the three annotation types did not significantly differ from each other, $F(2, 117) = 2.35$, $p = .10$. Three analytical contrasts (alpha level set at .017) comparing VBA against VSA, VBA against VVA, and VSA against VVA found no significant difference in each contrast, either, as can seen in Table 4.16. These findings seem to suggest that LWM learners tend to perform similarly in reading comprehension tests, no matter whether the texts are annotated verbally, visually, or both verbally and visually, that is, the type of annotation does not appear to produce differential effects on the reading performance of LWM learners.

MWM group

In the MWM group, the mean reading scores of VBA, VSA and VVA varied from 7.76 to 8.30. A repeated measures ANOVA showed that the difference among them did not reach statistical significance, $F(2, 243) = 1.90$, $p = .15$. Multiple comparisons (alpha level set at .017) displayed three analytical contrasts for the different combinations of the annotation types. The results indicated that there was no significant difference in the means between each pair of annotation types. Therefore, it appears that the type of annotation, be it verbal, visual, or both, does not bring about differential effects on reading comprehension for learners with medium WMC.

HWM group

Similarly, the mean reading scores of VBA ($M = 8.83$), VSA ($M = 8.60$) and VVA ($M = 9.35$) in this group showed small variations, and an ANOVA with repeated measures

found that the differences were not statistically significant, $F(2, 117) = 2.05$, $p = .13$. The results of multiple comparisons (alpha level set at .017) pointed to no significant difference in the means between each pair of the annotation types. Therefore, the findings suggest that no annotation types can make a difference to the reading comprehension for learners with high WMC.

Together, the above statistical results showed that within each of the three WM groups there was no statistically significant difference in the mean reading scores associated with annotation type. This indicates that the type of annotation does not appear to render any differential effects on reading comprehension for learners with similar WMC. In view of the fact that the HWM group outperformed the MWM group, which scored higher than the LWM group, this finding seems to suggest that it is WMC rather than annotations that predicts reading comprehension.

The second sub-question

This sub-question addresses whether the type of annotation makes any difference to reading comprehension for learners with different WMC. Each of the annotation types was examined across the three WM groups. The ANOVA and Post Hoc test results are recorded in Table 4.17 and Table 4.18 respectively. Each annotation type is reported individually.

Verbal annotation

As is shown in Table 4.12, the means of the HWM group ($M = 8.82$) was the best, followed by the MWM group ($M = 7.76$), and then the LWM group ($M = 7.20$). This appears to suggest that higher WM learners are likely to score higher on reading comprehension tests when exposed to reading texts with verbal annotations. The ANOVA results, presented in Table 4.17, showed that for this type of annotation there was a statistically significant difference in means across the three WM groups, $F(2,159) = 7.56$, p

= .001. A post-hoc test yielded the results (see Table 4.18) of multiple comparisons between the three WM groups. A statistically significant difference was observed between the LWM group and the HWM group, MD = -1.62, p = .00, and between the MWM group and the HWM group, MD = -1.06, p = .01, but not between the LWM group and the MWM group, MD = -.56, p = .40, with a 95% confidence interval for the mean differences.

Table 4.17

ANOVA Results for Annotation Types between WM Groups

		SS	*df*	*MS*	*F*	*Sig.*
VBA	Between Groups	55.48	2	27.74	7.56	.001
	Within Groups	583.30	159	3.67		
VSA	Between Groups	25.36	2	12.68	4.16	.017
	Within Groups	485.14	159	3.05		
VVA	Between Groups	41.04	2	20.52	7.78	.001
	Within Groups	419.45	159	2.64		

Visual annotation

With respect to this type of annotation, the means of the HWM group (M = 8.60) was higher than that of the MWM group (M = 8.07), and much higher than of the LWM group (M = 7.48), a pattern similar to that observed for VBA. An ANOVA calculation revealed a significant difference in the means, $F(2, 159)$ = 4.16, p = .017. The results of a Post Hoc test with multiple comparisons between the three WM groups manifested no statistically significant difference between HWM group and MWM group, MD = -.53. p = .36, or between MWM and LWM groups, MD = -.59, p = .23, but a statistically significant

difference between the LWM and HWM group, $MD = -1.13$, $p = .01$, with a 95%

confidence interval for the mean difference. The findings are somewhat different from

those obtained with VBA. It appears that VSA is likely to produce differential effects on

reading comprehension only between extreme-WM learners.

Table 4.18

Post Hoc Test Results for Annotation Types between WM Groups

	Comparison	*MD*	*SE*	*Sig.*	95% CI	
					LB	*UB*
	LWM - MWM	-.56	.37	.40	-1.45	1.04
VBA	LWM - HWM	-1.62	.43	.00	-2.66	-1.59
	MWM - HWM	-1.06	.37	.01	-1.96	-1.08
	LWM - MWM	-.59	.33	.23	-1.41	1.22
VSA	LWM - HWM	-1.13	.39	.01	-2.07	-1.18
	MWM - HWM	-.53	.34	.36	-1.34	1.29
	LWM - MWM	-.28	.31	1.00	-1.04	1.48
VVA	LWM - HWM	-1.33	.36	.001	-2.20	-1.45
	MWM - HWM	-1.05	.31	.003	-1.80	-1.29

Note. *CI = confidence interval; LB = lower bound; UB = upper bound.*

Verbal plus visual annotation

For this type of annotation, a similar pattern of mean difference was observed to that

of VBA and VSA: the HWM group received the highest mean score ($M = 9.35$), followed

by MWM group ($M = 8.30$), and then by LWM group ($M = 8.02$). The ANOVA results

demonstrated a statistically significant difference in the means, $F(2, 159) = 7.78$, $p = .001$.

A pos-hoc test showed a significant difference between the HWM and MWN groups, *MD*

$= -1.05$, $p = .003$, and between the HWM and LWM groups, $MD = -1.33$, $p = .001$, but no

significant difference between the MWM and LWM groups, $MD = -.28$, $p = 1.00$, with a

95% confidence interval for the mean differences. Therefore, for this type of annotation the

LWM and MLW groups performed similarly, and they were outperformed by the HWM

group.

The statistical results reported above showed that when the three WM groups

accessed the same type of annotation, the HWM group always significantly outperformed

the LWM group on the reading comprehension test and that the LWM and MWM groups

performed similarly. This indicates that each of the annotation types could help the LWM

learners to improve their reading comprehension to some degree, but generally none could

compensate for the reading performance differences from the HWM group.

Table 4.19

A Summary of Mean Differences between WM Groups by Annotation Type

	MD (LWM – MWM)	MD (MWM – HWM)	MD (LWM – HWM)
VBA	-.56	-1.06*	-1.62*
VSA	-.59	-.53	-1.13*
VVA	-.28	-1.05*	-1.33*

Note: * = *significant difference, p < .05.*

The findings related to the fourth research question are summarized as follows. First,

the type of annotation did not yield differential reading outcome within the same WM

group. The hypothesis for this question is thus accepted. This finding seems to suggest that

WM is a powerful predictor of L2 CALL reading comprehension. Second, the higher WM groups always received higher mean scores than the lower WM groups whatever the annotation type was, as indicated by the negative mean differences summarized in Table 4.19.

As reported, the mean differences between the HWM group and LWM group always reached significance; in contrast, the mean differences between the LWM group and MWM group never reached significance, that is, the LWM group was performing like the MWM group when they both were exposed to annotations. In addition, an ANOVA test showed that, as displayed in Table 4.20, the three WM groups differed greatly on the reading comprehension test when they received no annotations, $F(2,159) = 42.88$, $p = .00$, with a significant difference between the LWM group and the MWM group, $MD = -1.68$, $p = .00$, and between the LWM group and the HWM group, $MD = -3.95$, $p = .00$, as indicated by a post-hoc analysis (see Table 4.21). Thus, given that the LWM group could match the MWM group with the aid of annotations but neither could match the HWM group, it seems that annotations could assist LWM learners to the point that they were able to overcome limitations of WM to a certain extent. The overall findings related to Research Question 4 are shown graphically in Figure 4.4.

Table 4.20

ANOVA Results for NA between WM Groups

		SS	*df*	*MS*	*F*	*Sig.*
NA	Between Groups	314.69	2	157.85	42.88	.00
	Within Groups	585.25	159	3.68		

Table 4.21

Post Hoc Test Results for NA between WM Groups

	Comparison	*MD*	*SE*	*Sig.*	95% CI	
					LB	*UB*
	LWM - MWM	-1.68	.37	.00	-2.57	-.78
NA	LWM - HWM	-3.95	.43	.00	-4.99	-2.91
	MWM - HWM	-2.28	.37	.00	-3.17	-1.38

Note. *CI = confidence interval*; *LB = lower bound*; *UB = upper bound*.
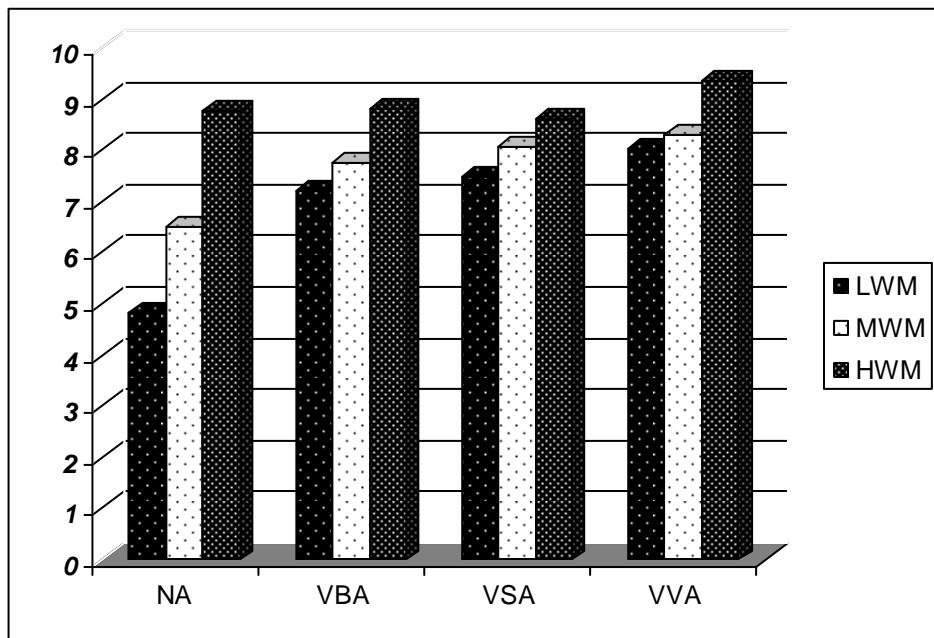
Figure 4.4 *Comparisons of Means across WM Groups by Annotation Condition*

4.3. Questionnaire Results

This section presents the results from the CALL reading questionnaire survey. It first describes the participants' general perceptions of annotations, then reports their perceptions of VBA, VSA and VVA respectively. When the results are reported in terms of percentage, they do not take the neutral choices (*Undecided*) into account.

4.3.1. General perceptions of annotation use

In the questionnaire, there were 12 items concerning the participants' general perceptions of annotation use. The statistical results based on these items are recorded in Table 4.22. It can be seen that the majority of the participants (85%) agreed or strongly agreed that multimedia reading was different from paper reading (Items 12 and 23: $M =$ 3.85, $SD = .62$). With respect to the effect of annotations on reading comprehension, most of them (72%) regarded annotations as helpful to reading comprehension (Items 1 & 27: $M = 3.64$, $SD = .68$). They gave generally positive views on the necessity of word annotations (Items 19: $M = 3.50$, $SD = .87$) and topic annotations (Items 2: $M = 3.17$, $SD = .56$). However, only 58% of the participants indicated a preference for reading with annotations positively (Items 6 and 26: $M = 3.44$, $SD = .71$) and 67% denied that annotations could help concentrate on reading (Items 4 and 8: $M = 2.86$, $SD = .75$). A plausible inference might be that their preference for annotations is conditional: when annotations helped solve problems in reading, they viewed them as being facilitative; otherwise, they considered them to be distracting. This is supported by their response to Item 11 ($M = 3.75$, $SD = .93$). In this study, they were required to click all the annotations, but presumably not all the annotations were necessary to all of them.

Among the 12 items, Item 15 was the only one where a Kruskal Wallis test (a non-parametric test analogous to one way ANOVA, performed based on ranked data) revealed a significant difference between the three WM groups, $X^2 = 10.38$, $df = 2$, $p = .006$. It appears that HWM learners ($M = 3.98$, $SD = .80$) are more inclined than MWM learners ($M = 3.39$, $SD = 1.03$) or LWM learners ($M = 3.45$, $SD = 1.04$) to depend on annotations.

4.3.2 Perceptions of different types of annotations

The statistical data on participants' perceptions of the three annotation types are presented in tables 4.23-4.25, and the results for each annotation type are described individually.

Table 4.22

Statistical Results for the General Perceptions of Annotations

Item	M	SD
12. Multimedia reading and paper reading make no difference. * 23. Multimedia reading is different from paper reading.	3.85	.62
6. I like reading a text with annotations more than one without. 26. I would rather read a non-annotated text than an annotated text. *	3.44	.71
19. The word annotations provided to the reading texts are necessary.	3.50	.87
2. The topic annotations in the reading texts are necessary.	3.17	.56
4. The use of annotations helps me to concentrate on reading. 8. I find annotations distract my attention from reading.*	2.86	.75
1. The annotations in the reading texts help reading comprehension. 27. The annotations in the reading texts don't facilitate the reading comprehension. *	3.64	.68
11. Usually I will turn to the annotations only when other methods fail.	3.75	.93
15. The dependence on the annotations will lead to a decline of my reading ability.	3.55	1.00

Verbal annotation (Table 4.23)

It was found that 61% of the participants liked reading texts with verbal annotations only (Item 13: $M = 3.52$, $SD = .97$), which indicates that generally they still favor the traditional annotation form. Overall, the participants confirmed the positive role of verbal

annotations for new words in reading comprehension (Item 7: $M = 3.72$, $SD = .95$), and

rated them as highly necessary (Item 22: $M = 4.25$, $SD = .86$). Regarding the design of

verbal annotations, more than 80% of the participants considered them to be intelligible

(Item 14 & 16: $M = 3.91$, $SD = .67$), indicating that the verbal annotations generally were

well developed, less likely to cause intrinsic cognitive load on WM. However, the mean

score of verbal annotations for text topics was relatively lower. One possible reason might

be that they tended to be long explanations, which were time-consuming when the primary

purpose of reading was comprehension. These findings were consistent across the three

WM groups, with no significant difference among them.

Table 4.23

Statistical Results for the Perceptions of Verbal Annotations

Item	*M*	*SD*
13. I like reading texts with verbal annotations only.	3.52	.97
7. The verbal annotations for new words assist the reading process.	3.72	.95
28. The verbal annotations for topics of the reading texts help in understanding the texts.	3.15	1.21
22. The verbal annotations in the reading texts are necessary.	4.25	.86
14. The verbal annotations in the reading texts are easy to understand. 16. The verbal annotations in the reading texts are difficult to understand.*	3.91	.67

Visual annotation (Table 4.24)

Compared to verbal annotations, a smaller proportion of participants (34%) expressed

their preference for visual annotations only (Item 20: $M = 2.99$, $SD = .99$). This suggests

that the participants showed a rather conservative attitude towards this type of annotations, as is evidenced by the data that about two-thirds denied the necessity of visual annotations for reading comprehension (Item 17: $M = 2.43$, $SD = .95$). However, 64% viewed the usefulness of visual annotations for words (Item 3: $M = 3.64$, $SD = .96$) positively, and 51% recognized the benefits of visual annotations for text topics (Item 9: $M = 3.31$, $SD = 1.25$). As regards the intelligibility of the pictures, only 54% the participants indicated that they could easily understand what the pictures alone were intended to mean (Items 24 and 29: $M = 3.50$, $SD = .84$).

Table 4.24

Statistical Results for the Perceptions of Visual Annotations

Item	M	SD
20. I like reading texts with visual annotations only.	2.99	.99
17. The visual annotations in the reading texts are necessary to reading comprehension.	2.43	.95
9. The visual annotations for topics are helpful to reading comprehension.	3.31	1.25
24. I do not know what the pictures alone are meant to explain. * 29. The pictures in the reading texts are easy to understand.	3.50	.84
3. The visual annotations for new words in the reading texts are useful for reading comprehension.	3.64	.96

Among the six items, Item 17 is the only one where a significant difference between the three WM groups was found from the results of a Kruskal Wallis test, $X^2 = 8.31$, $df = 2$, $p = .016$. The mean for the HWM learners ($M = 2.05$, $SD = .68$) was the lowest, suggesting that they are less likely to depend on visual annotations for reading

comprehension than MWM learners ($M = 2.55$, $SD = .97$), or LWM learners ($M = 2.55$, $SD = 1.06$).

Verbal plus visual annotation (Table 4.25)

It was found that 77% of the participants favored reading a text with verbal plus visual annotations (Item 10: $M = 3.85$, $SD = .92$), and 76% preferred them to be presented at the same time (Item 25: $M = 4.01$, $SD = .97$). However, only 9% indicated that they would read both types of annotation presented simultaneously even if either was unnecessary (Item 30: $M = 2.27$, $SD = .88$), against an opposite of 70%. This seems to evidence their antipathetic attitude towards unnecessary or redundant annotations. With

Table 4.25

Statistical Results for the Perceptions of Verbal plus Visual Annotations

Item	*M*	*SD*
10. I prefer reading a text both verbally and visually annotated than one annotated in a single way.	3.85	.92
25. I prefer both verbal and visual annotations to be presented at the same time.	4.01	.97
30. I will read both the verbal and visual annotations when they are presented at the same time even if either is unnecessary.	2.27	.88
5. The visual plus verbal annotations are helpful to reading comprehension. 21. The verbal plus visual annotations in the texts are not beneficial to reading comprehension. *	3.58	.71
18. The verbal plus visual annotations in the reading texts render negative effects on reading comprehension.*	3.93	.66

respect to the effect of this type of annotation, 65% agreed or strongly agreed that it could facilitate reading comprehension (Items 5 and 21: $M = 3.58$, $SD = .71$), and that a greater

percentage (82%) denied that it would affect reading comprehension negatively (Item 18: $M = 3.93$, $SD = .66$). For the six items, Kruskal Wallis tests did not find a significant difference among the three WM groups.

In summary, the questionnaire survey has revealed some substantive findings. Overall, the participants acknowledged the usefulness of annotations for reading comprehension. As to the types of annotations, verbal plus visual annotations were most favored, followed by verbal annotations only. Comparatively, they reported low preference for visual annotations though admitting their positive role in reading comprehension. Generally, they would click annotations only when necessary. The three WM groups only significantly differed in 2 out of the 30 items, suggesting that the students shared similar perceptions of annotations regardless of WMC.

4.4 Result of interviews

This section presents the results from the interviews conducted with 15 participants, 5 from each of the three WM groups. There are 7 questions concerning 3 topics: reading strategies; perceptions of annotations; views on the design of the annotations. The final question provided the interviewees with an opportunity to comment further on any aspect of CALL reading.

Data analysis generated a final coding of 4 categories and 19 corresponding themes, representative of 224 comments. The following is a detailed description of the categories and related themes. In reporting the data, a letter indicating their WM group and a number denoting their interview sequence in a WM group was assigned to each of the 15 interviewees, where, for example, H3 stands for the third student among the 5 HWM interviewees.

Category one: Reading strategies

The 15 interviewees reported a wide range of reading strategies. Six major themes emerged, including: guessing as the first strategy, using dictionary/annotation as the first strategy, using dictionary / annotation as the second strategy, focusing on key information, referring to background information, and reading the comprehension questions first.

Theme 1: Guessing as the first strategy

Nine comments by 6 interviewees across the three WM groups were coded under this theme, with the high WM interviewees generating 6 of the 9 comments. These comments reflect a tendency to use contextual clues or word-building knowledge as the first strategy to deal with vocabulary problems in reading. For instance, a student [H5] said:

I will directly skip over any new word or phrase that comes up during reading. If it is so important that I cannot go on reading without knowing its meaning, I will try to infer its meaning by using its contextual clues.

Theme 2: Using dictionary/annotation as the first strategy

Coded under this theme, eleven comments from the three WM groups related to using dictionaries or annotations as the first strategy when an unknown word comes up. These comments point to a great dependency on external aids to solve reading problems. Several reasons behind this strategy were mentioned by 7 students, including "*for in-depth reading*", "*for the exact meaning of a word*", etc. For instance, an interviewee [H2] described her strategy in this way:

I know look-ups will break the reading process, but you are not sure if your understanding is correct. So if possible, I will look up the new words, especially those of high frequency of occurrence in a text. I think they are probably key words, and they are critical to reading comprehension. I always have an e-dictionary on hand while reading.

The data showed that students turned to a number of sources for the meanings of new words, including paper dictionary, e-dictionary, online dictionary, and annotations.

Theme 3: Using dictionary/ annotation as the second strategy

Six interviewees across the three WM groups contributed 4 comments to this theme. The key point of the theme is: figuring out the possible meaning of a new word through guessing first, and then referring to a dictionary or annotation for confirmation. A typical practice reported by the students was that they would try to get a general idea of the text for the first reading, while underlining the unknown words, and for the second reading, they would look up these words in a dictionary or click the annotations to get the their exact meanings [H1, H2, H4, M1, and L1].

Theme 4: Focusing on key information

This theme generated the highest number (17) of comments throughout all the WM groups and was the most representative theme among all interviewees (10 out of 15 contributed to this theme). These comments showed a wide range of practice, including: "*ignore new words in examples*", "*only focus on information related to reading questions*", "*skip unimportant words*", and "*skip uncommon words*". The students' comments reflected a tendency to ignore problems that do not stand as barriers to fulfilling reading tasks. For example, a student [H3] said "*some new words are important to your comprehension, while others are not. If they appear in examples, usually they are not important, and I don't care about them.*"

Theme 5: Using background information

Seven comments from the three WM groups were connected with using background knowledge. Five interviewees related to consulting the Internet or annotations available in

case of an unfamiliar topic or insufficient background knowledge [H1, H2，M3, L2, and L5], suggesting the importance of background knowledge for successful reading. However, one student [M2] denied this importance, and another expressed reluctance to refer to background knowledge when taking a reading comprehension test.

Theme 6: Reading the comprehension questions first

Nine comments were linked to the strategy of reading the questions before reading a text. The reasons behind this strategy include orientation to the text, task reduction, and time saving measures. For example, the students reported the following: by reading the questions first, readers can get a rough idea of what the text is about in advance [M3]; with the reading tasks in mind it is easier to search for and locate the important information [H3]; it helps readers focus their attention on relevant information [M5]; it saves time [H1], and enables readers to do in-depth reading according to the reading questions [H2]. The strategy is exercised especially in a reading comprehension test.

Category two: General perceptions of annotations

Most of the interviewees presented balanced views on annotations, by pointing out both their merits and demerits, and some expressed views on when and how annotations should be provided or used. Specifically, three major themes emerged in regard to the general perceptions of annotations, including positive views of annotations, negative views of annotations, and annotations provided / used conditionally.

Theme 1: Positive views of annotations

Interviewees across the three WM groups contributed 26 comments regarding their perceptions of the benefits of annotations in a text. Three major points of the positive views were elicited from the data. First, annotations benefit reading performance, as they can lead to precise and accurate understanding [H2, M2, M5, L2, L3, and L5], saving

readers from making incorrect guesses or misunderstanding [H4, M4, H5, M5, L2, and L3]. Second, they facilitate the reading process because of their convenience of use [H3, M3, and M4], graphical presentation of ideas [M4 and L4], and improvement of reading speed [M3, M4, and L4]. Third, they have positive psychological effects on readers, for they can give readers confidence in reading [H2], motivate their interest in reading [H2, M3, and L3], satisfy their curiosity [M4], and keep them more focused on reading [L3].

Theme 2: Negative views of annotations

The key points of the negative views include: the use of annotations disrupts the reading process [M3, L2, L3, and L5] and distracts attention [H2], harms reading ability by preventing guessing based on contextual clues [H1 and M3] or by developing dependence in readers on annotations [M1, M3, and L2], leads to poor memory of the words or previous information, and reduces the time for active thinking [L2].

Theme 3: Annotations provided / used conditionally

The interviewees generated 12 comments in relation to the theme of providing or using annotations conditionally. As for when annotations should be provided, the key points are that annotations should not be provided (a) in an exam [H1], (b) to a simple text [H5 and L4], or (c) to uncommon or unimportant words but to key words. This indicates that students do not like a text being over-annotated. Nor do they like being forced to use annotations (as was the case in this study), but rather use them according to individual needs [M2 and L1].

Category 3: Perceptions of different annotation types

While the above categories concern annotations in general, this category focuses on the three specific annotation types. Sixty comments across 15 interviewees representing all WM groups contributed to this category, represented in 6 themes: (a) positive views of

verbal annotations, (b) negative views of verbal annotations, (c) positive views of visual annotations, (d) negative views of visual annotations, (e) positive views of verbal plus visual annotations, and (f) negative views of verbal plus annotations. That is, two opposite opinions emerged regarding each of the three annotation types.

Theme 1: Positive views of verbal annotation

Eight interviewees voiced 16 positive comments on verbal annotations from different perspectives. The key points are that they are easier to understand [H1, H2, and L1], clearer, more specific, and less distractive [M4 and L2] than pictures, look consistent and harmonious with the texts [M5], reflect a native speaker's way of thinking, and contain some cultural elements [L2].

Theme 2: Negative views of verbal annotation

On the other hand, the interviewees also mentioned the demerits of visual annotations. A major concern is unknown words in the annotations. Several interviewees [H5, M4, L3, and L5] said that any new word in a verbal annotation would make it hardly an annotation. Other concerns include that it is dull [H3 and L3] with words (*in annotations*) added to words (*in the text*), and takes time to read, thus slowing down reading speed [H5 and M1].

Theme 3: Positive views of visual annotation

Three students contributed 5 comments to this theme. For one student, visual annotations are "*vivid and lively*", and make the reading more interesting than otherwise [H2]; for another, pictures convey the visual meaning of a word, which is more "*direct" and intuitive*" [M3]. L3 added that pictures are "*motivati*ng" and "*stimulating*" and that viewing pictures can "*relax eyes to some degree*" in a lengthy reading process. These comments indicate the psychologically positive effects of visual annotations.

Theme 4: Negative views of visual annotation

In sharp contrast to the above theme, interviewees across the three WM groups

expressed many more (16) negative comments on visual annotations. They described this type of annotation in such negative ways as "*confusing*", "*ambiguous*", "*not meaning-focused*", "*not meaning-targeted*", leading to a general conclusion that some visual annotations fail to help. Other comments include "*interrupt reading process*" [H1], "*cause interference*" [M5], and "*add load to memory*" [M5].

Theme 5: Positive views of verbal plus visual annotation

Eight interviewees across the three WM groups expressed their preference for this annotation type. The main reason is that when a word is both verbally and visually annotated, the text and picture complement each other, so that the meaning of the word is conveyed clearly and exactly, thus leading to deeper understanding of the target word. Other comments relate to improvement of reading speed and a better memory of the annotated words.

Theme 6: Negative views of verbal plus visual annotation

Unlike all other themes under this category, only one student [M5] gave a comment. She preferred verbal annotation to any other type of annotation, and noted that when an English text was given annotations in English, the whole things (*text and annotations*) looked "*consistent and harmonious*", and a different medium, like a picture, would "*add [mental] load*" and "*cause interference to reading.*"

Category 4: Perceptions of annotation design

The fifteen interviewees reported a wide range of opinions of the design of annotations, from which four major themes emerged: convenient to use, emotionally supportive, informative and clearly intended, and concise and uncomplicated.

Theme 1: Convenient to use

Interviewees across the three WM groups contributed 16 comments on this theme. Six

comments pointed to a need to put a text on only one web page or on as fewer pages as possible [H3, M1, M2, M5, L2, and L4]. This is because with more pages it is difficult to search for information, and frequent page turning will lead to poor memory or decrease interest in reading. Other suggestions include that annotations should be provided overtly rather than covertly [M2, M4, L5, and L4], and that the Internet access should be permitted for more references [H4 and L5]. Interviewees also acknowledged the simultaneous presentation of word explanations and pictures [H2 and L2], and expressed complaints of being unable to take notes [L2 and L5]. These comments indicate that the design of annotation should be convenient to readers.

Theme 2: Emotionally supportive

There are 15 comments connected to this theme. The key points are that reading requires a calm mind, and therefore annotations should be "*comfortable*" and "*enjoyable*" to read [H2, H3, and M2], help focus attention rather than disrupt thinking [H3, M1, M3, M4, and L2], refresh and motivate readers [H1, H5, and M4] rather than make them feel tired and dull [H5, M1, M2, and M4]. As a student [H2] noted, "*Pictures should be vivid but not complicated. It is good to exclude sound, which distracts attention and upsets mood …. Reading requires a calm and undisrupted mind.*"

Theme 3: Informative and clearly intended

Twelve comments by 8 interviewees across the three WM groups suggested that annotations should be informative and clearly intended. The interviewees spoke of adding phonetic symbols and usage of a word to its annotation [H2 and L1], and more pictures to a word [L1, L2, and H4]. Two comments are linked to highlighting the intention of pictures [H1 and H3], as pointed out by H3, "*The focus of a picture should be highlighted, for example, by putting an arrow or circle, thus reducing the distraction by other information*

[in the picture]."

Theme 4: concise and uncomplicated

Ten comments by 7 interviewees representing all the WM groups are associated with the reduction of cognitive load in annotation design. The key points are that word definitions should be short, without unknown words [M5 and L4], and pictures should be easier to understand [H3 and M1]. They mentioned the principle of designing annotations, noting that they should be brief and concise, motivate readers and help focus their attention on reading [H1], facilitate the reading process [M4], avoid additional load on readers [M3], and do not interrupt thinking [M1, M4, and L2].

In a word, the interview data provide a description of the strategies that the interviewees generally adopt and their perceptions of annotations in CALL reading. The themes generated from the interviews provide some insights into the CALL reading process and some references for annotation design. Generally, the interviewees tried a wide range of approaches to the difficulties in CALL reading. This suggests that, on one hand, the students have some prototypical reading strategies, and on the other hand, their reading strategies vary according to certain contexts. Another major finding is the fact that, with the exception of the theme "negative views of verbal plus visual annotations", there are comments from all the three WM groups, suggesting that WM is not a distinguishing factor regarding learners' perceptions of annotation use. Therefore, no doubt, other individual learner factors may work together to influence their perceptions. This in itself is an interesting topic to explore in future research.

4.5 Summary

This chapter reports and summarizes the quantitative and qualitative results related to

the research questions of this study. The quantitative results indicated that annotations

provided at word and topic levels could help enhance reading comprehension, especially

when the texts were provided with both verbal and visual annotations. The results also

indicated that the effect of annotation use on reading comprehension made the greatest

difference to learners with low WMC and the least to learners with high WMC. There was

no evidence to suggest that any one annotation type was particularly more beneficial than

another according to WMC, which ascribes a significant role to WM in explaining

individual differences in L2 reading. The results of a questionnaire and interview surveys

showed that generally the learners, regardless of WMC, viewed annotations as useful to

reading comprehension and that they favored the verbal plus visual annotations the most.

The strategies they adopt in CALL reading, the perceptions they have of different types of

annotations, and the views they hold of annotation design suggest the complexity of CALL

reading on the one hand, and point to a need for further research into the CALL reading

and annotation design on the other hand. These results will be discussed in the next chapter.

5   DISCUSSION

This study sought to increase our understanding of the effects of annotations on L2 reading comprehension in a CALL context. Four specific research questions were addressed quantitatively and qualitatively in the previous chapter. This chapter will discuss the results for each of the research questions, making reference to the literature, and providing possible explanations for the findings. It unfolds in the order of the four research questions, and concludes with a summary of the whole chapter.

5.1 Discussion of the results for the first research question

The first research question investigates whether annotation use makes a difference to reading comprehension. The statistical results indicated that the mean reading score under each of the annotated reading conditions was significantly higher than that under the no-annotation condition. This suggested that annotation use is likely to lead to better reading comprehension. The following will provide possible explanations of this finding, followed by discussions of the different perceptions of annotation use.

Understanding the benefits of annotation use for reading comprehension

The benefits of annotation use for reading comprehension, as revealed by this study, can be interpreted with reference to the interactive model of reading (Rumelhart, 1977). According to this model, both top-down and bottom-up processes occur during reading, and comprehension results from the simultaneous interaction of the two operations. In this study, annotations were provided to the reading texts at both word level (vocabulary knowledge) and topic level (background knowledge), in verbal and visual forms, as related explanations to facilitating the reading process. On the one hand, vocabulary knowledge is

very important for reading comprehension and is the best predictor of success in reading

comprehension (Anderson & Freebody, 1981). If new words occur in a text, posing

difficulties for reading comprehension because of the missing information, readers might

need relevant explanations. The words annotated in the present study were identified

through piloting. It is postulated that the availability of the word annotations might have

facilitated reading comprehension in several ways. In the first place, according to the

noticing hypothesis (Schmidt, 1990), learners must first notice the linguistic features or

forms from the language input in order to learn new linguistic forms. Annotations give

salience to the unknown words, and the requirement of the readers to click the word

annotations resulted in their noticing of the annotated words. This noticing is assumed to

be beneficial to word processing. Some studies (e.g., Bowles, 2004; Yanguas, 2009) have

found that noticing of words with different types of annotations led to better learning of the

words. Therefore, where annotations promoted noticing, this could lead to better

processing of the unknown words, which in turn was likely to facilitate reading

comprehension. Secondly, with the missing information obtained quickly via annotations,

reading became easier, more enjoyable and interesting (Ercetin, 2003; Tozcu & Coady,

2004), and the readers were supported, with the ability to process the meaning of a

sentence where a new word resides. As a student [H2] noted:

> Generally, annotations are good for reading comprehension. If you know the
> meaning of the new words [*by referring to annotations*], you can comprehend the
> text better. Annotations give you confidence. If there are too many new words in a
> text, you just cannot read it, or even give it up. … Visual annotations are vivid and
> lively, and make the reading more interesting than otherwise.

Third, when vocabulary processing is automatic, more cognitive resources can be

given to top-down processing. As Chun and Plass (1996a) point out, when more words are

recognized or known, more attention can be paid to top-down interpretation and to overall comprehension. In this study, the easy access to the missing information of new words was likely to make up for insufficiently automatic bottom-up processes and thus allowed the readers to attend to top-down processes. The above postulation that word annotations facilitated reading comprehension gained support from Chun and Plass (1996a), Lomicka (1998), and Yeh & Wang (2003), who found that word annotations allowed for deeper processing, and thus effective reading.

On the other hand, providing background information through annotations could compensate for the lack of topic knowledge in the readers, and thus enhance their reading comprehension. The interactive model of reading suggests that reading comprehension depends not only on one's linguistic knowledge, but also on his or her general knowledge of the world and the extent to which that knowledge is activated during processing. Researchers in the field of reading have acknowledged readers' background knowledge to be one of the main contributors to successful comprehension (Alderson, 1984; Carrell, 1983; Rumelhart, 1980), and revealed that providing extra topical information through annotations could compensate for the lack of prior knowledge needed for successful reading (Akbulut, 2008; Ariew & Ercetin, 2004; Liu, 1993). In the present study, the text topics were found through piloting to be generally unfamiliar to the students. Therefore, the topic annotations provided for the reading texts might have benefited the readers in several possible ways. First of all, with the topic knowledge, readers were able to anticipate and check later text development and the author's perspective with respect to the information presented. In this way, prediction might allow the readers to evaluate the information, deciding whether it was useful or not. Regarding this point, Smith (2004) claims that prediction, as the core of reading, brings potential meanings to texts, reducing

ambiguity and eliminating irrelevant alternatives. Secondly, topic knowledge could

facilitate inference of word meaning during reading (Pulido, 2003, 2004). Thus, a student

who had limited vocabulary might rely on the topic knowledge that s/he had in deciphering

unknown words and read the text more effectively. Thirdly, with topic knowledge, readers

could concentrate on the global meaning of a text by reasoning the connections between

the isolated words, phrases and sentences. In addition, topic knowledge could help activate

other background knowledge in readers, which joined with new text information to

produce comprehension. On the contrary, where there is a lack of the prior knowledge

about the text topics, the interaction was less likely to take place, and comprehension might

be hindered. In the present study, since the participants were required to click both the

word and topic annotations, they were likely to use both levels of processing while reading,

although they might rely primarily on one at a given time. For example, a student [H2]

said:

> Before I read a text itself, I usually read the topic annotation first to know what the
> text is going to talk about. This helps the reading a lot. … When I came across an
> unknown word in the text, I clicked the annotation to make sure of its meaning. All
> these information helped me understand the text better.

When both the bottom-up and top-down processing were supported through the use of

annotations, their interaction was likely to be facilitated, resulting in better reading

comprehension. Most previous studies on the relationship between annotation use and

reading comprehension only considered word annotations, and only a few studies (Ariew

& Ercetin, 2004; Ercetin, 2003; Sakar & Ercetin, 2005) considered both word annotations

and topic annotations. However, none of the two studies found that, in contrast to the

finding of the present study, annotations led to better reading comprehension. The different

relationship that they revealed might be tentatively explained with reference to the

cognitive load theory. According to the theory, provision of additional input using a variety of presentation modes may create an extraneous cognitive load, and lead to redundancy or split attention effects, thus hampering reading comprehension. The annotations used in these studies included verbal, audio and visual modes, involving text, graphics, and video clips. It is possible that the multiple presentation modes which represented a duplication of information led to redundancy effects (Moreno & Mayer, 1999), overloading the readers' working memory (WM), or distracted them by drawing their attention away from the intended tasks, causing the reverse modality effect. The interviewees in Ariew and Ercetin's study reported that they tended to use almost all of the annotations, which provides evidence for the overuse of annotations. This possible explanation can find justification from the study by Aust, Kelley and Roby (1993), which revealed that when annotations were richly accessible, readers used them excessively, thus interfering with the reading process. Nevertheless, the finding of these studies have important implications for the design of multimedia annotations: it seems that annotations given in a variety of presentation modes can produce deleterious effects on reading comprehension. This will be further discussed later in this chapter.

From another perspective, the difficulty level of the reading texts (in relation to the participants' proficiency level) used in the present study might be a factor contributing to the positive relationship between annotation use and reading comprehension. The four reading texts were found to be demanding, as evidenced by the mean score under no-annotation condition ($M = 6.65$, 51.25% correct). It is envisaged that in general the more difficult a text is, the more likely that annotations are helpful, as pointed out by Nagata (1999) that the difficulty of a text determines in a sense the effects of annotations. However, few studies reported the difficulty of the reading materials vis a vis learners'

proficiency level. Further research may explore the extent to which the difficulty of reading texts impacts the efficacy of annotation use.

Understanding the different perceptions of annotation use

The data from questionnaires and interviews generally support the use of annotations in reading. As reported in Section 4.3.1, 72% of the participants felt annotations in the reading texts to be beneficial to reading comprehension. This finding is consistent with those revealed in some previous studies (e.g., Ariew & Ercetin, 2004; Ercetin, 2003; Lomicka, 1998; Roby, 1999; Sakar & Ercetin, 2005). However, it is worth pointing out that the participants did not hold "unanimously favorable" (Davis & Lyman-Hager, 1997, p. 67) attitudes towards annotation use. For example, it was found that 26% of them indicated a dislike for reading annotated texts, and some negative comments on annotations were elicited from the interviews, such as "*disrupt reading process*", "*distract attention*", "*lead to poor memory of the words or previous information*", and "*reduce the time for thinking.*" As a student [L2] said:

> …, but if you know the meaning of a word quickly [*by referring to its annotations*], it does not stay in your memory long. New words in a text can train your reading ability, so it [*the use of annotations*] affects your ability to think actively. Since you have less time for thinking, your reading comprehension is poorer.

Although the negative views were not the mainstream of the perceptions, they must be taken into account, for they might provide some valuable insights for the design and research of annotations.

It is argued that individual differences in reading strategy might be an important factor for the different perceptions of annotations. Barnett (1988) defines reading strategies as the mental operations involved when readers approach a text effectively and make sense of what they read. As reported in Section 4.4, the interviewees adopted a wide range of

strategies, represented in six major themes (guessing as the first strategy, using dictionary / annotation as the first strategy, using dictionary / annotation as the second strategy, focusing on key information, referring to background information, and reading the comprehension questions first). It is interesting to note that some students used guessing instead of annotations, while others would rely on annotations rather than guessing, as the first strategy to solve vocabulary problems. Those who believe in the value of this strategy and thus tend to apply the strategy in reading are less likely to think highly of word annotations. Several interviewees indicated that they would rely on contexts and word formation knowledge to deal with unknown words, and did not like a text being over-annotated, or being forced to use annotations. Therefore, if annotations are unnecessarily given to readers or readers are forced to access them, they may function adversely by breaking readers' thinking processes, or adding to cognitive load, or causing emotional discomfort. In these cases, they are interfering with, rather than scaffolding, reading comprehension. It is not to be assumed that certain reading strategies will necessarily be best for all students. As Williams and Burden (1997) point out, no one strategy is inherently better than another, but rather needs to be evaluated in terms of its effectiveness for the individual learner in the completion of the reading task at hand. The position here is that the trial of new strategies for annotation use should be encouraged, particularly if one has a very narrow strategy focus, or is performing badly on reading tasks. Besides, the successful use of annotations can promote interaction between the reader and the text, which in turn is more likely to result in the development of good reading strategies. Few studies have investigated CALL reading strategy in relation to annotations. While it was beyond the scope of this study to include an in-depth analysis, it suggests that reading strategy is likely an important component to consider in any description of the potential benefits of annotations.

To conclude this section, the results for the first research question pointed to greater benefits of annotation use for reading comprehension. This section advanced possible explanations for why annotations could assist reading comprehension. Understood from the perspective of an interactive model of reading, word annotations aided bottom-up process, topic annotations helped top-down process, and the simultaneous interaction of the two processes facilitated reading comprehension. Understood from the perspective of text difficulty, the students might rely more on annotations to comprehend the harder texts, and thus annotations had the potential to lead to a larger gain in reading performance. The qualitative data as a whole triangulated the quantitative findings, and the participants' different perceptions of annotation use are assumed to be partially due to their individual differences in reading strategy.

5.2 Discussion of the results for the second research question

The second question investigates whether annotation types yield differential effects on reading comprehension. The statistical results indicated that the mean reading score under VVA (verbal plus visual annotation) was the highest and that under VBA (verbal annotation) and VSA (visual annotation) was similar. This suggested that the combination treatment of providing VVA is likely to result in superior reading performance and the single treatment of providing VBA and VSA produce similar reading performance. The following will tentatively explain the results, considered from the perspectives of the cognitive theory of multimedia learning (Mayer, 2001) and individual differences in cognitive style, and discuss the learners' perceptions of different annotation types.

Understanding the difference in reading performance according to annotation type

The superior reading performance associated with VVA could be explained by

reference to the cognitive theory of multimedia learning (Mayer, 2001). The theory posits that verbal and visual inputs are two complementary modes, which can be integrated in working memory to build connections between mental representations (a verbal model and a visual model). According to the multiple representation principle (Moreno & Mayer, 1999) formulated for instructional design to foster multimedia learning, it is better to present an explanation in both words and pictures rather than solely in words or in pictures. In the present study, VVA provided both verbal and visual information for a given vocabulary item or a text topic. Theoretically, when they entered the readers' verbal and visual cognitive system respectively, the simultaneous activation of the two cognitive systems offered a chance for the integration of the information. As Akbulut (2008) contended, verbal and visual annotations, by supplying information via two complementary channels for bottom-up and top-down processing, could facilitate an interactive reading process, thus allowing for better comprehension. In addition, one representation could be used to constrain possible interpretations or misinterpretations in the use of the other. Thus the concurrent presence of visual and verbal annotations could help the readers construct accurate and deep understanding of the unknown words and thus the texts. It is envisaged that the learners benefited from the sum of the advantages.

The statistical findings do not support the hypothetical argument that the superior reading performance associated with VVA could have resulted from accessing only one type of annotation rather than two in VVA. If the participants had consulted only the word definition or the picture in a VVA, ignoring the other, their reading performance would have been similar to that under a single treatment condition. However, the advantage of VVA over both VSA and VBA in the mean reading score suggested that it was not any single type of annotation that contributed to the superior reading performance of VVA.

Another possible explanation is that VVA, as a combination of both a verbal and a visual annotation, could cater for different types of cognitive style. According to Riding and Rayner (1998), cognitive style is an individual's preferred and habitual approach to organizing, remembering and representing information. The construct of cognitive styles suggests that verbally oriented learners and visually oriented readers may differ in the extent of preferences for using textually or graphically presented information to construct meaning. Massa and Mayer (2006) propose a verbalizer-visualizer hypothesis, which claims that those who prefer verbal models of presentation tend to rely more on verbal inputs whereas those who prefer visual modes of presentation tend to rely more on pictorial inputs. Plass, Chun, Mayer and Leutner (1998) argue that readers might demonstrate better reading comprehension when they have the opportunity to receive their preferred information presentation mode. Therefore, it is possible that the verbalizers among the participants of this study benefited primarily from the verbal information in VVA, whereas the visualizers benefited primarily from the visual information. This might explain the smaller standard deviation score for VVA, as VVA provides a resource accessible to verbalizers and visualizers alike. It might also be assumed that the verbally oriented participants benefited as much from verbal annotations as visually oriented participants did from visual annotations. When only one type of annotation was available, it could accommodate one type of cognitive style, and thus the assignment of the participants to a treatment condition which did not cater for their cognitive styles might result in a level of cognitive load. For instance, visualizers who prefer to learn with visual information but only received verbal information might experience a high cognitive load that might prevent them from effectively processing information. Therefore, by exploiting the combination of verbal and visual annotations, readers were less likely to be limited by the weaknesses of

any single one. Accordingly, the perspective of individual differences in cognitive style might provide a tentative explanation of the similar reading performance under VBA and VSA as well as the superior reading performance under VVA.

However, cognitive style was not a focus of the present study, and there were no data on the participants' cognitive styles. In fact, there has not been enough research yet in multimedia environments to really understand to what extent (if any) cognitive styles mediate the effectiveness of annotation types, but it could be an avenue for future research, particularly in contrast to an understanding of the efficacy of annotation types from a multimedia learning theory perspective as discussed above.

The superior reading performance associated with VVA confirms the findings in some previous researchers (e.g., Plass et al., 1998; Yanguas, 2009). The qualitative data overall supported the use of VVA. Firstly, as reported in Section 4.2.3, the questionnaire survey indicated that generally the participants favored reading a text with verbal plus visual annotations, and agreed that this annotation type facilitated reading comprehension. Secondly, as reported in Section 4.4, the interviews elicited far more positive views on VVA than negative ones. The interviewees mentioned that words accompanied by pictures led to *"accurate understanding", "faster reading speed"* and *"a better memory of the information."* For example, a student [M1] said:

> I like verbal and visual annotations the most. Usually I look at the picture first to get a rough idea of a new word, and then read the definition for its exact meaning. To take the word "pupation" for an example, if you look at the picture alone, you are not quite sure what it is meant to mean exactly; if you look at the words [*definition*] alone, you are not quite clear about the meaning either, for it is somewhat abstract. In combination, you will quickly and accurately figure out the meaning of the word.

Thus, the participants' perceptions of the integrated mode of annotations triangulated their superior reading performance under this condition.

One concern raised in studies involving VVA is the split-attention effect. According

to Betrancourt and Bisseret (1998), a split-attention effect occurs in learning situations where multiple sources of information are presented separately that need to be integrated in WM to be understood. That is, physically separated multiple sources of information may lead to extraneous cognitive load. Using resources to process extraneous load will reduce the amount of resources available to process the intrinsic load, due to the limited cognitive resources. Thus especially when intrinsic load is high, materials should be designed so as to reduce the extraneous load. In the present study, the verbal and visual annotation in a VVA appeared in close proximity in a pop-up window simultaneously. It is less of a load than if they were far apart on the screen, because in that case the eye must track across the screen and take in the information separately. This picture-text adjacency design of VVA is in line with the Moreno and Mayer's (1999) contiguous principle: It is better to present corresponding words and pictures contiguously rather than separately when giving a multimedia explanation. In this, the split-attention effects were minimized. This presentation format was acknowledged in a view generated from the interviews that annotations should be convenient to use (Theme 1, Category 4) --- supporting the necessity to reduce extraneous cognitive load. For example, one student [M5] said "... *It is good that the picture and the definition are displayed on the same box [window], or you have to scan from one place to another. ...*".

Understanding the different perceptions of VBA and VSA

The qualitative surveys showed that the participants reported both positive and negative views of VBA and VSA. This is argued as being related to their cognitive styles. Although cognitive style was not investigated in this study as a variable, there was evidence that it might influence students' perceptions of annotations. For example, a

student [H1] said:

> … With visual annotations added, however, your attention is easily drawn to the pictures, which will interrupt your reading process and lead to poor understanding. … I like the visual annotations the least. Some pictures are not meaning-focused. For instance, the picture for "archipelago" --- you just don't know what you should focus on, the sea or the islands.

While another student [L3] said:

> A picture portrays the meaning of a word, that is, conveys its visual meaning. It is more direct and clearer in conveying meaning. … Pictures are motivating and stimulating, especially when you have been reading for a long time.

The two students are likely to have different cognitive styles, being on the opposite ends of the verbalizer-visualizer spectrum.

It is worth noting that the participants performed similarly under VBA and VSA despite their relatively low perceptions of VSA. The discrepancy seems to suggest that learners' perceptions may not directly affect their reading outcome. In fact, visual aids have long been assumed to be beneficial to language learning. This is because they can stimulate curiosity and thus attract learners' attention, which makes their associative meanings more memorable. Oxford and Crookall (1990) maintain that most learners can associate new information with concepts in memory by means of meaningful visual images, and that visual images make learning more efficient. Babaie (2010) also contends that pictures are better remembered than words, and might help learners better decipher the meanings of words, and thus learners might then arrive at a better understanding of reading passages. As reported in Section 4.4, the positive views of VSA elicited from present study point to its psychologically positive effect on readers. It is possible that although pictures might be less explicit than definitions and thus impose a high cognitive load, the germane cognitive load produced by the positive psychological effect on the reading process was likely to act

as a kind of tradeoff. This might partially explain the lower opinion but equal effect associated with VSA.

Summarising this section, the types of annotation can produce different effects on reading comprehension, with VVA resulting in the best reading performance. The superiority of combination treatment supports the cognitive theory of multimedia learning. The different perceptions of VBA and VSA were tentatively interpreted from the perspective of individual differences in cognitive style. The discussions, on the one hand, suggest a need to consider cognitive style in further research, and on the other hand, point to a complex picture of the potential of annotations for reading comprehension. As Al-Seghyer (2005) points out, the effect of individual learner differences on text comprehension in multimedia environments has not been fully studied and is in need of thorough investigation.

5.3 Discussion of the results for the third research question

The third research question explores whether annotation use makes a difference to reading comprehension based on learners' WM. The statistical results showed that access to each of the three annotation types is likely to lead to better reading performance for the low WM (LWM) and medium WM (MWM) learners, but is unlikely to yield differential reading performance for the high WM (HWM) learners. In addition, the effect size of each of the three annotation types was largest for the LWM group and smallest for the HWM group, that is, the efficacy of annotations was inversely proportional to the capacity of WM. This suggests that annotation use has differential benefits for reading comprehension according to readers' WMC. The following will explain why annotation use benefited the LWM learners the most and the HWM learners the least.

__Understanding the apparent benefits of annotation use for LWM learners__

The apparent benefit of annotation use for LWM learners is best understood with regard to the nature and role of WM in reading. As suggested by the theory of WM, WM has a limited capacity, which differs among individuals, and the difference can result in differences in reading performance (Alptekin & Ercetin, 2009; Baddeley, 2007; Cantor & Engle, 1993; Daneman & Carpenter, 1980). Generally when reading materials themselves are easy, the performance differences of readers with different WMC are small or insignificant, and these differences get larger when materials are difficult to understand (Budd, Whitney, & Turley, 1995). When annotations are provided to hard materials, reading is likely to become easier because readers have less amount of information that is unclear to them by resorting to annotations. In the absence of annotations, readers may have to interpret the input using bottom-up and/or top-down approaches, and may have to modify the interpretation when they confront new information inconsistent with their initial interpretation. If this inconsistent input is too frequent, reading becomes disjointed, creating a greater burden for WM, or requiring more WM involvement. This will be more disadvantageous to readers with lower WMC because they have less capacity to keep the earlier relevant information active in WM while processing the new information. Therefore, lower WM readers are more likely to rely on annotations to compensate for their capacity limitations. The following will discuss specifically why word annotations and topic annotations are more helpful to LWM readers.

First, word annotation is discussed in relation to cognitive load on WM. Unknown words in a text tend to induce an intrinsic cognitive load on readers' WM. Word annotations, however, provide context-specific meanings for unknown words, saving readers from working them out, or preventing them from making incorrect guesses.

Therefore, words annotations can in a sense alleviate text difficulty and reduce intrinsic cognitive load on WM. Meanwhile, the access to word annotations allows readers to pursue reading without disturbing their reading process, a problem usually caused by stopping to look up unknown words in a dictionary, thus reducing extraneous cognitive load on WM. Moreover, when less of readers' attention is spent on word identification, more germane cognitive load becomes available for text decoding. As Al-Shehri and Gitsaki (2010) suggest, the use of the online dictionary could improve students' WM by reducing intrinsic and extraneous cognitive load, and improve reading comprehension. This may explain why readers tend to look up unknown words immediately upon encountering them, particularly if this can be done simply (Chun, 2001). Accordingly, word annotations are assumed to be more beneficial to LWM readers, for the mental resources that can be expended on reading are relatively small, and the use of annotations can compensate for their limited WMC.

Second, topic annotation is discussed in relation to cognitive load on WM. The interactive reading model (Rumelhart, 1977) posits that reading involves an interaction between the reader and the text, where the reader constructs meaning in his or her WM based partly on the information extracted from the text and partly on the existing background knowledge that the reader has. WM reflects the ability to maintain information from the text and background knowledge in an active state. When background knowledge is missing or insufficient, the reader has to allocate more of his or her limited WMC to try to build content schema for the interaction. Topic annotations, which provide background knowledge for the reader, can therefore in a sense compensate for his or her WMC by reducing mental resources required for schema construction, and free WMC for other activities (Van Merrienboer & Ayres, 2005). For example, one of the reading texts in the

present study is about "dung beetle". The pilot study established the participants to be generally unfamiliar with the topic. Thus the intrinsic cognitive load caused by the text content was likely to be higher on the readers' WM. The topic annotation developed for the text, however, offers a brief introduction to the beetle's intelligence and behaviors, which could promote reader-text interaction. In addition, as mentioned in Section 5.1, background knowledge can contribute to inferring the meaning of unknown words, which in turn lessens WM load incurred by these words. Hence, topic annotations are supposed to be more beneficial to readers with fewer mental sources for text processing. This explanation is supported by Hambrick and Engle (2002), who, when investigating the interplay between relevant domain knowledge and WM on young adult listeners' comprehension, found that relevant background could compensate for low WM during comprehension.

Understanding the unapparent benefits of annotation use for HWM learners

However, the results for the third research question also showed that annotation use did not make a significant difference to the HWM learners' reading comprehension, which suggests that annotation use works least effectively for them. One possible explanation is that they are skilled readers, who were likely to resort primarily to guessing rather than annotations to address difficulties in reading, as guessing is argued to be a strategy for skilled language learners (Chall, 1987; Tomitch, 1996). Suppose a skilled reader encounters the unknown word "diminish" in the following sentence:

> At the present time, the world has plentiful supplies of stored-source energy, but a natural resource such as oil will be seriously *diminished* by the end of the century.

The reader might guess the meaning of the word with the help of word formation knowledge, collocation knowledge and contextual clues: (a) the prefix *di* usually indicates "reduce"; (b) the word is expected to have a negative value since it is linked with *seriously*;

(c) *but* suggests the word being in contrast with *plentiful supplies*. In this way, his or her

foundation knowledge of what to draw on helps figure out the meaning of the word

successfully. This postulation is partially supported by Daneman and Green (1986), who

found that skilled readers were better able to use contextual cues to infer the meaning of a

novel word in a text and make predictions. The ability to make inferences and predictions

can reduce dependence on external assistances like annotations in the reading process.

Although the participants in the present study were required to click annotations, it is

possible that the HWM learners did so not only for obtaining word meanings, but also for

testing their inferences, or just for fulfilling the click requirement. In these cases,

annotations were not functioning as effective scaffolds to them. There is some qualitative

evidence in this study to justify this possibility. First, the questionnaire survey (Item 15)

revealed that more HWM learners (87.50%) consider annotations a hindrance in reading

than MWM learners (43.70%) and LWM learners (27.50%). As further evidence, a HWM

student [H1] said in the interview:

> In my opinion, even in CALL reading with all kinds of annotations, it is better not to
> refer to the annotations first. Guessing should be the first policy. Some words are
> just new words, or uncommon words, and it is not necessary to make sense of them.
> Usually they do not show up as barriers to reading comprehension.

This hypothesis could be tested in future research by building an eye-tracking device

(Winke, Godfroid, & Gass, 2013) into a CALL reading program to record readers'

interaction with annotations for the investigation of the length of time spent on reading

annotations, or by adopting a think-aloud protocol to elicit their thought processes while

using annotations.

Another possible explanation is that the HWM learners might have already known the

meanings of the annotated words or possessed some relevant topic knowledge. That is,

they could automatically retrieve or activate information from preexisting knowledge or long-term memory, rather than from annotations. Hambrick and Engle (2002) claim that people with higher WM may be able to draw upon more preexisting domain knowledge during cognitive performance than people with lower WMC.

To summarize this section, the potential benefits of annotation use can depend on WM. Annotations can reduce WM load by supplying missing information, and the limited WMC can be allotted to other comprehension processes. Thus annotations are especially helpful to learners with smaller WMC, who might rely more on them to make up for capacity limitations while reading; while for HWM learners, their larger WMC is likely to reduce the need for annotations, and thus annotation use makes no apparent difference to their reading performance.

5.4 Discussion of the results for the fourth research question

The last research question investigates whether annotation type yields differential effects on reading comprehension according to WM. It was addressed by conducting a within-group analysis and an across-group analysis. The first analysis indicated that no one annotation type is likely to be significantly more beneficial than another to reading comprehension for learners with similar WMC. In particular, the advantage of VVA over VBA and VSA to reading comprehension for all the participants as a whole, as found in Section 4.2.2, disappeared in this analysis. This seems to point to WM as a determining variable in reading comprehension, and annotations probably functioning within the limited capacity of WM. That is, annotations only play a limited role in improving reading comprehension in contrast with the predictive power of WM. The second analysis showed that the HWM group always scored significantly higher than the LWM group in reading

comprehension for any of the annotation types, suggesting that no one annotation type can bridge the gap in reading performance associated with differences in their WM. It also showed that for any of the annotation types, the LWM group scored similarly to the MWM group, suggesting that annotations could help LWM learners to the point that they could overcome their WM limitations to a certain degree.

The finding of the present study that WM predicts L2 reading comprehension in a CALL context corroborates the results of some previous studies with hypertexts. For example, as reviewed in Chapter 2, Fontanini and Tomitch (2009) found that the WM scores of the participants were positively related to their scores on recall and comprehension tests of hypertexts. Lee and Tedder's (2003) study revealed that the participants with higher WMC received higher recall scores than those with lower WM no matter whether the reading material was in a traditional text format, or in a structured hypertext format, or in a networked hypertext format. Although the two studies did not use annotations in the same sense as the present study, the subtitles of the text in the former and the elaborating excerpts in the latter functioned similarly to annotations. The following will discuss the predictive power of WM to reading with annotations and the limited role of annotations in enhancing reading comprehension.

These findings can be partially explained, again, by reference to the theory of WM. Reading is a complex cognitive skill that draws on a variety of processes, such as word-level, sentence-level, and text-level processes. Abu-rabia (2003) argues that in any of these processes WM is involved in transitory processing and storage of information. WM offers sources for storage of the information, and the information has to be held in WM for the duration of simultaneous processing of new information. Readers whose WMC is larger are better able to maintain multiple, tentative interpretations in an active mode and

use text elements to test those interpretations (Lee-Sammons & Whitney, 1991), and hence

are more successful at interpreting the text than readers with smaller WMC. As such,

among individuals of similar L2 proficiency and educational background, those with

similar WMC are more likely to perform similarly on measures of reading comprehension,

as suggested by some previous studies (e.g., Daneman & Carpenter, 1980; Daneman &

Merikle, 1996; Osamu, 2006; Tomitch, 1996). In fact, the present study yielded more

evidence that WM makes a large contribution to explaining the differences in reading

comprehension. As reported in Table 4.21, when no aids were given through annotations,

the mean reading score of the HWM group ($M = 8.78$) was significantly higher than that of

the MWM group ($M = 6.05$), and much higher than that of the LWM group ($M = 4.83$). In

particular, this score (8.78) was higher than the highest mean score of the LWM group (M

$= 8.03$, received under VVA) and the MWM group ($M = 8.30$, received under VVA), a

piece of convincing evidence for the crucial role of WM in reading comprehension. Given

that previous studies rarely considered the contribution of annotations to reading

comprehension in relation to WMC, this is an important contribution made by the present

study. The results not only confirm the predictive power of WM for reading comprehension,

but importantly, they reveal that WM is a powerful predictor of the reading performance

associated with annotations.

　　　This finding is inconsistent with that by Chun and Payne (2004). As reviewed in

Chapter 2, they found no significant difference between the high- and low-span groups in

comprehension scores, which suggested that individual differences in WM did not affect

reading comprehension. Their finding, nevertheless, should be used with caution, because

of the small sample size ($n = 13$ undergraduates) and several missing data in the reading

comprehension test.

Another possible explanation of the limited role of annotations is that neither word annotations nor topic annotations, as used in this study, could address all that was measured in the reading tasks. On the one hand, the reading test measures the comprehension of the texts, not the content of the annotations, and the annotations that were provided were not based on the reading tasks themselves, but were developed based on the words that a pilot group of students reported finding difficult in the reading texts. For this reason, some of the annotations were unlikely to be related to the reading scores. On the other hand, some reading tasks required greater support than annotations. For one example, when a learner's lexical problem was resolved through looking up a lexical annotation, other problems, like syntactic or pragmatic ones, might still remain. For another example, some of the reading tasks required understanding global information, like choosing a suitable heading for a given paragraph, where to find the gist of the paragraph required the readers to be able to summarize the paragraph. In addition, although the role of topic annotations is to provide schematic knowledge about the topic, it might be difficult for the readers to integrate the information obtained from the annotations with the information embedded in the text in a short time, given the nature of the reading comprehension test. All these and other potential factors, like L2 proficiency, reading skills, etc., might in a sense account for the differences in reading performance. Therefore, annotation use, whatever the type might be, does not appear to be able to resolve all aspects of reading comprehension, despite the fact that all annotation groups outperformed the no-annotation group demonstrates the worth of annotation use.

In a word, the present study found that the types of annotation do not produce differential effects on reading comprehension according to WM. The similar reading performances by annotation type within the same WM group and the different reading

performances by annotation type across different WM groups suggest that WM, as a powerful predictor of reading performance, plays a central role in text processing and that different types of annotation play a limited role in improving reading performance. In other words, although annotation use can enhance reading comprehension, it seems that WM exerts some kind of central bottleneck effect, and thus the gap in reading performance deriving from WM difference cannot be completely closed through annotation use. Presently there is little relevant literature on this. The possible explanations provided suggest a need for further study on the interaction between WM and annotations, preferably with larger sample sizes.

5.5 Chapter summary

The four research questions discussed in this chapter aimed at developing an understanding of the effectiveness of annotation use for L2 reading comprehension. The results clearly pointed to the benefits of annotation use and the superiority of composite annotations, that is, providing both verbal and visual information (VVA). Annotation use was most effective for low WM learners and least for high WM learners, but the types of annotation did not yield differential effects on reading comprehension according to WM. WM is a significant predictor of CALL reading performance. Annotations can scaffold and promote reading to a certain extent, but they do not appear to make up for the difference in reading performance caused by WM. The overall findings are graphically demonstrated in the following Figure 5.1.

In the figure, there are three solid line boxes, which stand for the reading outcome of each WM group under no-annotation treatment. As can be seen, the HWM group has the largest box, which means its reading outcome is the best under that condition, suggesting the predictive role of WMC to reading comprehension. The dotted-line circle within each

box indicates the effect of annotations on reading comprehension. As can be seen, annotations produce the least effects on the reading performance of the HWM group, for its circle is the smallest. The three dotted line boxes, which are larger than their respective counterparts, stand for the improved reading outcomes when annotations are introduced. As can be seen, annotations benefit the LWM group the most, for its increase in size from the solid-line box to the dotted-line box is the biggest; the reverse is true for the HWM group, but its dotted-line box still remains the largest, suggesting that annotations play a limited role in improving reading comprehension, compared with the determinant role of WMC. Each circle is divided into three parts, each allocated to one of the three annotation types. As is shown, VVA in the three WM groups always gets the largest portion, for its mean score is the highest in each of the three WM group, although the difference does not reach significance. This division indicates that VVA contributes the most to the effect of annotations on reading comprehension, and VBA and VSA to a lesser but similar extent. It is hoped that the figure can explicitly demonstrate the interaction between annotations, WM and reading comprehension.
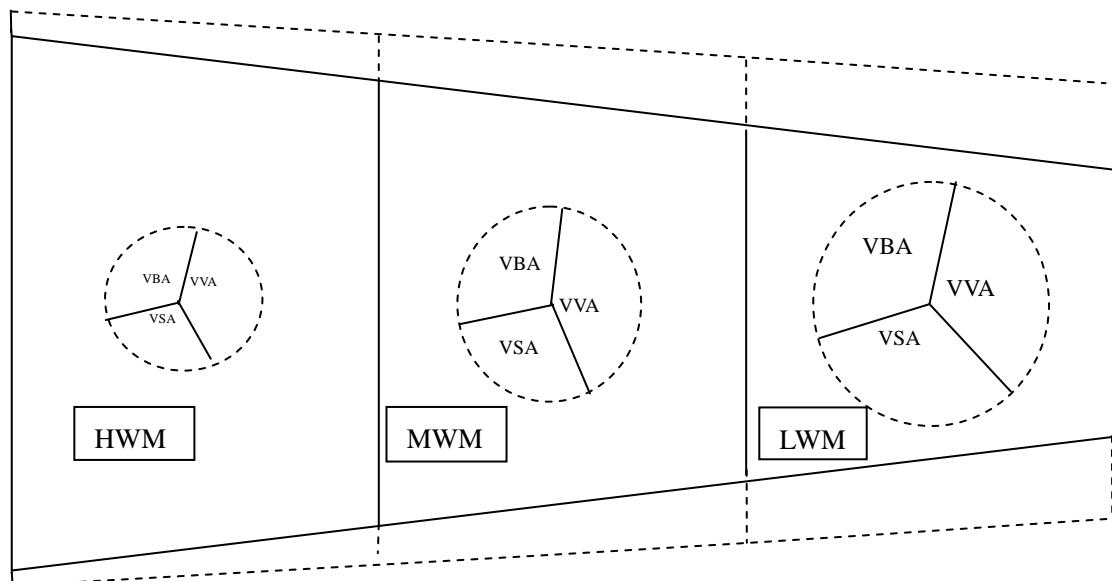

Figure 5.1 *Annotation, Working Memory and Reading Comprehension*

6 CONCLUSION

This chapter will present the major findings of the present study, discuss its possible contributions and pedagogical implications, point out its limitations, suggest avenues for further research, and provide final conclusions.

6.1 Major findings

The major objective of this study was to explore the relationship between working memory (WM), L2 learners' annotation use and reading comprehension in a CALL environment. The motivation for addressing this issue is two-fold. First, it stems from inconclusive findings regarding the effectiveness of annotation use for reading comprehension in a CALL environment. Second, there is very little research on WM as a variable in investigating the relationship, given recent findings that demonstrate the central role of WM in reading. Four research questions were formulated around the purpose of addressing the relationship between annotation use and reading comprehension, and the role of WM as a mediating factor in the relationship. The major findings are summarized as follows:

Firstly, annotation use was found to be positively associated with reading comprehension. It was argued that the word annotations and topic annotations provided for the reading texts assisted both bottom-up and top-down processing of information, thus facilitating the reading process and improving reading performance.

Secondly, the combination of verbal and visual annotations was associated with the best reading outcome. This was argued to be best explained with reference to the cognitive theory of multimedia learning. That is, with information from both verbal and visual

annotations held in learners' WM, a referential link was thus built between them, which could reduce cognitive load on WM and lead to deeper understanding. Furthermore, the dual presence of verbal and visual annotations could cater for learners of different cognitive styles, so that a learner with a preference for a verbal mode might rely more on verbal annotations, and a learner with a preference for a visual mode might rely more on visual annotations.

Thirdly, the efficacy of annotation use was mediated by individual differences in WM. Annotation use was most effective for low WM (LWM) learners, who might rely more on the annotations to compensate for their capacity limitations while reading. It made least difference to high WM (HWM) learners, who might rely primarily on their sufficient cognitive resources for comprehension.

Finally, annotation type did not yield differential effects on reading comprehension relative to WMC. Specifically, (a) learners within a similar WM range tended to perform similarly on reading comprehension measures, regardless of annotation type; (b) HWM learners tended to outperform LWM learners on reading comprehension measures, regardless of annotation type. This was arguably because WM played a crucial role in reading comprehension. That is, WM had a bottleneck effect on reading performance, and annotations probably functioned within the constraints of WMC, thus playing a limited role in improving reading comprehension.

6.2 The contributions of this research

This study brings some important insights and contributions to the research of CALL reading at theoretical and methodological levels. First, the improved reading performance through the use of annotations was predicted by two theories. The participants, when

exposed to L2 texts with both word and topic annotations, were found to be able to gain

more understanding. This is not surprising given the fact that these annotations were

provided to bridge the gap between the reader's lack of lexical and topical knowledge and

the texts themselves. This dual provision of annotations is consistent with the interactive

model of reading (Rumelhart, 1977), which recognizes the simultaneous interaction

between bottom-up and top-down processes throughout the reading process. Further, the

combined condition of both visual and verbal annotation was found to be more beneficial

to reading comprehension than either visual or verbal annotations alone. This result for the

superior benefit of dual provision of annotations is consistent with the cognitive theory of

multimedia learning (Mayer, 2001), which posits that the simultaneous and adjacent

presence of both verbal and visual inputs fosters an inferential meaning connection in WM,

and thus reduces the cognitive load on WM. Therefore, this study suggests a theory-driven

approach to annotation design as a promising means of developing effective annotations

for CALL reading.

Secondly, the present study supports L2 annotations as effective reading aids for

higher-proficiency L2 learners. The verbal annotations in the present study were provided

in L2 (English), as against L1 annotations as used in most previous studies (e.g., Akbulut,

2008; Bowles, 2004; Chun & Plass, 1996a; Nagata, 1999; Yanguas, 2009). Whether to

provide L1 or L2 annotations might pose a question to language instructors, as previous

studies that compared the effectiveness of L1 and L2 annotations yielded mixed results

(Chen, 2002; Jacobs, Dufon, & Hong, 1994; Lyman-Hager & Davis, 1996; Miyasako,

2002; Taylor, 2006; Yoshii, 2006). The present study adopted L2 annotations, as supported

by the concept mediation model proposed by Potter, So, Eckhardt, and Feldman (1984).

This model suggests that as learners' L2 proficiency progresses, they can link L2 directly

to concepts without the help of L1 interpretations. The choice of annotation language was not a focus of the present study, but it has methodological implications for future research that studies the relationship between annotation use, L2 proficiency, and L2 reading comprehension.

Thirdly, the present study makes a methodological contribution to measurement of WM in L2-related studies. Based on Waters and Caplan (1996) and Conway et al. (2005), the measurement of WM followed best practice by using different types of measures: it conducted three different WM tests (L1 RST, L2 RST and OST) to establish the readers' WMC. It was found that the three WM tests were positively correlated, with the strongest correlation found between the OST and the L2 RST. Therefore, this study adds to a growing body of research that supports OST and RST tests as representing a similar construct (Conway et al., 2005). In addition, it concurs with the positive link between L1 and L2 RST tests found in recent research (Berquist, 1997; Chung, 2008; Juffs, 2004). This has an implication for the measurement of WM in possible future research that studies the relationship between WM and L2 reading comprehension.

Perhaps most importantly, the present study indicates that WM is an important factor to consider in annotation research. Both the interactive model of reading and the cognitive theory of multimedia learning, which are two theoretical frameworks of this study, suggest that WM is an important aspect in CALL reading. However, the relationship between WM and reading comprehension has rarely been studied in relation to annotations. This study explored the interplay between WM and annotations, with the levels of the learners' WMC and types of annotation controlled for an in-depth analysis. Therefore, by filling a gap in the literature, it contributes to our understanding of how the two variables interact in influencing reading comprehension, and ultimately to our understanding of the nature and

complexity of L2 reading in a CALL environment.

6.3 Pedagogical implications

The findings of the present study carry some important pedagogical implications, which will be discussed with respect to annotation (software) design and reading strategy training.

Annotation (Software) design

In this respect, a number of areas need to be considered to make annotations more effective. First of all, instructional designers need to consider what technology can be used to develop annotations for the enhancement of reading context. Technological developments have made it possible for all kinds of learning aids, such as video, sound, picture, animation, to be integrated into computer delivered texts to assist learning. However, a misuse, or abuse of these tools could turn out to render CALL at best useless and at worst detrimental to learning. What is suggested here is not a technology-driven design, but a theory-driven one, where a medium is a tool to help achieve a learning goal, while a theory is the ground for the choice of the tools.

Secondly, instructional designers need to consider what annotation formats can promote learning. The present study provides support for incorporating verbal plus visual annotations in reading materials, which can reduce the cognitive load on WM and address learners' individual differences. Learners may prefer a typical mode of representation that works most effectively for them as individuals; thus, designers must never assume that specific types of multimedia annotations will have the same effect on all learners. Rather, they should assist learners in identifying their preferred modes, and encourage them to exploit the modes that they favor most.

Thirdly, it is important for instructional designers to consider cognitive loads that may be brought to CALL learning. For example, they should be aware that high cognitive loads may be caused by the intrinsic nature of the annotations, and for this reason, the information in the annotations should be clearly conveyed, without incurring additional loads to WM. A verbal annotation that contains unknown word(s) is a case in point, as indicated by some interviewees of this study. Furthermore, annotations should be developed in a manner that can arouse learners' interests, without splitting their attention, a crucial factor that can make annotation design a success or a failure. This is especially true of visual annotations, where a fascinating picture, for example, might on one hand motivate readers to actively think of its intended associative meaning, while on the other hand distract their attention, or interrupt the reading process, by making them stay unnecessarily longer. The psychologically positive effects of annotations should also be taken into consideration, as they can increase the germane cognitive load, a point revealed in the interview data of the present study. In a word, the general principle regarding annotation design is that annotations should eliminate intrinsic loads, minimize extraneous loads, increase germane loads, and accommodate individual differences, as suggested by Sweller (1988), Brunken, Plass, and Leutner (2003).

As a final point, it is worth pointing out that learners, in the presence of annotations but in the absence of teachers, become more user-centered in a CALL environment. They take control of their reading, and make decisions about their reading speed and strategies. In this, they are more actively involved in the reading process, and have more responsibility for their learning, which can foster autonomous learning. However this requires more careful course planning on the teachers' side. It should be emphasized that the role of a teacher in this context is equally important, if not more so, as that in

traditional learning environments. It is the teacher's job to provide learners with engaging materials and effective annotations.

CALL reading strategy training

The CALL environment adds a new dimension to reading. When readers are exposed to a widespread access to the multimedia aids used to assist reading, the process of reading becomes different, as acknowledged by 85% of the participants in this study. The traditional models of reading that emphasize interaction of bottom-up and top-down processes for successful reading comprehension is insufficient to explain text processing in a CALL environment (Sakar & Ercetin, 2005). Therefore, in contrast to paper reading, where strategies are relatively limited, more and different reading strategies are needed for multimedia reading, for example, the strategies to interpret pictures, as reported by the interviewees of this study. L2 learners are assumed to be not fully accustomed to this special reading environment, or not fully aware of the limitations of their habitual approaches to reading. Technology alone can never ensure the success of language learning and the use of technology does not automatically result in improved learning outcomes. Instructors need to train L2 readers to use reading software in the most beneficial way, by explaining to them, for example, what efficacy multimedia annotations can produce, when to consult annotations, how to establish the relations between the text and pictures and so forth. In this, learners are able to consciously think about how annotations can benefit their reading. In doing so, they actively participate in meaning-constructing and problem-solving situations, which involves the application of a wide range of strategies.

6.4 Limitations and suggestions

In spite of its useful findings, the present study has several limitations, which caution

us about the results obtained on the one hand, and provide possible routes for future studies on the other hand.

To begin with, the participants in this study were advanced L2 learners, sampled from a highly prestigious Chinese university. In China, exam-oriented education prevails among different educational institutions (due to inadequate high quality educational resources for a large population). As a result, Chinese students tend to be more score-serious about examinations. Some participants in the present study were likely to treat the reading tasks as "real" reading tests, although they had been advised that the study was conducted purely for research purposes, and their performance was not linked to any examination or evaluation. This is evidenced by the observation by the researcher that several students first clicked all the annotations in a hurry, then scrolled directly to the reading questions, and then answered each question by searching in the texts for the relevant information, as would be their strategy if taking an actual test. In this way, they did not pay serious attention to the annotations provided, which failed the researcher's expectation of them to really interact with the annotations. In addition, if there had not been a time limit, the students might have interacted with the texts differently. Therefore, the findings of the present study cannot be generalized to other contexts with different samples. A promising research method for the purpose of exploring annotation use further is eye-tracking (Winke et al., 2013). This can be a means of tracking eye movement to record learners' engagement with annotations. Another means is through the use of think-aloud protocols – learners record their thoughts while they are looking up annotations. Such methods will generate more exact data on their interaction with annotations. This is also valuable research that would complement this study.

Secondly, a likely contributing factor to the effect of annotations, given the nature of

the annotations provided, is individual differences in cognitive style. This is an area for future research, which may include an investigation of the interaction between WM, cognitive style and annotation use, as there is research evidence indicating a complex interplay of the first two variables (Riding, Grimley, Dahraei, & Banner, 2001, 2003). Besides, there may be other factors mediating the effectiveness of annotations for reading comprehension such as reading strategy, motivation, readers' proficiency, and text difficulties. These factors were not investigated in this study. Considering them in future research may add substantially to our understanding of the effectiveness of annotation use for reading comprehension.

Thirdly, the present study found that annotation use benefited reading comprehension according to learners' WMC. This evidence led the researcher to claim that (a) WM plays a crucial role in learners' reading of annotated texts, and (b) annotations, although they can improve reading comprehension, only have a limited role in this way. That is, differences in reading performance due to WM cannot be wholly compensated for by annotation use. However, this study has some peculiarities, such as forced annotation lookup and text difficulty. In particular, the English proficiency of the participants was not measured, and there might be a potential of conflating second language proficiency and WM. Therefore, the two claims should be taken as tentative, and utilized with caution. Currently, there is a lack of literature on the relationship between WM and annotation use in CALL reading, and this warrants more future research.

Finally, a longitudinal study which explores the effectiveness of annotations for reading comprehension over a longer period of time in a context where annotations are incorporated into a course curriculum would be more informative than one conducted within a shorter period of time.

6.5 Conclusion

As mentioned previously, researchers have studied how annotations may affect reading comprehension in a CALL environment. Despite the finding that annotations were generally perceived to be beneficial to reading, they were divided in their conclusions about the effectiveness of annotation use for reading comprehension; despite WM being a significant aspect to reading in a CALL setting, little research to date has examined whether it influences the effects of annotations on reading performance. The present study sought to address these issues. The results of the study demonstrate that the students viewed annotations positively, and that annotations can improve reading comprehension, with the combined format leading to the best reading outcome. However the role of annotations is limited in contrast to that of WM, which is found to be significantly predictive for reading performance and thus mediates the effects of annotations based on learners' WM capacity (WMC).

This study represents a step forward in the field of CALL towards a more comprehensive understanding of the way that computer technology can be applied to enhance language learning. It is the researcher's hope that the findings as well as discussions and suggestions can benefit CALL researchers, instructors and designers, and contribute to our understanding of how to support reading comprehension in a CALL environment.

APPENDICES

Appendix 1   Reading Comprehension Test

Text 1

<center>**Ant intelligence**</center>

When we think of intelligent members of the animal kingdom, the creatures that spring immediately to mind are apes and monkeys. But in fact the social lives of some members of the insect kingdom are sufficiently complex to suggest more than a hint of intelligence. Among these the world of the ant has come in for considerable scrutiny lately, and the idea that ants demonstrate sparks of cognition has certainly not been rejected by those involved in these investigations.

Ants store food, repel attackers and use chemical signals to contact one another in case of visual and auditory channels (as in religious chants, advertising images and jingles, political slogans and martial music) to arouse and propagate moods and attitudes. The biologist Lewis Thomas wrote, "Ants are so much like human beings as to be an embarrassment. They farm fungi, raise aphids as livestock, launch armies to war, use chemical sprays to alarm and confuse enemies, capture slaves, engage in child labour, exchange information ceaselessly. They do everything but watch television." However in ants there is no cultural transmission — everything must be encoded in the genes — whereas in humans the opposite is true. Only basic instincts are carried in the genes of a newborn baby, other skills being learned from others in the community as the child grows up. It may seem that this cultural continuity gives us a huge advantage over ants. They have never mastered fire nor progressed. Their fungus farming and aphid herding crafts are sophisticated when compared to the agricultural skills of humans five thousand years ago but have been totally overtaken by modern human agribusiness.

Or have they? The farming methods of ants are at least sustainable. They do not ruin environments or use enormous amounts of energy. Moreover recent evidence suggests that the crop farming of ants may be more sophisticated and adaptable than was thought.

Ants were farmers fifty million years before humans were. Ants can't digest the cellulose in leaves — but some fungi can. The ants therefore cultivate these fungi in their nests, bringing them leaves to feed on, and then use them as a source of food. Farmer ants secrete antibiotics to control other fungi that might act as "weeds", and spread waste to fertilize the crop.

It was once thought that the fungus that ants cultivate was a single type that they had propagated, essentially unchanged from the distant past. Not so. Ulrich Mueller of Maryland and his colleagues genetically screened 862 different types of fungi taken from ants' nests. These turned out to be highly diverse: it seems that ants are continually domesticating new species. Even more impressively, DNA analysis of the fungi suggests that the ants improve or modify the fungi by regularly swapping and sharing stains with neighboring ant colonies.

When we survey Mexico City, Tokyo, Los Angeles, we are amazed at what has been

<center>154</center>

accomplished by humans. Yet Hoelldobler and Wilson's magnificent work for ant lovers, *The Ants*, describes a supercolony of the ant *Formica yessensis* on the Ishikari Coast of Hokkaido. This "megalopolis" was reported to be composed of 360 million workers and a million queens living in 4,500 interconnected nests across a territory of 2.7 square kilometers.

Such enduring and intricately meshed levels of technical achievement outstrip by far anything achieved by our distant ancestors. We hail as masterpieces the cave paintings in southern France and elsewhere, dating back some 20,000 years. Ant societies existed in something like their present form more than seventy million years ago. Beside this prehistoric man looks technologically primitive. Is this then some kind of intelligence, albeit of a different kind?

Research conducted at Oxford, Sussex and Zurich Universities has shown that when desert ants return from a foraging trip, they navigate by integrating bearings and distances, which they continuously update in their heads. They combine the evidence of visual landmarks with a mental library of local directions, all within a framework which is consulted and updated. So ants can learn too.

And in a twelve-year programme of work, Ryabko and Reznikova have found evidence that ants can transmit very complex messages. Scouts who had located food in a maze returned to mobilize their foraging teams. They engaged in contact sessions, at the end of which the scout was removed in order to observe what her team might do. Often the foragers proceeded to the exact spot in the maze where the food had been. Elaborate precautions were taken to prevent the foraging team using odour clues. Discussion now centers on whether the route through the maze is communicated as a "left-right" sequence of turns or as a "compass bearing and distance" message.

During the course of this exhaustive study, Raznikova has grown so attached to her laboratory ants that she feels she knows them as individuals — even without the paint spots used to mark them. It's no surprise that Edward Wilson, in this essay, "in the company of ants", advises readers who ask what to do with the ants in their kitchen to:" watch where you step. Be careful of little lives."

Do the following statements agree with the information given in reading passage 1? In boxes 1-6 on your answer sheet, write

True   ***if the statement agrees with the information***
False   ***if the statement contradicts the information***
Not given   ***if there is no information on this***

1. ants use the same channels of communication as human do
2. city life is one factor that encourages the development of intelligence
3. ants can build large cities more quickly than humans do
4. some ants can find their way by making calculations based on distance and position
5. in one experiment, foraging teams were able to use their sense of smell to find food
6. the essay," in the company of ants", explores ant communication

Complete the summary using the list of words, A-O, below. Write the correct letter, A-O, in boxes 7-13 on your answer sheet.

Ants as farmers

Ants have sophisticated methods of farming, including herding livestock and growing crops, which are in many ways similar to those used in human agriculture. The ants cultivate a large number of different species of edible fungi which convert 7 ……… into a form which they can digest. They use their own natural 8 ……… as weed-killers and also use unwanted materials as 9 ……… Genetic analysis shows they constantly upgrade these fungi by developing new species and by 10 ……… species with neighboring ant colonies. In fact, the farming methods of ants could be said to be more advanced than human agribusiness, since they use 11 ……… methods, they do not affect the 12 ……… and don't waste 13……….

A. aphids   B. agriculture   C. cellulose   D. exchanging   E. energy
F. fertilizers   G. food   H. fungi   I. growing   J. interbreeding
K. natural   L. other species   M. secretions   N. sustainable   O. environment

Text 2

Choose the correct heading for paragraphs B and D–F from the list of headings below. Write the correct number i–viii in boxes 1- 4 on your answer sheet.

List of headings

1. Paragraph B	i	Effects of irrigation on sedimentation
2. Paragraph D	ii	The danger of flooding the Cairo area
3. Paragraph E	iii	Causing pollution in the Mediterranean
4. Paragraph F	iv	Interrupting a natural process
	v	The threat to food production
	vi	Less valuable sediment than before
	vii	Egypt's disappearing coastline
	viii	Looking at the long-term impact

Disappearing Delta

A. The fertile land of the Nile delta is being eroded along Egypt's Mediterranean coast at an astounding rate, in some parts estimated at 100 meters per year. In the past, land scoured away from the coastline by the currents of the Mediterranean Sea used to be replaced by sediment brought down to the delta by the River Nile, but this is no longer happening.

B. Up to now, people have blamed this loss of delta land on the two large dams at Aswan in the south of Egypt, which hold back virtually all of the sediment that used to flow down the river. Before the dams were built, the Nile flowed freely, carrying huge quantities of sediment north from Africa's interior to be deposited on the Nile delta. This continued for 7,000 years, eventually covering a region of over 22,000 square kilometers with layers of fertile silt. Annual flooding bought in new, nutrient-rich soil to the delta region, replacing what had been washed away by the sea, and dispensing with the need for fertilizers in Egypt's richest food-growing area. But when the Aswan dams were constructed in the 20th century to provide electricity and irrigation, and to protect the huge population centre of Cairo and its surrounding areas from annual flooding and drought,

most of the sediment with its natural fertilizer accumulated up above the dam in the southern, upstream half of Lake Nasser, instead of passing down to the delta.

C. Now, however, there turns out to be more to the story. It appears that sediment-free water emerging from the Aswan dams picks up silt and sand as it erodes the river bed and banks on the 800-kilometre trip to Cairo. Daniel Jean Stanley of the Smithsonian Institute noticed that water samples taken in Cairo, just before the river enters the delta, indicated that river sometimes carries more than 850 grams of sediment per cubic meter of water --- almost half of what it carried before the dams were built." I'm ashamed to say that the significance of this didn't strike me until after I had read 50 or 60 studies," says Stanley in Marint Geology." There is still a lot of sediment coming into the delta, but virtually no sediment comes out into the Mediterranean to replenish the coastline. So this sediment must be trapped on the delta itself."

D. Once north of Cairo, most of the Nile water is diverted into more than 10,000 kilometers of irrigation canals and only a small proportion reaches the sea directly through the rivers in the delta. The water in the irrigation canals is still or very slow-moving and thus cannot carry sediment, Stanley explains. The sediment sinks to the bottom of the canals and then is added to fields by farmers or pumped with the water into the four large freshwater lagoons that are located near the outer edges of the delta. So very little of it actually reaches the coastline to replace what is being washed away by the Mediterranean currents.

E. The farms on the delta plains and fishing and aquaculture in the lagoons account for much of Egypt's food supply. But by the time the sediment has come to rest in the fields and lagoons it is laden with municipal, industrial and agricultural waste from the Cairo region, which is home to more than 40 million people." Pollutants are building up faster and faster," says Stanley. Based on his investigations of sediment from the delta lagoons, Frederic Siegel of George Washington University concurs." In Manzalah Lagoon, for example, the increase in mercury, lead, copper and zinc coincided with the building of the High Dam at Aswan, the availability of cheap electricity, and the development of cheap electricity, and the development of major power-based industries." He says. Since that time the concentration of mercury has increased significantly. Lead from engines that use leaded fuels and from other industrial sources has also increased dramatically. These poisons can easily enter the food chain, affecting the productivity of fishing and farming. Another problem is that agricultural wastes include fertilizers which stimulate increases in plant growth in the lagoons and upset the ecology the area, with serious effects on the fishing industry.

F. According to Siegel, international environmental organizations are beginning to pay closer attention to the region, partly because of the problems of erosion and pollution of the Nile delta, but principally because they fear the impact this situation could have on the whole Mediterranean coastal ecosystem. But there are no easy solutions. In the immediate future, Stanley believes that one solution would be to make artificial floods to flush out the delta waterways in the long term an alternative process such as desalination may have to be used to increase the amount of water available." In my view, Egypt must devise a way to have more water running through the river and the delta," says Stanley. Easier said than done in a desert region with a rapidly growing population.

Do the following statements reflect the claims of the writer in Reading Passage 2? In boxes 18-23 on your answer sheet, write.

Yes   *if the statement reflects the claims of the writer.*
No   *if the statement contradicts the claims of the writer.*
Not GIVEN if it is impossible to say what the writer thinks about this.

5. Coastal erosion occurred along Egypt's Mediterranean coast before the building of the Aswan dams.
6. Some people predicted that the Aswan dams would cause land loss before they were built.
7. The Aswan dams were built to increase the fertility of the Nile delta.
8. Stanley found that the levels of sediment in the river water in Cairo were relatively high.
9. Sediment in the irrigation canals on the Nile delta causes flooding.
10. Water is pumped from the irrigation canals into the lagoons.

Complete the summary of Paragraphs E and F with the list of words A-H below. Write the correct letter A-H in boxes 11-13 on your answer sheet.

In addition to the problem of coastal erosion, there has been a marked increase in the level of 11………contained in the silt deposited in the Nile delta. To deal with this, Stanley suggests the use of 12 ………in the short term, and increasing the amount of water available through 13..………in the longer term.

A. artificial floods   B. desalination   C. delta waterways   D. natural floods
    E. nutrients   F. pollutants   G. population control   H. sediment

Text 3
Choose the correct heading for each section from the list of headings below. Write the correct number, i-vi, in boxes 1-4 on your answer sheet.

List of Headings
1. Section A   i. Causes of volcanic eruption
2. Section B   ii. Efforts to predict volcanic eruption
3. Section C   iii. Volcanoes and the features of our planet
4. Section D   iv. Different types of volcanic eruption
                        v. International relief efforts
                        vi. The unpredictability of volcanic eruption

Volcanoes --- earth-shattering news

A.   Volcanoes are the ultimate earth-moving machinery. A violent eruption can blow the top few kilometers off a mountain, scatter fine ash practically all over the globe and hurl rock fragments into the stratosphere to darken the skies a continent away.

But the classic eruption --- cone-shaped mountain, big bang, mushroom cloud and surges of molten lava --- is only a tiny part of a global story. Vulcanism, the name given to volcanic processes, really has shaped the world. Eruptions have rifted continents, raised mountain chains, constructed islands and shaped the topography of the earth. The entire ocean floor has a basement of volcanic basalt.

Volcanoes have not only made the continents, they are also thought to have made the world' first stable atmosphere and provided all the water for the oceans, rivers and ice-caps. What comes out volcanic craters is mostly gas. More than 90% of this gas is water vapor from the deep earth: enough to explain, over 3500 million years, the water in the oceans. The rest of the gas is nitrogen, carbon dioxide, sulphur dioxide, methane, ammonia and hydrogen. The quantity of these gases, again multiplied over 3500 million years, is enough to explain the mass of the world's atmosphere. We are alive because volcanoes provided the soil, air and water we need.

B.   Geologists consider the earth as having a molten core, surrounded by a semi-molten mantle and a brittle, outer skin. It helps to think of a soft-boiled egg with a runny yolk, a firm but squishy white and a hard shell. If the shell is even slightly cracked during boiling, the white materials bubbles out and sets like a tiny mountain chain over the crack --- like an archipelago of volcanic islands such as the Hawaiian Islands. But the earth is so much bigger and the mantel below is so much hotter.

Even though the mantle rocks are kept solid by overlying pressure, they can still slowly flow like thick treacle. The flow, thought to be in the form of convection currents, is powerful enough to fracture the eggshell of the crust into plates, and keep them bumping and grinding against each other, or even overlapping, at the rate of a few centimeters a year. These fracture zones, where the collisions occur, are where earthquakes happen. And, very often, volcanoes.

C.   These zones are lines of weakness, or hot spots. Every eruption is different, but put at its simplest, where there are weaknesses, rocks deep in the mantle, heated to 1,350°C, will start to expand and rise. As they do so, the pressure drops, and they expand and become liquid and rise more swiftly.

Sometimes it is slow: vast bubbles of magma --- molten rock from the mantle --- inch towards the surface, cooling slowly, to show through as granite extrusions. Sometimes --- as in Northern Ireland, Wales and the Karoo in South Africa --- the magma rose faster, and then flowed out horizontally on to the surface in vast thick sheets.

Sometimes the magma moves very swiftly indeed. It does not have time to cool as it surges upwards. The gases trapped inside the boiling rock expand suddenly, the lava glows with heat, it begins to froth, and it explodes with tremendous force. Then the slightly cooler lava following it begins to flow over the lip of the crater. By studying the evidence, volcanologists can read the force of the great blasts of the past.

The biggest eruption are deep on the mid-ocean floor, where new lava is forcing the continents apart and widening the Atlantic by perhaps five centimeters a year. Look at maps of volcanoes, earthquakes and island chains like the Philippines and Japan, and you can see the rough outlines of what are called tectonic plates. The most dramatic of these is the Pacific 'ring of fire' where there have been the most violent explosions – Mount Pinatubo near Manila, Mount St Helen's in the Rockies and El Chichon in Mexico about a decade ago, not to mention world-shaping blasts like Krakatoa in the Sunda Straits in 1883.

D.   But volcanoes are not very predictable. That is because geological time is not like human time. During quiet periods, volcanoes cap themselves with their own lava by forming a powerful core from the molten rocks slopping over the rim of the crater; later the lava cools slowly into a huge, hard stable plug which blocks any further eruption until the pressure below becomes irresistible. In the case of Mount Pinatubo, this took 600 years.

Then, sometimes, with only a small warning, the mountain blows its top. It did this at Mont Pelee in Martinique at 7.49 am on 8 May, 1902. Of a town of 28,000, only two people survived. In 1815, a sudden blast removed the top, 1280 meters of Mount Tambora in Indonesia. The eruption was so fierce that dust thrown into the stratosphere darkened the skies, canceling the following summer in Europe and North America. Thousands starved as the harvests failed, after snow in June and frosts in August. Volcanoes are potentially world news, especially the quiet ones.

Answer the questions below using **NO MORE THREE WORDS AND/OR A NUMBER** from the passage for each answer.
5. What are the sections of the earth's crust, often associated with volcanic activity, called?
6. What is the name given to molten rock from the mantle?
7. What is the earthquake zone on the Pacific Ocean called?
8. For how many years did Mount Pinatubo remain inactive?

Choose **NO MORE THAN TWO WORDS** from the passage for each answer. Write your answers in boxes 9-13 on your answer sheet.

Volcanic eruptions have shaped the earth's land surface. They may also have produced the world's atmosphere and 9……… Eruptions occur when molten rocks from the earth's mantle rise and expand. When they become liquid, they move more quickly through cracks in the surface. There are different types of eruption. Sometimes the 10 ……… moves slowly and forms outcrops of granite on the earth's surface. When it moves more quickly it may flow out in thick horizontal sheets. Examples of this type of eruption can be found in Northern Ireland, Wales, South Africa and 11………. A third type of eruption occurs when the lava emerges very quickly and 12……… violently. This happens because the magma moves so suddenly that 13……… are emitted.

Text 4

A remarkable beetle

Some of the most remarkable beetles are the dung beetles, which spend almost their whole lives eating and breeding in dung.

More than 4000 species of these remarkable creatures have evolved and adapted to the world's different climates and the dung of its many animals. Australia's native dung beetles are scrub and woodland dwellers, specializing in coarse marsupial droppings and avoiding the soft cattle dung in which bush flies and buffalo flies breed.

In the early 1960s George Bornemissza, then a scientist at the Australian Government's premier research organization, the Commonwealth Scientific and industrial Research Organization (SCIRO), suggested that dung beetles should be introduced to Australia to control dung-breeding flies. Between 1968 and 1982, the CSIRO imported insects from about 50 different species of dung beetle, from Asia, Europe and Africa, aiming to match them to different climate zones in Australia. Of the 26 species that are known to have become successfully integrated into the local government, only one, an African species released in northern Australia, has reached its natural boundary.

Introducing dung beetles into a pasture is a simple process: approximately 1500 beetles are released, a handful at a time, into a fresh cow pats in the cow pasture. The

beetles immediately disappear beneath the pats digging and tunneling and, if they successfully adapt to their new environment, soon become a permanent, self-sustaining part of the local ecology. In time they multiply and within three or four years the benefits to the pasture are obvious.

Dung beetles work from the inside of the pat so they are sheltered from predators such as birds and foxes. Most species burrow into the soil and bury dung in tunnels directly underneath the pats, which are hollowed out form within. Some large species originating from France excavate tunnels to a depth of approximately 20 cm below the dung pat. These beetles make sausage-shaped brood chambers along the tunnels. The shallowest tunnels belong to a much smaller Spanish species that buries dung in cambers that hang like fruit from the branches of a pear tree. South African beetles dig narrow tunnels of approximately 20 cm below the surface of the pat. Some surface-dwelling beetles, including a South African species, cut perfectly-shaped balls from the pat, which are rolled away and attached to the bases of plants.

For maximum dung burial in spring, summer and autumn, farmers require a variety of species with overlapping periods of activity. In the cooler environments of the state of Victoria, the large French species (2.5 cms long) is matched with smaller (half this size), temperate-climate Spanish species. The former are slow to recover from the winter cold and produce only one or two generations of offspring from late spring until autumn. The latter, which multiply rapidly in early spring, produce two to five generations annually. The South African ball-rolling species, being a subtropical beetle, prefers the climate of northern and coastal New Wales where it commonly works with the South African tunneling species. In warmer climates, many species are active for longer periods of the year.

Dung beetles were initially introduced in the late 1960s with a view to controlling buffalo flies by removing the dung within a day or two and so preventing flies from breeding. However, other benefits have become evident. Once the beetle larvae have finished pupation, the residue is a first-rate source of fertilizer. The tunnels abandoned by the beetles provide excellent aeration and water channels for root systems. In addition, when the new generation of beetles has left the nest the abandoned burrows are an attractive habitat for soil-enriching earthworms. The digested dung in these burrows is an excellent food supply for the earthworms, which decompose it further to provide essential soil nutrients. If it were not for the dung beetle, chemical fertilizer and dung would be washed by rain into streams and rivers before it could be absorbed into the hard earth, polluting water courses and causing blooms of blue-green algae. Without the beetles to dispose of the dung, cow pats would litter pastures making grass inedible to cattle and depriving the soil of sunlight. Australia's 30 million cattle each produce 10-12 cow pats a day. This amounts to 1.7 billion tones a year, enough to smother about 110000 sq km of pasture, half the area of Victoria.

Dung beetles have become an integral part of the successful management of the dairy farm in Australia over the past few decades. A number of species are available from the CSIRO or through a small number of private breeders, most of whom ere entomologists with the CSIRO's dung beetle unit who have taken their specialized knowledge of the insect and opened small businesses in direct competition with their former employer.

Do the following statements reflect the claims of the writer in the passage? In boxes 1-5 on your answer sheet write:

Yes   *if the statement reflects the claims of the writer.*
No   *if the statement contradicts the claims of the writer.*
Not GIVEN   if it is impossible to say what the writer thinks about this.

1. Bush flies are easier to control than buffalo flies.
2. Four thousand species of dung beetle were initially brought to Australia by the CSIRO.
3. Dung beetles were brought to Australia by the CSIRO over a fourteen-year period.
4. At least twenty-six of the introduced species have become established in Australia.
5. The dung beetles cause an immediate improvement to the quality of a cow pasture.

Label the tunnels on the diagram below. Choose your labels from the box. Write your answers in boxes 6-8 on your answer sheet.

French
Spanish
South Africa
Mediterranean
Australian native
South African ball roller

0

8

6

7

30

☐ Tunnel depth
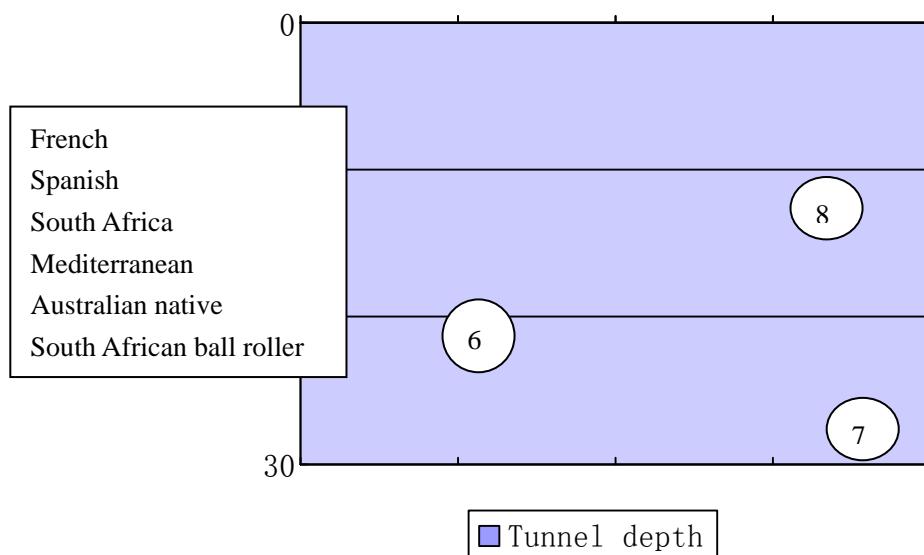
Appendix 2   Annotations and Examples

1. A list of the annotated words

Text 1	Text 2	Text 3	Text 4
fungi	delta	hurl	dung
aphid	delta	lava	marsupial
livestock	drought	basalt	pasture
fertilize	silt	mantle	burrow
swap	irrigation	archipelago	hollow
scout	canal	granite extrusion	larvae
propagate	replenish	treacle	residue
sparks	aquaculture	crater	algae
forage	pump	cone	entomologists
weed	sink	eruption	(cow) pat
maze	accumulate	squishy	predator
odor	sediment	bubble	pupation
bearing	lagoon	froth	smother

2. An example of topic annotation

[*Ant Intelligence*] Beetles are the group of insects which can be found in almost all habitats. They interact with their ecosystems in several ways. Some species are hunters of various animals including birds and mammals. Certain species are agricultural pests, while other species of beetles are important controls of agricultural pests. By burying and consuming dung, they improve nutrient recycling and soil structure. They also protect livestock, such as cattle, by removing the dung which, if left, could provide habitat for pests such as flies. Therefore, many countries have introduced the creatures for the benefit of animal husbandry.

163

3. A screen shot showing a verbal plus visual annotation

Appendix 3   Working Memory Test

(1) L1 Reading Span Task

Directions: This is a memory task. When the task begins, a sentence will appear in the center of the screen. ***You should read aloud the sentence, decide its sensibility, and try to remember the last word***. When the sentence disappears, put a tick if it makes sense or a cross if it does not, on the Answer Sheet. Then another sentence comes, and you repeat the same procedure. When you see "Please write down the words" on the screen, it signifies that this set of sentences is finished, and you should put the last word of each sentence on the Answer Sheet. If you cannot remember them in order, you can write them in any order, ***but you should not write the last word first***, unless it is the only one you can remember. We will be starting off with trials consisting of two sentences, but then the number of sentences per trial will change without any advance notice. Now you have a chance to practice some sentences. You can ask for another chance by raising your hand, until you feel quite confident for the formal test. ***Your serious and responsible participation is very important to the study and is highly appreciated.***
(*In the study, the instruction was given in Chinese*.)

<u>Practice</u>

Her loud shout made the two blackbirds flutter out of the tree.
The ocean shone brightly over the clear, blue sun.

The young woman sang the empty blue bottle.
I turned on the tap and splashed cold water on my face.
The below freezing temperatures caused the snowman to melt.

<u>Formal test</u>
The house quickly got dressed and went to work.
The cop spent a good half-hour questioning his trusted friend.

The castle sat in the refrigerator above the tiny village.
It wasn't all her fault that her marriage was in trouble.
When he reached the top of the heart, his mountain was beating.

As a full-time university student, he studied hard.

No one ever figured out what caused the crash to plane

His men now decided to refuse to continue with the journey.

As the ideas flowed, I wrote them down on some water.

Three of the beds were dead and he was next.

The red wine looked like blood on the white carpet.

The children put on their toilets and played in the snow.

I let the potato ring and ring, but still no answer.

He stood up slowly, stretching his arms above his head.

They talked about what the world would be like after the war.

He quickly put the carrot in the ignition and started the car.

The trousers set the table in the dining room, while I made dinner

The crowd stepped backwards, waiting for someone to pass through.

I took a bag from my knife, and began removing the earth.

The bar cried and began striking on the ape of the cage.

The young girl walked slowly along the wooden path.

The purpose of the course was to learn a new language.

The class homework was done by everyone in the history.

The men were all killed during a training flight near the base.

People are given by money at Christmas time.

She worked quickly but quietly while the others were asleep.

My escape out of the telephone was blocked by a fence.

His eyes were bloodshot and his face was pale.

She could not wait to go to the zoo to visit her cheese.

I waited for a long time, watching the loud silence.

The thought of going back in there made my skin swim.

The man was standing there nervously, once again checking his watch.

Clouds of cigar smoke fly into the open pencil.

He was pleased to receive so much love and attention.

Criminal offences have been increasing from the turn of the century

The story started as a joke, but soon got out of hand.

The lamp jumped, and sent the horse tumbling to the ground.

I could not believe he loved the oldest book in the trick.

Thick leaves surrounded him, and the air was heavy and still.

The wasted year passed sadly, driven by the wind and the tide.

The stove and oven stretched over the rapidly moving bridge.

The story started as a joke, but soon got out of hand.

Heavy rains and snows swept over the tiny deserted island.

He put the sandwich into the recorder and watched the movie.

Usually the visual images are the ones people remember best.

A blue-uniformed guard moved quickly out of the dog.

She wore a huge, white dress bigger than a camping tent.

They ran like the wind, but they would never get away.

The forest passed and the dead sound gained its quiet.

They ran until their lungs felt like they were going to burst.

As the flower talked about its busy life, it began to cry.

The sudden grizzly bear caused the noise to look in our direction.

The teeth were moving with music, theatre, and dance.

It was a foggy day and everything was dripping wet.

The girl was awakened by the strong wind blowing against the house.

The tower raced across the boat to the finish line.

Somewhere in the deep night, a man sang its evening song.

Women' roles remained the same as their roots are deep.

The umbrella grasped his bat and stepped up to the plate.

Then the sun had gone and the evening sky was purple.

(2) L2 Reading Span Task

Directions: This is a memory task. When the task begins, a sentence will appear in the

center of the screen. ***You should read aloud the sentence, decide its sensibility, and try to remember the <u>final noun phrase </u>in the quotation marks***. When the sentence disappears, put a tick if it makes sense or a cross if it does not on the Answer Sheet. Then another sentence comes, and you repeat the same procedure. When you see "Please write down the words" on the screen, it signifies that this set of sentences is finished, and you should put final noun phrases on the Answer Sheet. If you cannot remember them in order, you can write in any order, ***but you should not write the last noun phrase first***, unless it is the only one you can remember. We will be starting off with trials consisting of two sentences, but then the number of sentences per trial will change without any advance notice. Now you have a chance to practice some sentences. You can ask for another chance by raising your hand, until you feel quite confident for the formal test. ***Your serious and responsible participation is very important to the study and is highly appreciated.***

(*In the study*, *the instruction was given in Chinese*.)

<u>Practice</u>
诸葛亮是《水浒》里面的一个"人物"
蛋白质是人体所必须的营养"成分"。

熊猫最喜欢吃的食物是"白菜"。
猫头鹰习惯白天觅食晚上"睡眠"。
过度的肥胖不利于身体"健康"。

<u>Formal test</u>
春天是各种传染病流行的"季节"。
中国是处在南半球的一个"国家"。

光年是一种测量长度的"单位"。
爱因斯坦发现了万有引力"定律"。
苹果是一种营养丰富的"水果"。

我国实行九年义务教育"制度"。
张家界是中国著名的旅游"胜地"。
李自成领导了太平天国"运动"。
人的知识经验一生都不会"变化"。
冷空气和热空气相遇会形成"对流"。

绿叶能够利用阳光进行光合"作用"。
水分子是由氢原子和铁原子"组成"。
煤炭是一种不可再生的"资源"。
水和冰具有同样的化学"成分"。

适者生存是一种普遍的自然"规律"。
巴黎是美国的一个发达的"城市"
斜插入水的筷子会发生反射"现象"。
松花江是我国境内最长的"河流"。
声音可以通过空气进行"传导"。
英文字母和汉语拼音的读音"相同"。

大多数的花朵都是在春天"开放"。
智力测量可以测量人的"身高"。
鸟类的基本运动方式是"飞翔"。
铁片放在潮湿的空气中会"生锈"。

西藏是人口密度很大的一个"省份"。
《呐喊》是矛盾的一部代表"作品"。

改革开放是我国的基本"国策"。
人体吸入二氧化碳呼出"氧气"。
地震发生时在空旷地比较"安全"。
艾滋病被认为是一种遗传"疾病"。
蘑菇是一种有利于健康的"食品"。

绿色蔬菜中的纤维可以促进"消化"。
印地安人是典型的黄色"人种"
火山喷发是一种壮丽的自然"景观"。

遇热膨胀是物体的一种"属性"。
植物的叶子总是朝向有光的"方向"。
太阳光是由七种颜色的光"组成"。
两条平行线可以有一个"交点"。

突然发出的喊声会引起人的"注意"。

蝙蝠依靠眼睛判断物体的"位置"。

目前我国属于发展中的"国家"。

语言是人类交流的一种"手段"。

鸦片战争是我国近代史的"开端"。

和平与发展是世界的两大"主题"。

汽油机动车使用的主要"燃料"。

我国一共有五十个少数"民族"。

明朝是中国的最后一个"朝代"。

锌是人体必须的一种微量"元素"。

新加坡是世界上人口最多的"国家"。

北极星是天空南方的一颗"恒星"。

奥地利是世界著名的音乐"王国"

米饭和馒头中含有大量的"淀粉"。

人体吸入二氧化碳呼出"氧气"。

海水是一种可直接饮用的"液体"。

沙尘暴的波及范围正在"增大"。

大多数人对自己的名字较"敏感"。

正月十五是中国的元宵"佳节"

西部是我国目前重点建设的"地区"。

地震发生时在空旷地比较"安全"。

两点连接起来可以形成一条"线段"

(3) Operation Span Task

Directions: In this task, you will be presented with a series of mathematical equations, each followed by a word. Whenever an equation is presented to you, *you are to judge whether the stated solution is correct or not, and try to remember the word that follows*. When the equation disappears, put a tick if the solution is correct, or a cross if not, on the Answer Sheet. Keep doing this until you see "Please write down the words" on the screen, which signifies that this set of equations is over, and you write down the words in the order in

which they were presented. If you cannot remember them in order, you can write them in any order, ***but you should not write the last word first***, unless it is the only one you can remember. The number of equations in each set will not be informed in advance. We will be starting off with trials consisting of two equations, but then the number of equations per trial will change without any advance notice. Now you have a chance for practice. You can ask for another chance by raising your hand, until you feel quite confident for the formal test. ***Your serious and responsible participation is very important to the study and is highly appreciated***.

(*In the study, the instruction was given in Chinese.*)

Practice

$(2 \times 7) + 6 = 20$	good
$(6 \div 3) - 2 = 0$	wall
$(7 \times 5) - 8 = 27$	lake

$(5 \times 2) - 6 = 4$	goat
$(9 \div 3) + 6 = 8$	rice
$(5 \times 6) + 7 = 37$	card
$(6 \div 2) - 4 = -1$	video

Formal test

$(3 \times 6) - 5 = 12$	bottle
$(9 \div 3) + 4 = 7$	supper

$(2 \times 8) + 4 = 20$	music
$(6 \div 3) - 9 = -7$	foot
$(7 \times 6) - 4 = 38$	color

$(8 \div 2) - 4 = 1$	pencil
$(5 \times 2) + 9 = 19$	spring
$(6 \times 6) - 7 = 29$	time
$(4 \div 2) + 7 = 9$	center
$(5 \div 2) + 2 = 2$	police

$(5 \times 3) - 5 = 10$	room

$(9 \div 9) + 5 = 6$	milk
$(5 \times 3) + 9 = 24$	winter
$(6 \div 2) - 7 = -4$	light
$(2 \times 5) - 3 = 7$	help
$(4 \div 2) - 1 = 1$	horse
$(7 \times 6) + 5 = 48$	game
$(6 \div 2) + 8 = 10$	school
$(8 \div 4) - 5 = -2$	fire
$(9 \div 3) + 4 = 6$	bread
$(6 \times 9) + 8 = 63$	house
$(8 \div 4) + 8 = 13$	garden
$(6 \times 4) - 2 = 21$	road
$(4 \times 5) + 7 = 37$	bike
$(8 \div 1) - 5 = 3$	money
$(8 \times 6) + 4 = 37$	book
$(8 \div 4) - 3 = -1$	snow
$(4 \times 3) - 5 = 7$	town
$(4 \div 4) + 6 = 10$	plane
$(3 \times 6) + 6 = 24$	world
$(5 \times 4) - 6 = 14$	stair
$(4 \div 1) + 9 = 13$	moon
$(9 \times 3) + 2 = 29$	ball
$(6 \div 6) - 4 = -2$	hour
$(4 \times 4) + 8 = 25$	water
$(4 \div 2) - 6 = 8$	window
$(8 \div 2) + 7 = 11$	office
$(7 \times 3) - 7 = 15$	candy
$(6 \div 6) + 9 = 10$	forest
$(8 \times 5) - 6 = 34$	party

$(3 \times 9) - 8 = 16$ summer

$(7 \times 7) + 6 = 55$ work

$(6 \times 8) + 7 = 55$ river

$(8 \times 2) - 7 = 9$ door

$(6 \div 4) + 5 = 7$ flower

$(3 \div 1) - 8 = 5$ farm

$(4 \times 7) - 9 = 19$ night

$(4 \div 1) - 6 = 2$ bird

$(3 \times 6) + 9 = 26$ walk

$(5 \times 7) - 8 = 26$ street

$(8 \div 8) + 8 = 9$ table

$(9 \div 3) - 2 = 1$ friend

$(3 \times 7) + 5 = 28$ park

$(8 \div 4) + 6 = 9$ bank

$(2 \times 9) - 3 = 15$ tree

$(6 \div 3) + 9 = 12$ person

$(8 \times 4) + 3 = 36$ wind

$(6 \div 3) - 5 = -2$ year

$(9 \times 2) - 4 = 13$ paper

$(8 \div 2) + 4 = 8$ ship

Appendix 4   CALL Reading Questionnaire

(In the study, the questionnaire was written in Chinese)

Part One   Please fill in the table by writing in the spaces.

Reference No.		Gender	
Age		School/Department	
Years of learning English			

Part 2   Please indicate to what degree you agree with the following statements. Be careful with ones with an asterisk.

 (*1 = strongly disagree; 2 = disagree; 3 = undecided; 4 = agree; 5 = strongly agree*)

1. The annotations in the reading texts help reading comprehension.

2. The topic annotations in the reading texts are necessary.

3. The visual annotations for new words in the reading texts are useful for reading comprehension.

4. The use of annotations helps me to concentrate on reading.

5. The visual plus verbal annotations are helpful to reading comprehension.

6. I like reading a text with annotations more than one without.

7. The verbal annotations for new words assist the reading process.

8. I find annotations distract my attention from reading.*

9. The visual annotations for topics are helpful to reading comprehension.

10. I prefer reading a text both verbally and visually annotated than one annotated in a single way.

11. Usually I will turn to the annotations only when other methods fail.

12. Multimedia reading and paper reading make no difference. *

13. I like reading texts with verbal annotations only.

14. The verbal annotations in the reading texts are easy to understand.

15. The dependence on the annotations will lead to a decline of my reading ability.

16. The verbal annotations in the reading texts are difficult to understand.*

17. The visual annotations in the reading texts are necessary to reading comprehension.

18. The verbal plus visual annotations in the reading texts render negative effects on reading comprehension.*

19. The word annotations provided to the reading texts are necessary.

20. I like reading texts with visual annotations only.

21. The verbal plus visual annotations in the texts are not beneficial to reading comprehension.*

22. The verbal annotations in the reading texts are necessary.

23. Multimedia reading is different from paper reading.

24. I do not know what the pictures alone are meant to explain.*

25. I prefer both verbal and visual annotations to be presented at the same time.

26. I would rather read a non-annotated text than an annotated text. *

27. The annotations in the reading texts do not facilitate the reading comprehension.*

28. The verbal annotations for topics of the reading texts help in understanding the texts.

29. The pictures in the reading texts are easy to understand.

30. I will read both the verbal and visual annotations when they are presented at the same time even if either is unnecessary.

Appendix 5   Interview Questions

(In the study, the interviews were conducted in Chinese)

1. How did you deal with any difficulty in reading before the CALL reading program was introduced?

2. How do you usually use annotations while reading?

3. Does annotation use have any effects on your reading? Why or why not?

4. Which annotation type do you prefer, and which not? Why or why not?

5. Is there a particular annotation type that you feel is more useful than another? Why?

6. How do you find the design of the annotations? Please give an example.

7. Do you think the CALL reading software can be improved in any way? If so, in what way?

8. Is there anything that you want to add concerning CALL reading?

LIST OF REFERENCES

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning, 21*(3), 199 - 226.

Abu-Rabia, S. (2003). The Influence of working memory on reading and creative writing processes in a second language. *Educational Psychology, 23*(2), 209-222.

Abuseileek, A. F. M. (2008). Hypermedia annotation presentation: Learners' hypermedia annotation preferences and effect on EFL reading comprehension and vocabulary acquisition. *CALICO Journal, 25*(2), 260-275.

Abuseileek, A. F. M. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computer & Education* (57), 1281-1291.

Akbulut, Y. (2008). Predictors of foreign language reading comprehension in hypermedia reading environment. *Journal of Educational Computing Research 39*(1), 37-50.

Akyel, A., & Ercetin, G. (2009). Hypermedia reading strategies employed by advanced learners of English. *System, 37*, 136-152.

Al-Seghayer, K. (2001). The effectiveness of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning and Technology, 5*(1), 202-232.

Al-Seghayer, K. (2005). The Effects of Verbal and Spatial Abilities on Reading Comprehension Task Performance in Multimedia Environments with Respect to Individual Differences among Learners. *CALL-EJ Online, 7*(1).

Al-shehri, S., & Gitsaki, C. (2010). Online reading: a preliminary study of the impact of integrated and split-attention formats on L2 students' cognitive load. *ReCALL, 22*(3), 356-375.

Alderson, J. C. (1984). Reading in a Foreign Language: a Reading Problem or a Language Problem. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a Foreign Language*. London: Longman.

Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.

Alderson, J. C., & Urquhart, A. (1988). This test is unfair: I'm not an economist. In P. Carrell, J. Devine & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 168-182). Cambridge: Cambridge University Press.

Alptekin, C., & Ercetin, G. (2009). Assessing the relationship of working memory to L2 reading: Does the nature of comprehension process and reading span task make a difference? *System, 37*, 627-639.

Anderson, C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research Reviews* (pp. 77-117). Newmark: International reading Association.

Anderson, R. C. (1985). Role of the reader's schema in comprehension, learning, and memory. In H. Sinder & R. Ruddell (Eds.), *Theoretical models and processes of reading. 3rd ed*. Newark, DE: International Reading Association.

Anderson, R. C., & Pearson, P. D. (1984). A Schema-theoretic view of basic processing in reading. In P. D. Pearson, R. Barr, M. L. Kamil & P. Mosenthal (Eds.), *Handbook of reading research*. White Plains, NY: Longman.

Andreassen, J., & Braten, I. (2007). *Working memory influences on reading comprehension assessed with different measures*. Paper presented at The 12th EARLI Biennial Conference.

Ariew, R. (2006). A template to generate hypertext and hypermedia reading materials: its design and associated research findings. *The Reading Matrix, 6* (3).

Ariew, R., & Ercetin, G. (2004). Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning, 17*(2), 237-259.

Arnold, N., & Ducate, L. (2006). CALL: Where are we and where do we go from here? *CALICO Monograph Series, 5*, 1-20.

Ashworth, D. (1996). Hypermedia and CALL. In M. C. Pennington (Ed.), *The Power of CALL*. Hong Kong: Athelstan Publications.

Aust, R., Kelley, M. J., & Roby, W. (1993). The use of hyper-reference and conventional

dictionaries. *Educational Technology Research and Development, 41*(4), 63-73.

Babaie, H. (2010). Image modalities in multimedia vocabulary instruction: Does dual coding of vocabulary occur at different degrees? *Journal of English Language Studies, 1*(2), 83-102.

Baddeley, A. (1986). *Working Memory*. Oxford: Oxford University Press.

Baddeley, A. (1992). Working Memory. *Science, 255*, 556-559.

Baddeley, A. (2000). The episodic buffer: A new component of working memory. *Trends in Cognitive Sciences, 4*, 417-423.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Baddeley, A. (2007). *Working Memory, Thought, and Action*. Oxford: Oxford University Press.

Baddeley, A. (2010). Working memory. *Current Biology, 20*(4), 136-140.

Baddeley, A., & Hitch, G. (1974). Working memory In G. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A., & Logie, R. H. (1999). Working memory: the multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.

Barnett, M. A. (1988). Reading through Context: How Real and Perceived Strategy Use Affects L2 Comprehension. *The Modern Language Journal, 72*, ii.

Beaty, K. (2005). *Computer-Assisted Language Learning*. Beijing: Foreign Language Teaching and Research Press.

Bell, T. (2005). Behaviors and attitudes of effective foreign language teachers: Results of a questionnaire study. *Foreign Language Annals, 38*(2), 260-270.

Bensoussan, M. (1998). Schema effects in EFL reading comprehension. *Journal of*

Research in Reading, 21, 213-227.

Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical and classroom perspectives*. NJ: Ablex Publishing.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics, 16*(1), 15-34.

Berquist, B. (1997). Memory models applied to L2 comprehension: a search for common ground. In G. Taillefer & A. K. Pugh (Eds.), *Reading in the University: First, second and foreign languages* (pp. 29-44). Toulouse: Presses de L'Universite des Sciences Sociales.

Betrancourt, M., & Bisseret, A. (1998). Integrating textual and pictorial information via pop-up windows: An experimental study. *Behavior and Information Technology, 17*, 263-273.

Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Hispania, 87*(3), 541-552.

Braul, b. (2006). ESL Teacher Perceptions and Attitudes toward Using Computer-Assisted Language Learning: Recommendations for Effective CALL Practice. Unpublished PHD dissertation. University of Alberta.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

Budd, D., Whitney, P., & Turley, K. J. (1995). Individual differences in working memory strategies for reading expository text. *Memory and Cognition, 23*(6), 735-748.

Butler, Y., & Hakuta, K. (2006). Cognitive factors in children's L1 and L2 reading. *Academic Exchange Quarterly*.

Cantor, J., & Engle, R. (1993). Working memory capacity as a long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1101-1114.

Carnevale, D. (2004). Educational technology fails to deliver on its promises, report says. *Cited in Murday, Kimmaree, Ushida, Eiko and Chenoweth, N. Ann (2008)*

Learners' and teachers' perspectives on language online. Computer Assisted Language Learning, 21(2), 125 -142.

Carpenter, P., & Just, M. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 120-150). Hillsdale: Erlbaum.

Carrell, P. L. (1983). Some issues in studying the role of schemata or background knowledge in second language comprehension. *Journal of Reading in A Foreign Language, 1*(2), 81-92.

Carrell, P. L. (1984). Schema theory and ESL reading: classroom implications and applications. *The Modern Language Journal, 68*(332-343).

Carrell, P. L. (1988). Some causes of text-boundedness and schema interference in ESL reading. In P. L. Carrell, J. Devine & D. E. Eskey (Eds.), *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press.

Chall, J. S. (1987). Two vocabularies for reading: Recognition and meaning. *The nature of vocabulary acquisition*, 7-17.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction 8*(4), 293-332.

Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *The Modern Language Journal, 81*, 209-227.

Chen, S. (2002). A cognitive model for non-linear learning in hypermedia programs. *British Journal of Educational Technology, 33*(4), 449-460.

Chen, S. Q., & Fan, L. (2007). Compensation between Chinese Reading Ability and English Language Proficiency in English Reading. *Journal of Zhejiang University of Science and Technology, 1*(19), 56-59.

Chen, Z. (2006). 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. Unpublished PhD dissertation. University of South Florida.

Chun, D. M. (2001). L2 reading on the Web: Strategies for accessing information in hypermedia. *Computer Assisted Language Learning, 14*(5), 367-403.

Chun, D. M. (2006). CALL technologies for L2 reading. *CALICO Monograph Series, 5*(69-98).

Chun, D. M., & Payne, J. L. (2004). What makes students click: working memory and look-up behavior. *System, 32*, 481-503.

Chun, D. M., & Plass, J. L. (1996a). Facilitating reading comprehension with multimedia. *System, 24*(4), 503-519.

Chun, D. M., & Plass, J. L. (1996b). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*(2), 183-198.

Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology, 1*(1), 60-81.

Chung, W. M. N. (2008). Attention control, working memory, and second language experience in second language performance. Unpublished PhD dissertation. Concordia University.

Clarke, M. A. (1979). Reading in Spanish and English: evidence from adult ESL students. *Language Learning, 29*, 121-150.

Clarke, M. A. (1988). The short circuit hypothesis of ESL reading -- when language competence interferes with reading performance. In P. L. Carrell, J. Devine & D. E. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 114-125). Cambridge: Cambridge University Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769-786.

Cooledge, S. (2004). Interpreting a Second Language Text: A Study of L2 Reading Behavior and Comprehension in a Hypertext Environment. Unpublished PHD

dissertation. The University of Arizona.

Cummins, J. (1979). Cognitive/academic language proficiency. Linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism, 19*, 197-205.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Carpenter, P. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 561-584.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1-18.

Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology, 130*, 208-223.

Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomicl Bulletin & Review, 3*, 422-433.

Davis, N. (1989). Facilitating effects of marginal glosses on foreign language reading. *The Modern Language Journal, 71*(1), 41-48.

Davis, N., & Lyman-Hager, M. (1997). Computer and L2 reading: Student performance, student abilities. *Foreign Language Annals, 30*(1), 58-72.

De Ridder, I. (2000). Are we conditioned to followed links? Highlights in CALL materials and their impacts on the reading process. *Computer Assisted Language Learning, 13*(2), 183-195.

De Ridder, I. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language Learning and Technology, 6*(1), 123-146.

Deanel, P., Sheehan, K., Sabatini, J., Futagi, Y., & Kostin, I. (2006). Differences in text structure and its implications for assessment of struggling readers In F. Manis (Ed.),

The Official Journal of the Society for the Scientific Study of Reading (pp. 241-255).

Debski, R., & Gruba, P. (1999). A Qualitative Survey of Tertiary Instructor Attitudes towards Project-Based CALL. *Computer Assisted Language Learning, 12*(3), 219-239.

DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*, 1616-1641.

Dornyei, Z. (2005). *The Psychology of the Language Learner: Individual differences in second language acquisition* Mahwah, N.J.: L. Erlbaum Associates.

Ellis, R., & Kurniawan, S. (2000). Increasing the usability of online information for older users: A case study in participatory design. *International Journal of Human Computer Interaction, 12*(2), 263-276.

Engle, R., Cantor, J., & Carullo, J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 972-992.

Ercetin, G. (2003). Exploring ESL Learners' use of Hypermedia Reading Glosses. *CALICO Journal, 20*(2), 261-283.

Eskey, D. E. (1988). Holding in the Bottom: An interactive approach to the language problems of second language readers. In P. L. Carrell, J. Devine & D. E. Eskey (Eds.), *Interactive Approaches to Second language Reading*. Cambridge: Cambridge University Press.

Eskey, D. E. (2002). Reading and the Teaching of L2 Reading. *TESEL Journal, 11*(1), 5-9.

Feinberg, S., & Murphy, M. (2000). Applying cognitive load theory to the design of web-based instruction. *Technology & Teamwork*, 353-360.

Field, A. (2005). *Discovering Statistics Using SPSS* (2nd ed.). London: Sage Publications.

Fontanini, I., & Tomitch, L. (2009). Working memory capacity and university students' comprehension of linear texts and hypertexts *International Journal of English Studies, 9*(2), 1-18.

Fotos, S., & Browne, C. (2004). The development of CALL and current opinions. In S. Fotos & C. Browne (Eds.), *New Perspectives on CALL for Second Language Classrooms* (pp. 1-14). Mahwah, NJ: Lawrence Erlbaum

Friedman, D., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*, 136-158.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, 37*, 581-590.

Gathercole, S. (2006). Non-word repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513-543.

Gathercole, S., & Baddeley, A. (1993). *Working Memory and Language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goodman, K. (1967). Reading: A psycholinguistic guessing game. In F. K. Goolasch (Ed.), *Language and Literacy: The Collected Writings of Kenneth S. Goodman* (Vol. 1). London: Routledge.

Goodman, K. (1987). Reading as A Psycholinguistic Process. Unpublished lecture, cited from C. Wallace, 1992.

Gough, P. B. (1972). One second of reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by Ear and Eye* (pp. 331-358). Cambridge, MA: MIT press.

Grabe, W. (1991). Current development in second language reading research. *TESEL Quarterly, 375-406*(3), 2-33.

Grabe, W. (2004). Research on Teaching Reading. *Annual Review of Applied Linguistics, 24*, 44-69.

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading*. Harlow: Pearson Education

Grimley, M., & Banner, G. (2008). Working memory, cognitive style, and behavioral predictors of GCSE exam success. *Educational Psychology, 28*(3), 341-351.

Guan. (2007). The activation and long-term memory of predictive inferences: The role of working memory constraint and text elaboration. Unpublished PhD dissertation. The Florida State University.

Hambrick, D. Z., & Engle, R. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*, 339-387.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*, 103-128.

Harrington, M. (1992). Working memory capacity as a constraint on L2 development. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals*. Amsterdam: Elsevier.

Harrington, M., & Sawyer, M. (1990). Working memory in L2 reading: Does capacity predict performance? *ERIC, NO: ED320446*.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.

Hong, Y. (2007). The relationship between perception, memory and reading. Unpublished MA thesis. Capital Normal University. China.

Hwang, W.-Y., Wang, C.-Y., & Sharples, M. (2007). A study of multimedia annotation of Web-based materials. *Computers and Education, 48*, 680-699.

Jacobs, G., Dufon, P., & Hong, F. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effects for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17*(1), 19-28.

Ji, L. (2003). Listening activities for effective top-down processing. *The Internet TESL Journal 9*.

Jobst, J., & Dillon, A. (2005). Multimedia learning with hypermedia. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 569-588). Cambridge: Cambridge University Press.

Jonassen, D. H. (1996). *Computers in the classroom: Mindtools for critical thinking*.

Englewood, NJ: Prentice Hall

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*(2), 199-225.

Juffs, A. (2006). Working memory, second language acquisition and low-educated second language and literacy learners. In I. van de Craats, J. Kuvers & M. Young-Scholten (Eds.), *Low educated adult second language learners and literacy acquisition. LOT Occasional Papers* (pp. 89-105): Netherlands Graduate School of Linguistics.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(1), 137-166.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction *Applied Cognitive Psychology, 13*, 351-371.

King, J., & Just, M. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30, 580-602.*

Kirby, J. R. (1993). Collaborative and competitive effects of verbal and spatial processes. Learning and Instruction, 3(201-214)
.

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1-10.

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal, 78*(3), 285-299.

Kobayashi, M. (1995). Effects of text organization and test format on reading comprehension test performance. Unpublished PhD dissertation. Thames Valley University.

Kobayashi, M. (2002). Method Effects on Reading Comprehension Test Performance: Text Organization and Response Format. *Language Testing, 19*(2), 193-220.

Kong, A. (2006). Connections between L1 and L2 readings: Reading strategies used by four Chinese adult readers. *The Reading Matrix, 6*(2), 19-44.

Kost, C., Foss, P., & Lenzini, J. (1999). Textual and pictorial gloss: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32*(1), 89-113.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*.

La Pointe, L. B., & Engle, R. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*(16), 1118-1133.

Lee-Sammons, W., & Whitney, P. (1991). Reading perspectives and memory for text: an individual differences analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 1074-1081.

Lee, M., & Tedder, M. (2003). The effects of three different computer texts on readers' recall: based on working memory capacity. *Computer in Human Behavior, 19*, 767-783.

Leslie, L., & Caldwell, J. (2009). Formal and informal measures of reading comprehension. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 403-427). New York, NJ: Routledge.

Levy, M. (1997). Computer-assisted Language Learning: Context and Conceptualization. *Oxford, Oxford University Press*.

Li, L. H. (2005). On the preference representation of college students of verbal and imagery cognitive style in memory system. Unpublished PHD dissertation. Jilin University.

Liu, Y. (1993). The effects of prior knowledge and vocabulary knowledge on reading comprehension process and some instructional implications. *Journal of Guangxi Normal University, 29*(4).

Lomicka, L. L. (1998). To gloss or not to gloss: An investigation of reading comprehension online. *Language Learning & Technology, 1*(2), 41-50.

Lusting, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*, 199-207.

Lyman-Hager, M., & Davis, J. F. (1996). The case for computer-mediated reading: Une vide de boy. *The French Review, 69*(5), 775-790.

Marshall, C. C. (1997). Annotation: from paper books to the digital library. from http://www.csdl.tamu.edu/~marshall/dl97.pdf

Martinez-Lage, A. (1997). Hypermedia Technology for Teaching Reading. In D. Michael & T. Robert M (Eds.), *Technology Enhanced Language Learning* (pp. 121-163). Lincolnwood, IL: National Textbook Company.

Marzban, A. (2011). Investigating the role of multimedia annotations in EFL reading comprehension. *Social and Behavioral Sciences, 28*, 72-77.

Massa, L. J., & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-imager cognitive style? *Learning and Individual Differences, 16*, 321-335.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition, 27*, 759-767.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions. *Educational Psychologist, 32*(1), 1-19.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31-48). New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. C. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490.

McCormick, T. (1988). *Theories of Reading in Dialogue: An interdisciplinary study*. New York: University Press of America.

Milim, k., & Shen, H. Z. (2006). Integration of computers into an EFL reading classroom. *ReCALL, 18*(2), 212-219.

Miyake, A., & Friedman, D. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic studies on training and retention*. Mahwah, NJ: Lawrence Erlbaum Associates.

Miyake, A., & Shah, P. (1999). *Models of working memory: mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology, 39*, 1-20.

Moreno, R., & Mayer, R. E. (1999). Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals, 32*(4), 469-479.

Nambiar, R. (2009). Cross linguistic transfer between L1 and L2 texts: Learning strategies used by bilingual Malay Tertiary learners. *European Journal of Social Sciences, 7*(3), 115-125.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House Publishers.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing* (pp. 3-13). Amsterdam: John Benjamins.

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts. http://www.victoria.ac.nz/lals/about/staff/paul-nation.

Newell, A. (1973). Production systems: Models of control structure. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Nicholson, T. (1984). *The Process of Reading*. Sydney: Martin Educational.

Nielsen, J. (1995). *Multimedia and Hypertext: The Internet and Beyond*. Boston: Morgan Kaufmann Publishers.

Numminen, H. (2002). Working memory in adults with intellectual disability. *Famr, Research Publications, 85*.

Oberauer, K., & Sub, H. (2000). Working memory and interference: A comment on Jenkins, Myerson, Hale, and Fry (1990). *Psychonomic Bulletin & Review* (7), 727-733.

Oberauer, K., Weidendeld, A., & Hornig, R. (2006). Working memory capacity and the construction of spatial mental models in comprehension and deductive reasoning. *The Quarterly Journal of Experimental Psychology, 59*(2), 426-447.

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society, 3*(4), 287-289.

Osamu, I. (2006). L1 and L2 working memory: An investigation into the domain specificity and processing efficiency issues. *Bulletin of the Faculty of Education of Ehime University, 53*(1), 113-121.

Overstreet, C. (2006). Reading authentic text in the hypermedia environment: The effects of question glosses on comprehension processes of intermediate learners of German as a foreign language. Unpublished PhD thesis. University of Florida.

Oxford, R., & Crookall, D. (1990). Vocabulary learning: A critical analysis of techniques. *TESL Canada Journal, 7*(2), 9-30.

Paas, F., Renkel, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Renkel, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science 32*, 1-8.

Paivio, A. (1990). *Mental representations: A dual coding approach*. NY: Oxford University Press.

Patterson, N. (2000). Hypertext and the changing roles of readers. *English Journal, 90*, 74-80.

Penkl, A., & Atkinson, R. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15-22.

Pennington, M. C. (Ed.). (1996). *The Power of CALL*. Hong Kong: Athelstan Publications.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second language multimedia learning environment. *journal of Educational Psychology, 90*(1), 25-36.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*(2), 221-243.

Potter, M. C., So, K., Eckardt, V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23-38.

Preece, J. (1993). Hypermedia, multimedia and human factors. In C. Latchem, J. Williamson & L. Henderson-Lancett (Eds.), *Interactive Multimedia: Practice and Promise* (pp. 135-150). London: Kogan Page.

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language learning, 53*(2), 233-284.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning, 54*(3), 469-523.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Qian, J. C. (2007). Constructivism and multimedia learning. *Journal of North China Institute of Water Conservancy and Hydroelectric Power (Social Sciences Edition), 23*(4), 126-128.

Qian, J. C. (2011). Cognitive style and reading comprehension. *Teaching and Management*(24), 61-63.

Richards, J., & Rodgers, T. (2001). *Applications and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Rickman, J., & Grudzinski, M. (2000). Student expectations of information technology use in the classroom. *Educause Quarterly*, 1(1), 24-30.

Riding, R., Grimley, M., Dahraei, H., & Banner, G. (2001). Working memory, cognitive style and academic attainment. In R. Nata (Ed.), *Progress in education* (Vol. 5, pp. 1-19). New York: Nova Science Publishers Inc.

Riding, R., Grimley, M., Dahraei, H., & Banner, G. (2003). Cognitive style, working memory and learning behavior and attainment in school subjects. *British Journal of Educational Psychology, 73*(2), 149-169.

Riding, R., & Rayner, S. (1998). *Cognitive styles and learning strategies*. London: Fulton.

Roblyer, M. D., Edwards, J., & Havriluk, M. A. (1997). *Integrating Educational Technology into Teaching*. Englewood, NJ: Prentice Hall.

Roby, W. B. (1999). What's in a gloss? *Language Learning & Technology 2*(2), 94-101.

Rouet, J.-F., & Levonen, J. J. (1996). Studying and learning with hypertext: Empirical studies and their implications. In J. F. Rouet, J. J. Levonen, A. Dillon & R. J. Spiro (Eds.), *Hypertext and cognition* (pp. 9-23). Mahwah, NJ: Lawrence Erlbaum.

Rowland, S. (2008). Applying Instructional Design Principles to Web Site Design. www.westga.edu/.../Report1_Stacey_Rowland_ABED4118.doc

Rumelhart, D. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and Performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rumelhart, D. (1980). The building blocks of cognition. In R. J. Spiro, B. C. Bruce & W. F.

Brewer (Eds.), *Theoretical Issues in Reading Comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rumelhart, D., David, E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro & W. W. Montague (Eds.), *Schooling and the Acquisition of Knowledge* (pp. 99-135). Hillsdale, NJ: Lawrence Erlbaum Associates.

Russell, M., Bebell, D., O'Dwyer, L., & O'Connor, K. (2003). Examining teacher technology use: Implications for preservice and inservice teacher preparation. *Journal of Teacher Education, 54*(4), 297-310.

Sakar, A., & Ercetin, G. (2005). Effectiveness of hypermedia annotations for foreign language reading. *Journal of Computer Assisted Learning, 21*, 28-38.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schrodt, M., & Turman, P. (2005). The impact of instructional technology use, course design, and sex differences in students' initial perceptions of instructor credibility. *Communication Quarterly*, 53(2), 177-196.

Schwienhorst, K. (2000). Book Reviews: CALL environments: research, practice, and critical issues. *System, 28*, 447-465.

Seng, G. (2006). Use of L1 in L2 reading comprehension among tertiary ESL learners. *Reading in a Foreign Language, 18*(1).

Singhl, M. (1998). A Comparison of L1 and L2 Reading: Cultural Differences and Schema. *The Internet TESL Journal, 4*(10).

Smith, F. (2004). *Understanding Reading (6th ed)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Spodark, E. (2005). Technoconstructivism for the undergraduate foreign language classroom. *Foreign Language Annals, 38*(3), 428-435.

Stern, H. (1983). *Fundamental Concepts of Language Teaching*. London: Oxford University Press.

Strehler, A. (2008). The relationship between cognitive load, cognitive style and multimedia learning. Unpublished PHD dissertation. University of Pretoria.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

Taillefer, G. F. (1996). L2 reading ability: Further insight into the short-circuit hypothesis. *Modern Language Journal, 80*(4), 461-477.

Tal, N., Siegel, L., & Maraun, M. (1994). The role of question type and reading ability in reading comprehension. *Reading and Writing, 6*(4), 387-402.

Taylor, A. (2006). The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal, 23*(2), 309-318.

Taylor, P., & Gitsaki, C. (2004). Teaching WELL and loving it. In S. Fotos & C. Browne (Eds.), *New perspectives on CALL for second language classrooms* (pp. 131-147). Mahwah, NJ: Lawrence Erlbaum.

Tomitch, L. (1996). Individual differences in the text organization perception and working memory capacity. *Revista da ANPOLL, 2*, 71-93.

Tomitch, L. (1999). Individual differences in working memory capacity and the recall of predicted elements in the text. *Lenguas Modernas, 26*(31-51).

Tozcu, A., & Coady, J. (2004). Successful learning of frequent vocabulary through CALL also benefits reading comprehension and speed *Computer Assisted Language Learning, 17*(5), 473-495.

Turner, M., & Engle, R.(1989). Is working memory capacity task dependent? *Journal of Memory & Language, 28*, 127-154.

Unsworth, N., Heitz, R., Schrock, J., & Engle, R.(2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505.

van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External presentation of

argumentation in CSCL and the management of cognitive load. *Learning and Instruction, 12*(1), 121-138.

van den Noort, M., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*(4), 289-298.

Van Merrienboer, J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for E-learning. *Educational Technology Research and Development, 53*(3), 5-13.

Wallen, E., Plass, J. L., & Brunken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research and Development, 53*(3), 59-71.

Wang, F. (2006). Compensation between Chinese Reading Ability and English Language Proficiency in English Reading. Unpublished MA thesis. Shanxi Normal University.

Ward, M. (2004). The additional uses of CALL in the endangered language context. *ReCALL, 16*(2), 345-359.

Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), *Multimedia Language Teaching* (pp. 3-20). Tokyo: Logos International.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching 31*(57-71).

Waters, G., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Journal of Experimental Psychology, 49*(1), 51-79.

Wen, Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics, 22*(1), 1-22.

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes, 14*, 133-145.

Williams, M., & Burden, R. L. (1997). *Psychology for Language Teachers: A Social Constructivist Approach*. Cambridge: Cambridge University Press.

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Eye-Movement Recordings in Second Language Research. *Studies in Second Language Acquisition* (35), 205-212.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*, 345-376.

Xu, J. (2010). Using multimedia vocabulary annotations in L2 reading and listening activities. *CALICO Journal, 27*(2), 311-327.

Yamashita, J. (2002). Influence of L1 reading on L2 reading: Different perspectives from the process and product of reading. *Studies in Language and Culture, 23*, 271-283.

Yamashita, J. (2004). Reading attitudes in L1 and L2, and their influence on L2 extensive reading. *Reading in a Foreign Language, 16*(1).

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology, 13*(2), 48-67.

Yao, Y. M. (2006). The effect of different presentation formats of hypertext annotations on cognitive load, learning and learner control. Unpublished PhD dissertation. University of Central Florida.

Yeung, A. S., Jin, P., & Sweller, J. (1997). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology, 23*, 1-21.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology, 10*, 85-101.

Young, D. J. (2004). When good technology means bad teaching. *The Chronicle of Higher Education, 51*(12), A31.

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology, 80*, 351-361.

Zakaluk, B. L. (1998). *A theoretical overview of the reading process: factors which influence and implications for instruction*. Winnipeg: University of Manioba.

Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal, 21*(1), 7-27.