



**The Effects of Hypermedia Annotations on Young Kuwaiti
EFL Learners' Vocabulary Recall and Reading
Comprehension**

Zahra Amir

**Thesis Submitted for the Degree
of Doctor of Philosophy**

Integrated Ph.D. in Educational and Applied Linguistics

**Newcastle University
School of Education, Communication and Language Sciences**

August 2015

Abstract

Although there is a wealth of literature exploring the benefits of annotations on vocabulary learning and reading comprehension, very little exists on whether this tool is useful to young learners. Studies with adult learners show that annotations are, by and large, beneficial but, nevertheless, can be problematic for those learners with a lower working memory capacity. For these reasons it is necessary to establish whether hypermedia annotations are suitable for young learners who naturally have a lower working memory capacity than adults. The effects of annotations must be examined using a theoretical framework and research design that are suitable for this age group.

This study investigates the impact of annotations on young learners' immediate vocabulary recall and reading comprehension. The research was conducted on 112 Kuwaiti EFL learners studying in public schools at year 4. An experimental research approach was adopted using two treatment groups and one control. A within-subjects design was set in place with a counterbalance method whereby all participants sat for three reading conditions: Treatment A (Picture + L2 audio), Treatment B (Picture + L1 translation) and control (no annotations). Quantitative data were gathered using pre-tests, post-tests and software log files. Qualitative data were gathered through focus groups which took place during the last week of the experiment.

The findings suggest that reading with annotations promotes immediate vocabulary recall but does not affect reading comprehension either positively or negatively. No significant differences were found between the two treatments for vocabulary recall and reading comprehension. Qualitative data from the focus group indicated that the participants may have experienced some disorientation as they reported difficulty concentrating on both annotations and text. The quantitative data, however, from pre-tests and post-tests did not substantiate these claims. It was, therefore, hypothesised that participants exerted an extra mental effort which compensated for any disorientation. Further research is needed to verify whether this is indeed the case and to determine what, if any, are the long-lasting effects.

Acknowledgements

The completion of this thesis was only made possible with the help and motivation of those around me.

I am grateful to my supervisor, Mr William Scott Windeatt, from Newcastle University, whose continuous support and guidance have been invaluable. His positive and constructive feedback was crucial, particularly at times when I was doubtful of my abilities.

I am deeply indebted to my belated father who passed away during my PhD journey at Newcastle University. He instilled in me the love of learning and the importance of pursuing a degree. I cannot begin to thank my mother whose support at all times has been unequivocal. To her I am forever grateful.

I would like to thank the principals, secretaries and teachers from the two schools that participated in this experiment. They created a pleasant and supportive environment for my data collection.

Contents

Abstract	ii
Acknowledgements	iii
List of Tables	viii
List of Figures	ix
Chapter 1 Introduction	1
1.1 Introduction and Background to the Study	1
1.2 The Context of the Study	3
1.3 Research Gap	4
1.4 Research Questions	5
1.5 Outline of the Chapters	6
Chapter 2 Review of the Literature	7
2.1 How do children learn?	7
2.2 L2 Vocabulary Learning and Reading	9
2.3 What are Hypermedia Annotations?	11
2.4 Why use Hypermedia Annotations?	11
2.4.1 Hypermedia annotations for vocabulary learning	11
Eliminate or reduce pre-teaching time	11
Minimize inaccurate guessing of word meaning	11
Promote active language processing	12
Encourage learner autonomy	13
Replace traditional dictionaries	13
2.4.2 Hypermedia annotations for reading comprehension	14
2.5 Problems with Annotations	15
Summary	16
2.6 The Theoretical Framework	16
2.6.1 The Interaction Hypothesis	17
2.6.2 The Generative Theory of Multimedia (GTM)	18
The Working Memory	23
The Cognitive Load	26
Summary	33
2.7 Empirical Research Findings	33
2.7.1 Empirical research with young learners	38

2.7.2 Learner proficiency.....	44
2.7.3 Learner interest and prior knowledge.....	47
Interest.....	48
Prior Knowledge.....	49
2.7.4 CALL tracking.....	50
Summary.....	51
2.8 The Rationale and Focus of this Thesis.....	53
Chapter 3 Methods.....	54
3.1 Research Questions.....	54
3.2 Research Paradigm: Positivism.....	56
3.3 Experimental Research.....	57
3.4 Participants of the Study.....	58
3.5 Study Design.....	60
3.5.1 Texts and target words (Annotations).....	64
3.5.2 Electronic story design.....	66
3.6 Controlling the Variables.....	67
3.7 Research Instruments.....	69
3.7.1 Pre-test.....	69
3.7.2 Post-test.....	70
3.7.3 Log files.....	71
3.7.4 Focus groups.....	71
3.8 Piloting the Study.....	73
3.8.1 Pre-piloting the tests.....	74
3.8.2 Piloting the tests.....	74
3.8.3 Piloting the electronic stories.....	75
3.9 Validity and Reliability of the Study.....	75
3.9.1 Validity.....	76
Face validity.....	76
Content validity.....	76
Construct validity.....	77
3.9.2 Reliability.....	77
3.9.3 Trustworthiness.....	78
3.9.4 Triangulation.....	79
3.10 Experiment Procedures.....	80

3.10.1 Focus group procedures	81
3.11 Ethical Considerations	82
Chapter 4 Results	85
4.1 Vocabulary Outcomes.....	85
4.1.1 Preparing the data for R-Q 1	86
4.1.2 Testing Hypothesis 1	87
4.1.3 Preparing the data for Sub-Q 1	89
4.1.4 Testing Hypothesis 2	90
4.2 Reading Comprehension.....	93
4.2.1 Testing Hypothesis 3	93
4.2.2 Testing Hypothesis 4	94
4.3 Reading Behaviour	95
4.3.1 Focus group data.....	96
4.3.2 Weekly performance data	98
Week1	98
Week2	99
Week3	100
Summary of weekly performance	101
4.3.3 Clicking behaviour.....	102
Testing clicking frequency.....	102
Testing correlations.....	103
Hypothesis 5.....	103
Hypothesis 6.....	103
Summary.....	104
Chapter 5 Discussion	105
5.1 Effect of Annotations on Immediate Vocabulary Recall	105
5.1.1 Effect of annotation type on immediate vocabulary recall.....	108
5.2 Effect of Annotations on Reading Comprehension.....	109
5.2.1 Effect of annotation type on reading comprehension.....	111
5.3 Learners' Reading Behaviour and Preferences	112
5.3.1 Learners' reading strategies and use of annotations.....	112
5.3.2 Learner preferences for reading condition	117
5.3.3 Emerging themes	119
Chapter 6 Conclusion.....	121

6.1 Brief Summary of the Study	121
6.2 Theoretical and Pedagogical Implications	123
6.3 Limitations and Evaluation of the Study	123
6.4 Evaluation of the Study.....	124
6.4 Conclusion and Recommendations for Future Research	126
Appendix I	127
List of target words (annotations).....	127
Appendix II.....	128
Pre-tests.....	128
Appendix III.....	133
Post-tests	133
Appendix IV.....	141
Sample log file	141
Appendix V	146
Curriculum excerpt	146
Appendix VI.....	147
Focus groups transcripts	147
References.....	174

List of Tables

Table 3.4: Random assignment of participants	59
Table 3.5a: Repeated measures experimental design	61
Table 3.5b: The independent and dependent variables of the experiment.....	61
Table 3.9: Internal consistency of tests (KR-20)	78
Table 4.1a Preparing data to answer R-Q1	86
Table 4.1b Combining treatments to obtain a ‘reading with annotations’ score	87
Table 4.1c: Descriptive statistics for the pre-tests and post-tests across the reading.....	90
Table 4.1d: Paired samples t-Test.....	91
for pre-tests and post-tests in three reading conditions	91
Table 4.2a: Data preparation for reading comprehension post-tests	93
Table 4.2b: Descriptive statistics for reading comprehension scores in post-tests.....	95
Table 4.3: Tabulation of focus group data	97

List of Figures

Figure 2.6a: Paivio's Dual-coding model.....	19
Figure 2.4b: Paivio's Dual-coding model (Bilingual).....	22
Figure 3.5a: Outline of the research design.....	60
Figure 3.5b: Sample of the control reading condition.....	62
Figure 3.5c: Sample page of the Treatment A reading condition.....	63
Figure 3.5d: Sample page of the Treatment B reading condition.....	64
Figure 3.5e: Sample title page for one of the electronic stories	66
Figure 4.1a: Vocabulary pre-tests and post-tests means for reading with annotations	88
Figure 4.1b: Vocabulary post-tests means for reading with and without annotations	89
Figure 4.1c: Vocabulary post-test means for Treatment A, Treatment B and Control....	92
Figure 4.2: Reading comprehension post-test means (with and without annotations)	94
Figure 4.3a: Participants' performance during week 1.....	99
Figure 4.3b: Participants' performance during week 2	100
Figure 4.3c: Participants' performance during week 3.....	101

Chapter 1. Introduction

1.1 Introduction and Background to the Study

With the spread of technology in all fields of life, there is an increasing pressure on teachers to incorporate computers into teaching and learning. In many countries, computers and computer-assisted language learning (CALL) material have become ubiquitous in language learning environments. If CALL materials are to enhance learning English as a Foreign Language (EFL), they need to offer more than simply an electronic version of existing printed material. “CALL designers, users, and researchers need to be able to theorize not only the "normal" process of acquisition but also how to modify this normal process in hopes of helping students to learn faster and better.” (Chapelle, 2009:742). Researchers are, therefore, looking for ways to integrate second language acquisition (SLA) principles into CALL materials in order to enhance the linguistic input which appears to learners on screen.

One method that has been researched is the use of electronic ‘annotations’ to enhance vocabulary learning and reading comprehension. Traditionally, annotations appear in printed books either in the margin or at the bottom of the page. They provide peripheral information for words and terms which the author feels the reader may not comprehend. This method of explanation has found its way to digital reading as it is common to find web pages that include highlighted words that are linked to other texts or websites (hypertext and hyperlinks). It is thought that EFL readers too can benefit from this process of connecting words to other forms of input. In particular, providing clarification for words through annotations is thought to aid reading comprehension and vocabulary learning.

The provision of annotations enables learners to understand new words and reduces their dependency on the teacher. Meaning is explained to learners through annotations as they read the text. The teaching time is reduced since learners select the words they do not understand rather than relying on the teacher to pre-teach all the potentially difficult words prior to reading. Annotations also aid with the mental process of deciphering meaning that takes place when learners encounter unfamiliar words. The help provided for words encourages learners to integrate the new

information with their current knowledge in order to comprehend the reading material. Thus, learners are not left to derive meaning entirely on their own but are supported through annotations. One can argue that annotations have created a middle ground between spoon feeding learners vocabulary knowledge prior to reading and leaving learners to figure out meaning on their own which may result in incorrect meaning inferences, confusion and de-motivation.

The crucial question remains though as to what electronic annotations can offer that is superior to the ones found in printed material. Mayer's (1997) Generative Theory of Multimedia (GTM) has featured prominently in much of the research carried out on the efficacy of electronic annotations. The theory argues that learning is more effective when information is presented in more than one mode. For example, a difficult word that is explained through a text definition with an illustrative visual is better understood than if presented with a text-only or visual-only explanation. This is because multiple modes activate different corresponding parts in the brain and this creates stronger connections in the mind. The use of computers has made it easy for researchers to create an electronic text and attach annotations that include additional text, images, audio and/or video for clarification purposes. This diversity renders electronic annotations superior to traditional ones. Annotations that include multiple forms of media are referred to as 'hypermedia annotations' and are the subject area of this study.

This research is carried out within the theoretical framework of the GTM and its two sub-theories: The Dual-coding Theory and The Cognitive Load Theory. The Dual-coding Theory developed by Paivio (2007) states that there are two separate systems in the brain for processing verbal and non-verbal information. The GTM builds on this by calling for the simultaneous presentation of two types of information, such as visual and textual, in order to activate the corresponding sensory modes in the brain. This is believed to foster better retention of information (Mayer, 1997 & 2001). The Cognitive Load Theory centres on the idea that electronic material should be designed in accordance with learners' limited cognitive processing resources (Gerjets et al., 2009). If the design of material is too complex, a 'cognitive overload' may occur and learning outcomes are consequently affected (Kirschner, 2002).

1.2 The Context of the Study

“Whereas in the past, research in SLA had a limited following, research regarding how learners use technology for L2 learning and particularly how to optimize this learning holds interest to a much wider set of people, including teachers, learners, business people, and governments.” (Chapelle, 2009: 750).

In recent years, The Ministry of Education in Kuwait has taken measures to promote reading among Kuwaiti school children. The ministry launched a reading programme called “Today a Reader, Tomorrow a Leader” which aims to instil the love of reading in young children through various reading activities at schools. Alongside the ministry’s initiative is the award winning reading initiative carried out by the British Council in Kuwait called ‘Kids Read’. It aims to “encourage children to move from teacher led to independent reading” (British Council, 2014) and the project is in its third consecutive year. Schools from all six educational districts in Kuwait are involved and the participating primary schools each receive a set of approximately 50 popular UK story books. These can be used by teachers in class or self-accessed by students. The initiative also involves teachers and trainers who undergo workshops on how to use stories in class and then cascade this knowledge to others. In addition, parents are invited to storytelling events in schools and/or public venues whereby they are shown ways in which they can encourage and develop their children’s reading skills at home. The ministry of Kuwait has also invested in developing the computer skills of pupils throughout the school grades. Each school is equipped with a computer lab and students take a computer science subject as part of the curriculum. In addition to this, students in primary schools are taken to the lab for lessons that supplement other school subjects such as English, Science, Arabic, Religion ...etc.

While both these reading initiatives are undoubtedly supportive of reading, they do not incorporate electronic reading. From my point of view, promoting electronic literacy is also of vital importance especially since children nowadays are active online and enjoy the use of computers and other electronic devices. Maynard and McKnight (2001) view the electronic environment as motivational for children who are reluctant to read printed text. It is also claimed that the features provided by electronic texts make them superior to printed ones. It would be beneficial to make use of this technology to further instil the

love of reading in young learners, and reading with annotations may well aid in the Ministry's aim to move Kuwaiti learners towards independent reading. Nevertheless, empirical research that is both supportive of claims and relevant to the target group is needed to establish whether electronic reading is worth the investment.

1.3 Research Gap

This PhD research aims to build on the current literature pertaining to the effects of 'hypermedia annotations' on vocabulary learning and reading comprehension. Although there is a wealth of studies in this area, the need for this study stems from two observations. Firstly, the literature on annotations has focused primarily on adult learners but the use of CALL material is not only restricted to adults as more and more children are using technology in school and at home. The prevalence of visual literacy amongst young learners renders electronic books a more attractive reading medium than printed ones (Maynard and McKnight, 2001). Only a few studies, however, have involved the use of annotations with young learners. The studies by Tabatabaei and Shams (2010) and Gerjets et al. (2009), for example, both involved teenagers. Only one study that I have come across involves children and looks specifically at the dual-coding theory (Acha, 2009). One cannot simply extrapolate the findings from research with adults to children since their developmental needs and cognitive abilities differ greatly. Thus, before embarking on creating reading texts with annotations for young learners, it is necessary to design and carry out research in a manner that is appropriate for their age group. This would provide relevant empirical evidence to help determine whether the use of annotations is beneficial for young learners. The second reason for doing this research is to contribute to the GTM and The Cognitive Load Theory. Research suggests that annotations are more suitable for intermediate and advanced learners of English because their cognitive abilities are better equipped to handle the complexity of non-linear reading. There is a suggestion that the learner control provided in hypermedia reading is only favourable for learners with high prior knowledge because they are more capable of navigating and employing deeper processing strategies to produce better learning outcomes (Gerjets et al., 2009). If cognition is a pre-requisite for the success of annotations, then the implication is that children are likely to be at a disadvantage since they naturally possess lower cognitive abilities than adults. This research sets out to investigate whether this is the case.

The studies on the efficacy of hypermedia annotations with adults differ considerably in their results. This study discusses the possible reasons for this by shedding light on how experiment variables and seemingly insignificant software design features can alter the learning outcomes of using hypermedia annotations. The findings from these were then used as the basis for the design of this study and its research instruments.

1.4 Research Questions

The aim of this study is to examine the effect of hypermedia annotations on young Kuwaiti EFL learners' immediate vocabulary recall and reading comprehension. The research also looks at the young learners' attitudes towards reading with annotations and uses computer tracking software to investigate their use of annotations. It is hoped that the findings from this investigation will broaden our understanding of the use and design of CALL materials with young learners and encourage further research in this field.

The research questions are as follows:

***R-Q 1:** Do hypermedia annotations facilitate vocabulary learning for young Kuwaiti learners?*

***Sub-Q 1:** Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for L2 vocabulary recall?*

***R-Q 2:** Do hypermedia annotations facilitate reading comprehension for young Kuwaiti learners?*

***Sub-Q 2:** Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for reading comprehension?*

***R-Q 3:** How do young Kuwaiti learners use annotated texts and how do they feel about this type of reading?*

This is an experimental piece of research which uses a primarily quantitative approach to answer the above questions. In TEFL, texts are used as a platform for teaching vocabulary and structures both explicitly and implicitly. The first research question aims to examine whether enhancing an electronic text through the use of annotations can improve vocabulary learning. The type of learning tested here is ‘receptive’ vocabulary knowledge rather than ‘productive’ knowledge. The distinction between the two terms is addressed in section 2.2. The second research question looks at reading comprehension. The decision to test this language skill stems from the need to determine whether the presence of annotations throughout an electronic text may distract learners and affect their overall comprehension of the text. Statistical analysis of pre-tests and post-tests is used to examine the efficacy of annotations on both vocabulary learning and reading comprehension. A computer tracking device is used to gather quantitative data for the third research question and qualitative data is also gathered through focus groups. These data will be used to answer question 3 and supplement the findings of questions 1 and 2.

1.5 Outline of the Chapters

This study is organised into six chapters. Chapter 1 has given an introduction and some background information on the research topic. It has also highlighted the significance of the research and presented the research questions. Chapter 2 provides a more detailed account of the use of annotations and places this within a theoretical framework. The chapter also reviews the empirical research on annotations. Since this is largely based on research with adults, a discussion takes place throughout highlighting why it is not feasible to generalise the findings to young learners. Each section of chapter 2 includes a summary of the main ideas and findings. Chapter 3 is dedicated to discussing the methodology of this study. This covers the study design, participants, research instruments and procedures, measures taken for validity and reliability, and ethical considerations. Chapter 4 presents the results of the data analysis and answers the research questions. Chapter 5 discusses the data analysis findings in light of previous research and also presents some theoretical and pedagogical implications. The conclusion to the research is addressed in the sixth and final chapter.



Chapter 2. Review of the Literature

This chapter opens with an overview of how children learn, then focuses on L2 vocabulary learning and reading and shows how the two skills are related. Hypermedia annotations are then defined and the pedagogical reasons for using them in computer-assisted language learning are listed. A discussion on why the use of annotations may be problematic also takes place. The theoretical framework that underpins this research is then presented and followed by a review of the literature. The review begins with the empirical findings from research with adults. The findings are discussed and analysed in light of the theoretical framework. A review of empirical findings from research with young learners then follows. Since there is a lack of research on annotations with young learners, this section was supplemented with other CALL-related and Human Computer Interaction (HCI) research to provide further insight into how children use CALL material. The role of learner proficiency is looked at with respect to how it impacts the efficacy of annotations. Similarly, learner interest and background knowledge are examined. A summary for the key points raised is given at the end of each section. The chapter concludes with a section on the importance of using computer-tracking in CALL material.

2.1 How do children learn?

From a constructionist point of view, children are best able to learn in environments which foster discovery and participation rather than through lectures or step-by-step instruction (Alesandrini and Larson, 2002). With the help of an adult, teacher or peer, a child is able to make connections between various concepts and build on his/her current knowledge (ibid). Thus, in essence, the theory of constructionism is based on the notion that knowledge is created rather than transmitted (Candy, 1991) and that learning is most effective when the child is engaged in a task that is meaningful (Sabelli, 2008) and tangible (Ackerman et al., 2009). The main criticism of constructionism, however, is that it does not provide a concrete, computational account of how learning in such an environment would lead to mental representations (Marcus, 1998). In other words, the theory focuses on the social context of learning but fails to shed light on the biological workings of the mind. Some researchers (Elman, et al., 1996), however, believe that

‘connectionist’ models of learning work hand in hand with constructionism and can make up for its shortcomings.

‘Connectionism’ offers a behavioural approach to understanding learning. The theory argues that language acquisition is a by-product of engaging in a meaningful task. It is “something that develops in the course of learning how to perform the primary communicative tasks of comprehension and production” (Seidenberg and MacDonald, 1999: 571). Within a connectionist model, mental processing involves interacting with input and forming neural connections. “All knowledge is embodied in a network of simple processing units joined by connections which are strengthened or weakened in response to regularities in input patterns” (Gasser, 1990: 179). McClelland et al. (1986) describe these processing units as either ‘possible hypotheses’ or ‘possible goals and actions’. When a learner encounters a new word, for example, possible hypotheses may be formed about the word based on its’ morphological or syntactical properties (McClelland et al., 1986). If the word appears to be a noun, the mental networks relative to nouns will be activated while those relating to verbs will be inhibited. In other cases, the processing units account for goals or actions. Looking at CALL as an example, a change in the shape of the cursor from  to  may alert the learner of a change in the cursor’s function. This, in turn, may prompt the learner to take action (physical clicking on mouse) to achieve the new goal of the cursor. Thus, in connectionism, learning is seen as a mental network of connections and associations which expand as the learner is exposed to more and more linguistic input (Gass and Selinker, 2008). This implies that the type of linguistic input to which learners are exposed is crucial to learning (Gass and Torres, 2005). For example, providing input through two types of modes (i.e. reading and listening), is more conducive to learning than a single mode of input because it activates more networks in the brain, thus strengthening the neural connections.

As with adults, children’s mental networks expand as they become more experienced with life and gain a better grasp of language. A young child may first deduce that ‘cat’ refers to a four-legged animal and will subsequently refer to all four-legged animals as ‘cat’. The child in this case has ‘over-generalised’ the concept of the word ‘cat’. Similar processes occur in language acquisition when children begin to learn the past tense (Elman, 2001). Since most verbs in the English language are regular and require the addition of ‘ed’ to achieve the past tense (play – played), a child is likely to over-

generalise this rule by adding 'ed' to irregular verbs to produce words such as 'goed' instead of 'went'. From a connectionist perspective, these inaccurate verbs are produced because the mental networks of a child are not developed or experienced enough to produce the past tense effectively (McClelland and Rumelhart, 1986). Nevertheless, while children are not given explicit rules on language, they are able to build multiple networks which over time become capable of producing correct forms of past tense verbs, both regular and irregular (McClelland and Rumelhart, 1986).

Although the discussion above has shed light on the workings of the mind, it is important to look into how the working memory impacts this process. The properties and role of the working memory within the connectionist model will be discussed in section 2.6.2 in light of the computer-assisted learning model (The Generative Theory of Multimedia) used in this experiment. The differences between the ways in which adults and children learn will be discussed further throughout the thesis in relation to the working memory and theoretical framework of the study.

2.2 L2 Vocabulary Learning and Reading

A complex, two-way relationship exists between reading and vocabulary learning. On the one hand, the more words a learner knows, the more likely he/she will be able to understand a text. On the other hand, the more learners read, the more likely they will pick up vocabulary from the reading context. A distinction is often made between receptive knowledge and productive knowledge (Nation, 2001 and Thornbury, 2002). With receptive knowledge, a learner can identify a particular word when encountered in spoken or written form and can recall the meaning of that word, albeit to varying degrees of accuracy. Productive knowledge, however, involves the ability of a learner to use or 'produce' a word in speech or writing. This inherently requires a deeper cognitive process and a better understanding of a word. Nation (2005) demonstrates that to know a word involves an understanding of its meaning, form and use. He refers to the complexities of these aspects as the 'learning burden' of a word. Repeated exposure to a word in different contexts enables learners to gradually build up their knowledge of that word and move from a basic receptive understanding to a more productive one.

When reading, learners may employ 'top-down' and 'bottom-up' strategies to make sense of a text. Bottom-up processing is characterized as a low-level cognitive process

(Antonenko and Niederhauser, 2010) whereby the reader makes use of information at the most basic level to reach a more general understanding of the text (Nunan, 1999). At a linguistic level, the reader may decode characters and recognise words and parts of sentences (Antonenko and Niederhauser, 2010). A reader may also turn to other clues available such as visuals or the layout of the text in order to form a better understanding of the reading material (Nunan, 1999). Top-down processing is regarded as a higher-level of cognitive processing in which a reader's background knowledge and expectations are applied to guide the process of understanding a text. The new information is then integrated into existing knowledge (Myong, 2005). Top-down and bottom-up processes will be revisited in section 2.6.1 in relation to the Interaction Hypothesis.

Before giving learners a text to read, a teacher may choose to pre-teach the vocabulary items that will appear. This is done to facilitate the reading and comprehension of the text, or even to activate the students' current knowledge and provoke interest. Alternatively, a teacher may not pre-teach the new words in a text, but encourage students to decipher their meanings through context or use of dictionary. Arguments in favour of the 'pre-teaching' approach centre on the belief that learners need to have a certain basic level of vocabulary understanding before reading a text (a vocabulary threshold). If a text has too many new and difficult words, readers will feel disoriented and de-motivated. A counter argument in favour of deciphering vocabulary meaning from context is based on The Involvement Load Hypothesis proposed by Laufer and Hulstijn (2001). It centres on the belief that learners are more likely to remember words when extra mental effort is exerted for understanding. Deciphering meaning from context or consulting a dictionary requires mental effort which is thought to contribute to the retention of the new words.

While one can debate as to which approach is better for learning, the introduction of CALL has, in many ways, altered the dynamics of reading and vocabulary learning. The electronic text is a significant new medium that offers additional value to the printed book by incorporating other forms of media alongside text on its pages (Maynard and Mcknight, 2001). The addition of visuals, sounds and hypertext create an element of interactivity within an electronic text making it distinguishable from its printed counterpart (ibid). The following sections demonstrate the role of annotations in

electronic reading and the ways in which they promote learning and offer alternative perspectives on SLA theories.

2.3 What are Hypermedia Annotations?

Annotations, also referred to as glosses, are “[added] comments or notes about difficult words, phrases, or ideas in order to provide their definitions or meanings in a particular context” (AbuSeileek, 2008: 260). This extra information usually appears in the margins or at the bottom of pages in printed books. When words in electronic texts are linked to textual information, the outcome is ‘hypertext’. When linked to additional forms of media, such as visuals and audio recordings, the result is ‘hypermedia’ annotations.

2.4 Why Use Hypermedia Annotations?

Hypermedia annotations are deemed effective for language learning and there are a number of pedagogical reasons for incorporating them into electronic texts. The following sections demonstrate how hypermedia annotations can promote both vocabulary learning and reading comprehension:

2.4.1 Hypermedia annotations for vocabulary learning

Eliminate or reduce pre-teaching time

It is common practice for teachers to ‘pre-teach’ potentially difficult vocabulary items in a text. This is done to facilitate the reading process for the students by eliminating any confusion which may arise from the presence of unfamiliar words. Although this is useful for learners, the process can become lengthy and tedious as the new words are usually presented in isolation. With hypermedia annotations, new words are learned while reading and meaning is reinforced through context. Learners can consult annotations according to their own needs and this is particularly advantageous when compared to being exposed to an entire list of words in the pre-teaching phase, some of which a learner may already know.

Minimize inaccurate guessing of word meaning

In contrast to pre-teaching vocabulary, learners are sometimes encouraged to infer meanings of new words from context. This is done to increase their mental effort which

is believed to have a positive effect on vocabulary retention (Laufer and Hulstijn, 2001). Research findings (Parel, 2004, Smidt and Hegelheimer, 2004) indicate that it is necessary to equip learners with strategies for lexical inferencing. Teachers may, therefore, train their students to use strategies for decoding word meaning. It is argued that the cognitive strategies used for inferring meaning from context are of vital importance for vocabulary development (Ellis, 1995). While more proficient readers are usually capable of inferring the meaning of an unknown word correctly (Chun and Plass, 1996), less proficient learners may not always be able to deduce meaning from context (Peters et al, 2009). The provision of annotations, in this case, caters for those who need help with deducing meanings of unknown words. As for the more proficient learners who are accustomed to using inferencing strategies, they can use annotations to check the accuracy of their guesses (Lenders, 2008). Thus, the use of annotations can prevent or minimize incorrect inferences and guesses about word meaning (Paribakht and Wesche, 1997 and Lenders, 2008).

Promote active language processing

One might assume that the provision of annotations would override the beneficial mental effort normally exerted when inferring meaning from context. Chapelle (2003), however, argues that allowing learners to work alone on CALL material can promote deep mental processing. The presence of annotations encourages an active approach to reading because readers are able to create their own knowledge representations (Antonenko and Niederhauser, 2010). Thus, the desirable mental effort involved in the construction of meaning is still taking place. With traditional printed annotations, learners may feel they are obliged to read the information because it is present alongside the text (Taylor, 2009). With hypermedia annotations, however, learners decide which annotations to consult. They are in charge of identifying their own knowledge gaps and allocating their attention resources to information that they perceive to be relative and important. This type of active processing can also be encouraged if the annotations are designed to include demands or tasks for the learners to perform such as answering comprehension questions about the annotated words.

Annotations are beneficial for 'bottom-up' processing (Gettys et al., 2001). Through annotations, a learner can understand difficult words which will then facilitate the understanding of the text as a whole. Top-down processing can also be facilitated by

creating annotations that include links to other topics or background information relevant to the text (Antonenko and Niederhauser, 2010). It is argued that providing effective bottom-up assistance in the form of annotations will in turn facilitate top-down processing (Taylor, 2009) and vice versa.

Encourage learner autonomy

In addition to promoting active language processing, Huang (2013) suggests that encouraging learners to take charge of their own learning will boost their motivation. According to Lenders (2008), consulting annotations promotes learner autonomy and guides students towards an active approach to vocabulary learning. This is because learners are able to isolate words from the text and focus on them individually. They have the freedom to request help for the words they see necessary. This freedom extends to the pace of reading and is particularly useful for learners who struggle to keep up with the pace of reading in class (Ariew and Ercetin 2004 and AbuSeileek, 2011). Hence, reading with annotations may provide a sense of security for some learners and increase their motivation.

Replace traditional dictionaries

Another benefit of annotations is that they are embedded in texts, thereby ensuring that meanings of words are automatically reinforced through context. Learners can see how a word is used naturally rather than in isolated examples in dictionaries. Since annotations are tailor-made for texts, it means that the information included for a word is more relevant and accurate than a dictionary (Taylor, 2009). Annotations are also accessed much faster than looking up a word in a dictionary (Dziemianko, 2010; Nation 2001). The help that is provided at the click of a mouse makes annotations superior to dictionary consultation which is time-consuming and disruptive to the reading process. Furthermore, looking up words in dictionaries causes a shift of attention away from the text which, in turn, consumes the learner's attentional resources that are needed for understanding the text as a whole. Dziemianko's (2010) study compared the benefits of monolingual dictionaries in paper and electronic form for receptive and productive tasks to see which is more useful for retention of meaning and collocations. Sixty four students of English at upper-intermediate level and advance level took part. She used 3 tests: a pre-test (before the experiment), a test proper (during the experiment) and a surprise delayed recall test (after the experiment). The test proper included a receptive

and an active recall task. Participants had to complete the test using the electronic or paper version of COBUILD6. The pre-test was identical to the test proper and was administered immediately before. The post-test given two weeks later was also identical to the test proper but with a re-shuffling of the item sequencing to avoid recall issues. Dziemianko's results revealed that the electronic dictionary group outperformed the paper group significantly in both receptive and productive tasks. Similar significant differences were observed in the retention test in favour of the electronic group. It was speculated that the quick retrieval of information from electronic dictionaries might lead to shallow processing and poor retention of the looked-up words. The study, however, did not find this to be true. The researcher believes that the form of electronic presentation was more "captivating" and less time-consuming than the paper form and this led to better results in the tests. Although this study does not look at annotations per se, it does demonstrate the positive effects of easy and speedy access to meaning which are typical of annotations. The author of the study believes that the attractive nature of the electronic form also had a positive influence on the study's outcome. CALL material appeal and learner motivation are discussed further in section 2.7.3 as they are relevant to the third research question in this study.

2.4.2 Hypermedia annotations for reading comprehension

It is argued that second language learners need to know 95-98% of the words in a text in order to gain a sufficient understanding of its meaning (Lenders, 2008). This entails that an overall comprehension of a foreign text can be achieved once learners have reached a vocabulary 'threshold'. This refers to the amount of vocabulary learners have in the target language which would enable them to understand a text (Nation, 2001). Some researchers suggest that once learners have grasped the 2000 most frequently used words in English, they are able to recognize and use 84% of words in different texts (non-academic) (Parel, 2004). Other researchers recommend a vocabulary threshold of 3000 word families (Thornbury, 2002). It is important to note, however, that the one hundred most common words in English mainly include function words (Thornbury, 2002). With young learners, these are learned at a later stage as children begin by learning concrete, rather than abstract, vocabulary (Cameron, 2001).

An opposing view, however, exists as researchers argue that it is not the threshold alone that facilitates reading, rather the ability of a learner to infer meanings of words through

different strategies (Parel, 2004). Another factor to consider when estimating the vocabulary threshold is the distinction between receptive and productive word knowledge (Nation, 1990). The incrementality principle posed by Nagy and Scott (2000) suggests that word knowledge exists on a continuum and that learners move from no knowledge, to different degrees of partial knowledge until they reach a more complete understanding of a word's meaning.

Annotations have now further challenged the importance of a threshold. The peripheral help provided by annotations simplifies the reading material which, in turn, grants learners greater access to authentic texts (Akbulut, 2007; Nation, 2001; Jacobs et al., 1994 and Long, 2007). The reading material is simplified because the peripheral information can help with difficult words or concepts in the text. It can, therefore, be argued that the use of inference strategies and the provision of annotations render a vocabulary threshold less significant for reading comprehension.

2.5 Problems with Annotations

Although the benefits of annotations may be evident, they are not entirely problem-free. One issue concerning using annotations with learners is that of authenticity. This refers to “the degree of correspondence between an L2 learning task and tasks that the learner is likely to encounter outside the classroom” (Chapelle, 2001:56). When browsing the World Wide Web, it is not uncommon to find authentic electronic texts which are annotated, such as those available on wikipedia.com. EFL vocabulary annotations, however, are usually custom-made for particular groups of learners. The type of information provided by EFL-specific annotations, can, therefore, be quite different from that encountered previously by learners. Additionally, it would be unwise to get learners accustomed to using annotations to the point where they become fully dependent on them for establishing meaning. Learners need to develop other strategies for deducing meaning when annotations are not readily available. A recent study by Huang (2013) demonstrated that the most beneficial text is one that includes both annotated words and other words that require readers to use inferencing strategies

Another issue with annotations is that their design process is time-consuming. It is, therefore, worthwhile carrying out research to examine whether the learning impact of this technology outweighs its drawbacks (Lenders, 2008). Examining the efficacy of

annotations is not a straight forward process, however. The different types of information provided through annotations, in addition to their design, can result in drastically different learning outcomes as shall be discussed in section 2.7.

Summary

The previous sections have examined the advantages of using annotations for vocabulary learning and reading. Not only do annotations minimize the need for teachers to pre-teach vocabulary items, they also reduce inaccurate guesses of word meaning and distraction caused by traditional dictionary consultation. Reading with annotations provides students with an environment in which they can work at their own pace and take charge of their learning. Learners with reading difficulties or a low vocabulary threshold can also benefit from the help provided by annotations. Since EFL annotations are not commonly found outside the classroom, it is necessary to examine whether the time and effort invested in creating them is worthwhile. Empirical research is needed to determine how to optimise the use of this CALL reading tool.

2.6 The Theoretical Framework

It is important to examine the efficacy of annotations in light of theoretically-grounded constructs. The distinct and varying designs of annotations create multiple variables that ultimately affect research outcomes. The theoretical framework informs the design and execution of research on annotations. Not only does it enable one to examine the characteristics and effects of annotations more accurately, but also allows researchers to reciprocate by contributing to the understanding of the theory. “As a consequence, all approaches to SLA that theorize a role for linguistic input need to consider the way that technology changes linguistic input and how learners' access to new forms of input might affect acquisition.” (Chapelle, 2009:750). This section discusses the rationale for using annotations from a theoretical stand-point. The two key theories discussed are The Interactionist Hypothesis and The Generative Theory of Multimedia (GTM). The section looks at how these theories can inform the design of electronic reading with annotations. Section 2.6 concludes with a summary of the key information discussed.

2.6.1 *The Interaction Hypothesis*

The use of hypermedia annotations falls under an interactionist paradigm of second language acquisition. This paradigm operates on the basis that meaningful learning takes place when learners are encouraged to notice particular aspects of the linguistic input (Long, 1996) during meaning-oriented tasks (Chapelle, 2009). Explicit attention to input is believed to encourage deep mental processing which in turn results in better retention of information (Schmitt, 2000). In traditional classrooms, a teacher may, for example, adopt a change in tone to alert students to a particular structure or word in spoken discourse. A teacher may also underline particular words in books or on the board (Chapelle, 2003). Psychological research has shown that drawing learners' attention to particular linguistic features promotes linguistic acquisition (Schmidt, 1990 and Skehan, 1998). In addition to noticing input, the concept of 'negotiation of meaning' is central to the Interaction Hypothesis (Long, 1996). When learners find 'a gap' in their knowledge, they engage in a process of seeking clarification from their colleagues or teacher. In a conversation, this may involve the learner asking the speaker to repeat a difficult word or explain its meaning. With text, however, interaction or negotiation of meaning takes place through top-down and bottom-up processes (see section 2.2). When learners encounter new words, they may resort to their background knowledge to help them understand (top-down). If the text they are reading is an expository text about computer mechanics, for example, they can expect to encounter new technical terms. Conversely, learners may use a bottom-up approach by looking at the linguistic features of unfamiliar words. For example, if a word is preceded by the indefinite article 'a', the learner can deduce that the word is a noun, or if the auxiliary 'very' precedes it, they can assume the word is an adjective and so on.

When comparing The Interaction Hypothesis to connectionism, the two theories share the importance of deep mental processing but differ in their explanation of the mechanics behind this process. The Interaction Hypothesis argues that processing occurs through noticing, focusing and negotiating meaning, while connectionism views deep mental processing as the activation of multiple neural networks in the brain in order to consolidate meaning.

CALL researchers and educators are exploring ways to create a similar digital environment which fosters noticing and negotiation of meaning. In her book, Chapelle

(2003) discusses how CALL materials can be designed to achieve this interactionist form of learning. She lists three means of enhancing linguistic input; salience, modification and elaboration. Salience involves directing learners' attention to particular areas in the input presented on screen. Words or sentences that are highlighted or underlined, for example, become salient and will encourage learners to notice and begin negotiation of meaning. This process is further developed when modification is available for these highlighted forms. Modification involves providing the learner with additional information to make the input more comprehensible. Chapelle's (2003) third means of input enhancement is elaboration. This involves adding semantic detail to the salient input, usually in the form of meta-language, such as defining a relative clause or pronoun...etc. With hypermedia annotations, it is possible to meet Chapelle's (2003) criteria for interactive software. Annotated words in an electronic text are made prominent through highlighting, underlining or a change of font (salience). A learner then notices these words and clicks on them to receive clarification (modification and elaboration) in the form of textual, visual and audio input.

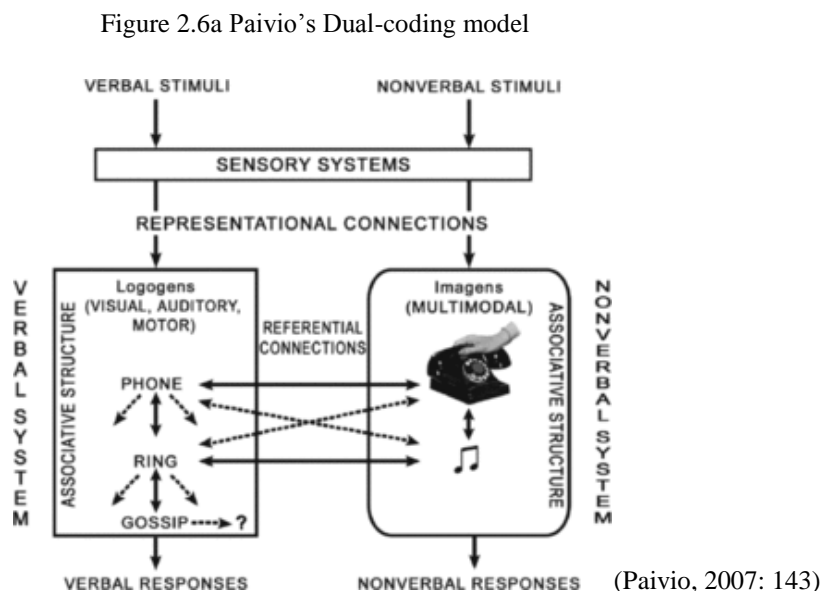
Although the Interaction Hypothesis provides direction for the design of computer-based materials, it is important to note that it is not a theory that was developed specifically for computer instruction. It is, therefore, necessary to incorporate this theoretical framework with another that is built on computer-assisted language learning. The following section discusses the Generative Theory of Multimedia (Mayer, 1997) which echoes the interactionist view of SLA and is closely associated with studies on hypermedia annotations.

2.6.2 The Generative Theory of Multimedia

The Generative Theory of Multimedia emerged when Mayer (1997) became puzzled as to why his students were not able to answer questions relating to an apparently straightforward text they had read on the computer screen. He found that although the text explains a scientific process clearly, students were not able to use the information to answer a follow-up transfer question. Mayer investigated the reasons for this and his research led him to conclude that to help his students, he needed to adopt a *multimedia* approach. Instead of presenting information in a single mode, he made use of multiple modes. Textual information was presented hand-in-hand with other modes such as visual and/or audio input. By combining different types of input, learners were better

able to grasp the concepts in a text. Mayer's Generative Theory of Multimedia is based on two theories of learning; the dual-coding theory and the generative theory.

The dual coding theory by Paivio (1986) is neuropsychological in its origin but later developed as a behavioural explanation and a psychological theory of the mind. Neuropsychological data initially showed a distinction in the brain between a verbal system and a non-verbal system. Behavioural studies later built on this by determining the specific functions of these two systems and by demonstrating that they operate independently of one another. The dual-coding theory states that the cognitive processes of learning occur through these distinct information processing systems. Figure 2.4a illustrates the two systems in the brain.



The verbal system processes verbal information whether written or spoken. The words 'phone' and 'ring' may be encountered in written form or spoken form (verbal stimuli) and will subsequently be processed through the verbal system. Alternatively, an image of a 'phone' as an object or the sound of it ringing (nonverbal stimuli) is processed in the non-verbal system in the brain.

Paivio's theory goes against the connectionist approach in terms of brain compartmentalisation. The idea that the brain is made up of separate, distinctive faculties is rejected within connectionism (Jordan, 2004) as the theory views the brain as one large web of interconnected neurological networks. The evidence for the division of

the brain is, however, compelling. According to Paivio, the verbal system residing in the left hemisphere of the brain processes linguistic information both in written and spoken form. Research has shown that injuries in this hemisphere result in language impairment and the precise location of the injury within the hemisphere affects the degree of impairment (Paivio, 2007). The non-verbal system (the right hemisphere) is responsible for processing information in the form of visuals or non-linguistic sounds. This is not to say, however, that images are processed exclusively by one side of the brain, it is more a matter of dominance. Although the systems are independent of one another, they are connected in the sense that the processing of input in one system prompts the activation of the other and thus 'referential connections' are made. If one sees the word 'phone' written, for example, the non-verbal system may also be activated with an image or prototype of the object appearing in the mind and vice versa. Thus, decoding words is both verbal and non-verbal (Sadoski, 2005). Paivio's (1986) dual coding theory features prominently in studies on the effect of annotation types. The inclusion of both visual and verbal information in an annotation aims to activate the two corresponding systems in the brain. Consequently, the reader unites the two types of information to make sense of the annotated word. It is argued that the mental representation for an unknown word is more likely to be retained due to the dual coding of information.

Mayer's Generative theory of Multimedia incorporates the dual-coding aspect by arguing that meaningful learning encompasses the activation of both verbal and non-verbal cognitive processing systems (Mayer, 1997 & 2001). More importantly, he calls for the *simultaneous* activation of these systems in the short-term memory for effective learning to take place. He suggests that text and illustrations, for example, be presented in near proximity rather than successively or separately. This echoes 'the contiguity principle', a subsidiary of the dual-coding theory, which calls for presenting information simultaneously in order to avoid 'split attention' (Kirschner, 2002). This is a concern with hypertexts which require the reader to hold information from the text in the working memory while also process new information encountered in a new, separate node (e.g. annotations) (Antonenko and Niederhauser, 2010). The split attention effect is further discussed later in this section. In terms of designing electronic reading material, the contiguity principle implies that it is best to have annotations appear in near proximity to the corresponding words. Another subsidiary of the dual-coding theory is the 'modality

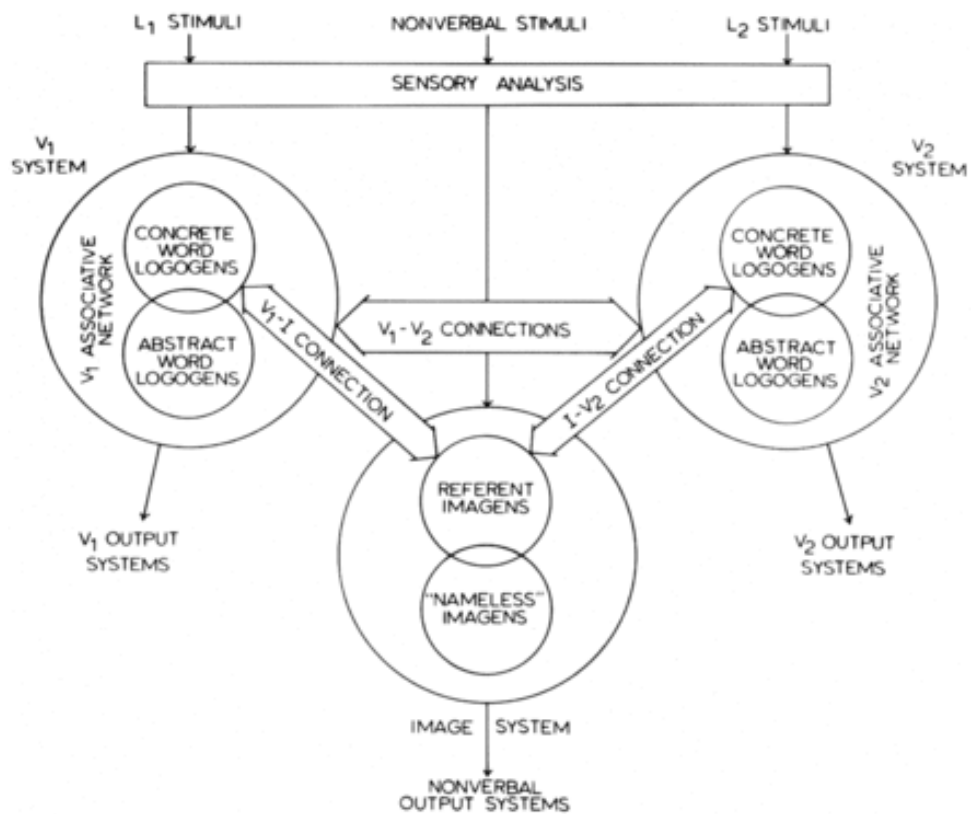
principle' which proposes that, as opposed to visual text on the screen, it is more beneficial to present words or text as an auditory narration to accompany visual images. This is because the combination of animation + audio was found to be more effective than animation + text (Moreno & Mayer, 1999 and Mayer, et al., 2003). These studies were not carried out in an EFL context, however, and the combinations tested were in relation to an expository animation that explained a process to adult learners of science. It is essential to examine whether the principle holds true for annotations.

The second component of Mayer's (1997) theory is the generative approach to learning. It suggests that meaningful learning takes place through three key processes; selecting, organising and integrating information. When learners are presented with textual and visual input, they first need to *select* textual information to create a textual base and select illustrations to create a visual or pictorial base. The learner then *organises* the selected textual base to form a verbal-based model in the short-term memory. Similarly, the visual base is transformed into a non-verbal model in the short-term memory (Kirschner, 2002). The final step is *integration* whereby the learner connects the verbal-based model with the visual-based one and maps it onto existing knowledge to construct a coherent mental representation. The process of integration and mapping new information onto existing representations in the mind is often referred to in connectionist contexts as 'schema mapping' (Cook, 1997). 'Schemata' can be defined as the networks of data and meaning that are stored in memory (Sadoski, 2005). When a learner encounters a new word or concept, the relevant schemata are activated in order to make sense of the new input. The freedom that accompanies the selection, organisation and integration of presented information is believed to better enable learners to form a coherent mental representation of the input. The learner is regarded as a knowledge constructor (Mayer, 1997) and is actively involved in negotiation of meaning. The design of multimedia learning materials will inevitably determine the degree to which learners are able to select and process visual and verbal information to construct a mental representation.

Although the Generative Theory of Multimedia emerged from a computer-assisted learning context, it is not oriented specifically to language learning. Paivio's dual-coding model (figure 2.4a above) that features in Mayer's theory is a mono-lingual one and as such does not take into account the additional complexity posed by the L2. To gain a better understanding of how the GTM fits in with EFL learning, it necessary to

examine how Paivio incorporated bilingualism in his dual-coding theory. Figure 2.6b presents Paivio's bilingual model of the dual-coding theory. Unlike the monolingual model (Figure 2.6a), there are two verbal systems (v1 and v2) which correspond respectively to a person's first language (L1) and second/target language (L2). The single, non-verbal system is shared by the two verbal ones.

Figure 2.6b Paivio's Dual-coding model (Bilingual)



(Paivio, 2007:112)

It is apparent from figure 2.4b that the bilingual model is more complex than its monolingual counterpart. In particular, a heavier strain is placed on the short-term memory (working memory) because it is not only processing a new concept but also a foreign language. If a learner encounters a word in the L2 (L2 stimuli), the corresponding word is activated in the V1 system. Hence, the two verbal systems are connected (V1-V2 connections) through translation equivalents. Figure 2.6b shows that concrete words in the L1 and L2 systems are connected to the imagery system. The image system is also considered an alternative route to translation and thus a learner may resort to images to understand input in the L2, rather than find an L1 equivalent.

When an EFL learner reads an electronic annotated text he/she will encounter new words. The learner will try to process these words using clues in the text and the help provided by annotations. Engaging in this process involves the operation of three systems in the brain; the learner's L1 system (background knowledge), the L2 system (knowledge of the target language) and the non-verbal system (audio or visual input). The three systems come together to establish connections and make sense of both the new words and the reading text as a whole. Based on the sophistication of the process, it would be unwise to apply the Interaction Hypothesis and the Generative theory of multimedia to research on annotations without examining the role of the working memory. The following section analyses the make-up of the working memory and how it may determine the learning outcomes of reading with annotations. The concept of 'cognitive load' is also defined and discussed in light of the working memory.

The Working Memory

The working memory is a significant component of The Interaction Hypothesis and The Generative Theory of multimedia. It is also referred to as short-term memory and can be defined as "the set of mental resources that people use to encode, activate, store, and manipulate information while they perform cognitive tasks" (DeStefano and LeFevre, 2007:1618). It is "the system for temporarily storing and processing information while performing higher order cognitive tasks such as comprehension, learning and reading" (Juffs and Harrington, 2011:137). From these definitions, two key functions of the working memory emerge which are the storage and processing of information. Hence, the working memory is not memory per se in the sense of only storing information temporarily, but rather a system, with a limited capacity, which controls and regulates behaviour. For this reason, researchers are inclined to relate the working memory to cognition rather than memory (Juffs and Harrington, 2011). It is assumed that only a limited amount of information can be processed simultaneously in the working memory (DeStefano and LeFevre, 2007). Miller (1956) and Baddeley (1992) state that the working memory is limited to approximately 7 items or pieces of information at any one time. The definition of an 'item' is controversial, however, as it is possible for people to remember phone numbers and credit card numbers of more than 7 digits (items). It is, therefore, likely that the retention of information is dependent on the context or task at hand. Kirschner (2002) argues that since the working memory has the additional

responsibility of organising, comparing and processing information, it may only, in effect, process two or three items of information at once. Kirschner, therefore, makes a distinction between holding information in the working memory and processing or comprehending information. The working memory model is comprised of three factors (Juffs and Harrington, 2011):

1. The phonological loop which handles phonological and verbal information
2. The visuo-spatial sketchpad which handles visual and spatial information
3. The episodic buffer which is the area where different types of information are temporarily stored and integrated.

These three elements are governed by the ‘central executive’, a system which controls the flow of information between them and other cognitive processes (Baddeley and Hitch, 1974). If the processing of new information is successful, it will be passed on to the long-term memory (Juffs and Harrington, 2011). Henson et al. (2000) outline the basic processes thought to be involved in the verbal short-term memory from a connectionist point of view. They discuss ‘recoding’, a process in which visual items are transformed into a phonological form. The word ‘xylophone’, for example, would be transcoded into ‘zy-lo-fone’. This phonological form is then stored in the working memory where it becomes subject to loss over time or through interference. Storage is also facilitated through the process of ‘rehearsal’ whereby the phonological form is articulated sub-vocally. The last process is labelled ‘grouping’ whereby “rhythmic parsing of a sequence of items improves retention of their temporal order” (Henson et al., 2000: 426). This process would explain why people are capable of storing and recalling more than 7 or 8 items from the working memory, i.e. phone numbers (see previous paragraph).

Research suggests that individual differences in learners’ working memory may be responsible for L2 learning and use (Walter, 2004 and Leeser, 2007). Research has also shown that children have a lower working memory capacity than young adults (Fry and Hale, 1996 and Salthouse, 1994). The term ‘lower’ in this case does not refer to the storage property of the working memory per se, but the speed of processing information. In other words, young learners would need more time to process input in the working memory than adults. It is also important to note that children under the age of 7 cannot make use of storage and retrieval strategies like those aged 7-9. For children aged 10

and over, these storage and retrieval strategies become spontaneous and the children are able to use processes such as verbal labelling, rehearsal and retrieval cues to aid memory search (Rose et al., 2009).

The phonological memory of the working memory model above is of particular importance in L2 research. This is because it processes and stores, temporarily, both familiar and new phonological information (Baddeley, 1986). The most crucial role of the 'central executive' is that it controls the selective attention required to stay focused and inhibits irrelevant information or that which might distract from successful task completion (Juffs and Harrington, 2011). The central executive is, therefore, primarily concerned with controlling attention. With L2, the assumption is that the higher the working memory capacity, the more likely learning will be successful. Having more working memory will enable learners to process both form and meaning (Juffs and Harrington, 2011). If this holds true, the Interaction Hypothesis and enhancing linguistic input (Chapelle, 2007) are more likely to be fruitful for learners with a higher working memory capacity. Annotations would more likely be successful with adults as they have a higher working memory capacity than young learners. However, there may be other factors such as a learner's L1 or motivation which prove to be more influential in determining the success of learning. Ultimately, empirical evidence is needed to support any claims about whether the interaction hypothesis or the linguistic enhancement method is of more benefit to adults or those with a higher working memory capacity.

In contrast to the short-term or working memory, the long-term memory is the repository for more permanent knowledge and skills. The 'schema theory' states that knowledge is stored in the long-term memory in a network known as schemata. The presence of this network (background information) can reduce the load on the working memory (Kirschner, 2002) through top-down processing (see section 2.2). Learners' top-down and bottom-up processing skills play a part in text comprehension with the latter being mediated by working memory (Juffs and Harrington, 2011). Studies have shown that L2 learners rely more heavily on top-down processing to overcome gaps in morpho-syntactic knowledge (Clahsen and Felser, 2006 and Leiser, 2007). The Cognitive Load Theory suggests that a limited working memory is linked to an unlimited long-term memory. Hence, instruction should be designed so as not to overload the working memory capacity (Kirschner, 2002). When there is too much information for the working memory to handle, the result is a 'cognitive overload'.

The Cognitive Load

The cognitive load is defined as the mental burden that is imposed on learners from performing a task (Sweller, 1988). The cognitive load is generally classified into three types. These are discussed below and related to annotations:

1. Intrinsic load – also known as ‘information interaction’, is the inherent complexity of a task and is, therefore, not subject to change (Antonenko and Niederhauser, 2010; and Kirschner, 2002). Learning foreign words one by one, for example, is considered a ‘low information interaction’ task. The content in this case can be learned more easily than in a ‘high interaction’ task that involves the construction of sentences since it inherently requires both word and syntactic knowledge (Antonenko and Niederhauser, 2010). Although annotations in a text are meant to facilitate reading, some would argue that it involves greater information interaction than regular texts. This is because the non-linear nature of hypermedia reading requires additional mental effort from readers to synthesise and interpret input and this may adversely affect comprehension. This is debatable, however, as shall be demonstrated by the research findings in section 2.7.
2. Germane load – this type of load is relative to the individual reader, i.e. it is dependent on the reader’s prior knowledge and reading skills. Hence, it is described as a ‘variable’ load rather than the inherent or immutable nature of the intrinsic load. With the germane load, the learner’s mental effort is exerted on the construction of schemata (Kirschner, 2002) and managing the level of the load is crucial for learning (Antonenko and Niederhauser, 2010). If the learning material is not challenging enough cognitively, the germane load is low and may consequently dampen the reader’s motivation and interest in learning. Alternatively, if the material is too advanced, the germane load is high and can similarly crush motivation and impair learning. For this reason, it is vital that learning material be at an appropriate level of difficulty (Antonenko and Niederhauser, 2010). In a hypermedia reading environment, the germane load can be affected by the text itself, the number of annotated words, clarity of images ...etc.
3. Extraneous load – this type of load includes the cognitive demands that do not come about from the material itself, but from extrinsic factors. In CALL material, this can include the layout of the software, navigation requirements, or even the basic operation requirements of computers (Antonenko and Niederhauser, 2010). In other

words, the extraneous load is “the effort required to process poorly designed instruction” (Kirschner, 2002:4). An increase in this load can leave the learner feeling disoriented and distracted from the task (Antonenko and Niederhauser, 2010). In contrast, effective instructional design decreases the extraneous cognitive load while increasing the germane within the limits of the overall cognitive load (Kirschner, 2002). The importance of software design emerges again, particularly because reading with annotations is a non-linear process. The layout of a text, the design of the annotations, and the navigation process between the two ought to be as simple and clear as possible in order to reduce the unwanted extraneous load. Thus, the aim is to design CALL materials which increase the germane cognitive load while keeping both the intrinsic and extraneous loads at a minimum through proper design (Hollender et al., 2010 and Kirschner, 2002).

In light of the working memory and cognitive load, there are two problems that can arise from reading with hypermedia. The first is that this non-linear type of reading may cause disorientation, and the second is that a cognitive overload may occur (Madrid et al., 2009). In printed-text, the author attempts to create a logical argument or sequence of events for the reader to comprehend. In hyper-reading, however, the learner is the one responsible for choosing information and making connections. In order to do so, the reader needs to hold conceptual representations in the working memory while simultaneously trying to connect them with the additional linked input (Antonenko and Niederhauser, 2010). Antonenko and Niederhauser, (2010) argue that non-linear reading “places additional cognitive demands on the reader” compared to reading printed-text. They state that the problem with processing input through annotations is that the interruption in reading may hinder the learners’ ability to integrate the new concepts residing in the working memory into the long-term memory. With every disruption (annotation consultation), the new information needs to be held in the working memory while the reader continues to read the text and build connections. The overlapping of comprehension, navigation and spatial tasks (remembering visually presented shapes) may well adversely affect performance. This can result in a ‘split-attention’ effect. DeStefano and LeFevre (2007) similarly suggest that following links in hypertext may disrupt the integration process. The assumption is, therefore, that consulting external information creates an interruption of the ongoing reading comprehension process and a possible disorientation effect.

There is evidence to suggest that concerns of disorientation and cognitive overload are reasonable ones and may indeed render learning through hypermedia no more effective than reading traditional books. In a study by Lee and Tedder (2003), readers with a high working memory capacity were better able to recall the content of hypertext than those with a low capacity. DeStefano and LeFevre (2007) investigated the hypothesis that hypertext-specific activities can increase cognitive load and impair learning. The mental load imposed by task demands during hypertext reading and navigating was examined in light of the theoretical construct of working memory. It was hypothesized that choosing and following links would result in separation and hindrance to information integration. DeStefano and LeFevre (2007) predicted that the more links attached to a hypertext, the greater the cognitive load. They also looked at evidence for the effect of decision-making on increased cognitive load, i.e. the more links available, the more decisions participants would have to make. Their review included 38 studies published between 1990 and 2004. Participants were all college or university undergraduate students and all the texts (linear and non-linear) included in the review were electronic. Their review reported studies which found that reading linear texts results in better factual recall than reading with hypertext. The number of links in a hypertext was also found to be an important variable, with fewer links being more favourable for learning. For future research, DeStefano and LeFevre (2007) suggested approaching the examination of the cognitive load by looking at it at an individual level. Those with a stronger verbal working memory are more skilled at reading and comprehension and thus the researchers hypothesise that reading hypertext is more challenging for those with less working memory capacity. Since DeStefano and LeFevre's (2007) review displays results for adult participants at university level, the question arises again as to the effects of annotations on young learners. With children, stories are often read by an adult or a teacher who encourages them to construct meaning, make connections between information in the text, images ...etc. and relate this to their general or prior knowledge. The teacher can also check the learners' comprehension and provide additional support where needed. When young learners read alone, however, they will have the responsibility of doing this by themselves. The provision of annotations aims to present the learners with options that can assist in establishing connections but may, inadvertently, lead to a cognitive overload and a strain on the working memory. Indeed, Paolucci (1998) found that his 5th grade participants learned more about ecosystems

when navigating a hypertext with 50 links as opposed to their counterparts who had access to 173 links.

Despite the evidence against reading with annotations, various studies have found no such negative effect on working memory or cognitive overload. Madrid et al. (2009) set out to test the effects of two types of coherence on the cognitive load. They defined 'Text coherence' as the extent to which ideas in a text are clearly related and 'reading coherence' as the order in which learners read information in a text. The researchers tested the hypothesis that the more links available to readers, the higher the imposition on the cognitive load. This is because learners may choose links that are not entirely relevant and this incoherent order of reading may increase the cognitive load. They also tested the hypothesis that learners who receive navigational support will experience less cognitive load during link selection and reading. Their study included 42 university students who read an electronic text of 4,440 words spread over 21 hypertext pages. The experimental design compared the effects of providing 3 links vs. 8 links in addition to the effects of providing navigational support vs. no support. Madrid et al. found that the hypotheses were only partially supported. Learners with access to more links chose a less coherent order of reading but this did not impede their learning. Furthermore, the number of links and the provision of navigational support had no significant effect on their participants' cognitive load. The text coherence, however, had a direct effect on the cognitive load and learning outcomes. The higher the coherence, the less the cognitive load and the more able learners are to integrate the new information in the text with their prior knowledge.

Since studies have produced mixed findings, it is worth investigating whether there are other factors that affect the working memory and cognitive load. Researchers have looked into ways of minimizing disorientation by providing navigational support (Madrid et al., 2009) and reducing the cognitive load through the provision of 'leads'. These are advance organisers that provide the reader with a preview of what is to come, i.e. the abstract of an article or the contents page. Studies have shown that providing these leads promotes better retention of knowledge, particularly when the content is unfamiliar and when learners are less experienced in the subject matter (Antonenko and Niederhauser, 2010). Implementing leads with annotations may come in the form of a pop-up window that gives readers a preview of what will appear if they were to click on a particular annotated word (Antonenko and Niederhauser, 2010). In their study with 22

female participants aged 18-23 years, Antonenko and Niederhauser (2010) searched for evidence of the split attention effect by examining their participants' brain waves when reading hypertext. They found that the use of leads reduced the brain wave activity in their participants. The authors imply that an increase in these waves may be associated with split attention and that providing good leads can reduce the negative effects of extraneous cognitive load and split attention. It is important to highlight, however, that increased brain wave activity is not necessarily an indication of a negative effect. The Involvement Load Hypothesis (see section 2.2), for example, encourages the exertion of extra mental effort as it is thought to aid comprehension and retention of information.

As for children reading with annotations, one might argue that there is no need for leads per se as clicking on annotated words does not move readers to different pages nor does it provide lengthy information and, therefore, there is no real threat of split attention occurring. Nevertheless, every effort must be made to ensure that split attention is kept to a minimum, particularly because the cognitive abilities of young learners are lower than adults. For example, Naumann et al. (2008) argue that not all learners will benefit from strategy training because reading strategies are 'resource-demanding', especially when they are newly acquired. As a result, reading skills and strategy use both fight for working memory capacity. Hence, learners with a low working memory capacity or poor reading skills may be overtaxed. In contrast, experienced learners with efficient reading skills will have more working memory free for strategy use. The study by Naumann et al. (2008) included 64 participants at undergraduate level who read an expository hypertext with 230 annotations. Prior to this, the participants had undergone either cognitive strategy training or metacognitive strategy training or no training at all. Naumann et al. speculated that learners with a lower working memory capacity and those with poor reading skills will struggle to simultaneously comprehend the text and apply the newly-learned strategies. The results of the reading did indeed reveal that learners with a high working memory capacity had better learning outcomes from the cognitive and metacognitive strategies. On the other hand, the learning outcomes of learners with a low working memory capacity deteriorated as a result of the two types of training. Based on these findings, it cannot be said with certainty that strategy training will have a direct positive effect on young learners reading with annotations.

Aside from strategy training, there are other factors that can affect the learning outcome of reading with annotations. The type of information presented in annotations may, for

example, play a role in determining the level of interruption or distraction. An annotation with a single word translation or picture would, in theory, be less distracting than one that contains a sentence, definition, or video. This is because the processing of single-mode annotations takes less time and mental effort away from the text. Nevertheless, confusion can also occur with single-mode annotations if the visuals presented are not clear and this will, in turn, increase the cognitive load. This is the case particularly with abstract words (DeStefano and LeFevre, 2007). However, if single-mode annotations are processed faster than multiple-mode ones and are less of an imposition on the working memory then the dual-coding theory may not be appropriate for young learners or those with a low working memory capacity. Only empirical research can determine if this is the case.

Looking at studies on the design of annotated texts, it is apparent that even seemingly insignificant factors can affect the reading process. Abuseileek (2008) found that annotations in the form of a pop-up window yielded a lower performance rate on post-tests than marginal annotations. He speculated that with lower-level readers, glancing at annotations outside the reading line (pop-up) may distract from concentrating on the text. In a later study on the location of annotations, Abuseileek (2011) concluded that the closer the annotation is to the word (spatial proximity), the better it is for avoiding split attention. Similarly, Cheng and Good (2009) found that participants who had access to in-text annotations were better able to recall vocabulary than those who accessed annotations in other locations. They suggest that this was due to the ease of location. The findings of these three studies affirm 'the contiguity principle' of the GTM discussed earlier in section 2.6.2. Contrary to these findings, Chen and Yen (2013) concluded that in-text annotations were detrimental to the reading process. They did not find any significant difference in vocabulary recall amongst their 83 participants who were subjected to 3 experimental reading conditions. Their study looked at pop-up annotations, in-text annotations, and glossary ones which appeared at the end of the reading passage. They did find, however, that pop-up annotations were the most beneficial for reading comprehension. The authors argue that in-text annotations are harmful and disruptive to the reading process because they force readers to process some unessential information and consequently increase memory load. The issue with the findings of Chen and Yen (2013), however, is that both the in-text and glossary annotations had a permanent appearance on the screen, i.e. the participants had no

control over requesting their appearance. This is problematic because if there are multiple annotations on each page, participants with glossary annotations would have to browse through the list of annotations to find the corresponding meaning. Additionally, as Taylor (2009) mentioned, learners may feel obliged to read annotations that are permanently present. However, if a single annotation appears in the margin or bottom of the passage upon request, then arguably, the split attention effect is significantly reduced. With the in-text annotations, as Chen and Yen (2013) suspected, their presence caused an increase in the cognitive load. It is unfair to conclude though that this is a problem which arises from the location of annotations, but the nature of their permanent presence. The readers in this case may have felt obliged to read the in-text annotations because they were permanently present.

Given the issues and restrictions imposed by the working memory, the question arises as to whether it is possible to create a learning environment with hypermedia annotations that would overcome the potential effects of split-attention and cognitive overload. While one might argue that eliminating hypermedia annotations completely will undoubtedly reduce the extraneous load, this would also minimize the desirable germane load which is beneficial for learning. It is, therefore, imperative to draw from the theories of SLA to assist in the effective design of CALL reading material. There are a number of implications that can be drawn from the Generative theory of multimedia and Interaction Hypothesis. The restrictions of the working memory and the risk of a cognitive overload occurring are the main factors that influence the efficacy of reading with annotations. Thus, it is of crucial importance to consider the following when creating electronic texts. Firstly, the researcher needs to ensure that the design of the material is clear and simple in order to avoid increasing the extraneous load. The more complex the design is, the more likely it will increase the unwanted extraneous load. Giving a demonstration of the software before learners read will also help in this respect. Schema activation prior to learning assists in reducing the load on the working memory (Pollock, et al., 2002). By activating learners' prior knowledge, they become more prepared to tackle the new information and make connections, i.e. their germane load is enhanced. One might argue that the purpose of reading with annotations is to eliminate or reduce the pre-teaching time that usually precedes tasks. However, schema activation need not be lengthy, particularly since learners will see the new vocabulary in context. The activation can simply be a reminder of the format of a story or can be achieved by asking learners to predict the events of the story from the title page.

Summary

This section has discussed two key SLA theories that form the basis for much of the research on reading with annotations. The Interaction Hypothesis highlights the importance of getting learners to notice input in order to focus on meaning and initiate learning. The Generative Theory of Multimedia encompasses Paivio's (2007) Dual-coding Theory and calls for the presentation of input in multiple modes to aid understanding and retention of input. Since the two theories largely revolve around working memory, it was vital to take into consideration its holding capacity and its subsequent effects on learning through annotations. The suggestion throughout the literature, by and large, has been that the higher the working memory the more successful the learning. This has raised the question as to whether annotations can be used successfully with young learners who naturally have a lower working memory.

The working memory capacity will naturally vary from one individual to the next but practical implications stipulate that material ought to be designed in such a way that it will not cause a 'cognitive overload'. This is particularly important for young EFL learners or those at a lower level of proficiency. Clear and simple material designs, in addition to software demonstration, can reduce the cognitive load. Similarly, schema activation prior to reading can assist in enhancing the desirable germane load. Nevertheless, in order to determine with confidence whether reading with annotations is a help or hindrance empirical research is needed. The following section presents the findings of studies on annotations.

2.7 Empirical Research Findings

This section reviews the findings of empirical research on reading with hypermedia annotations. It discusses how different types of annotations can impact reading comprehension and vocabulary learning to varying degrees. The majority of studies in this area have been conducted with adult learners and the findings from these are presented first. Section 2.7.1 then discusses the findings from research with young learners. There are other factors that can affect the learning outcome of hypermedia reading which include learner proficiency, learner motivation and background knowledge. These are discussed in sections 2.7.2 and 2.7.3 respectively. The section

concludes with a discussion on the importance of using computer tracking devices to better understand the mechanics of reading with annotations. The findings of the studies throughout this section are analysed in light of the SLA theories discussed earlier. The literature is rich with studies that have examined the effects of annotations but only those directly relevant to this study and the GTM are presented. Studies examining incidental vocabulary learning, for example, were excluded from the discussion, in addition to those on long term vocabulary retention.

Many studies have tested the GTM and Interaction Hypothesis by designing texts with different annotation types and testing their efficacy using pre-tests and post-tests methods. Annotations can present information in a single mode such as pictures, audio, text, video...etc, or combine two or more of these modes (picture + text, audio + text... etc). Chun and Plass (1996) compared the benefits of combining two types of information in the annotations as opposed to one. Their findings indicated that the combination of text and visual for annotated words lent itself to better reading comprehension than a text-only annotation. Turk and Ercetin (2014) also found their participants who received simultaneous visual and verbal information in the annotations performed better on the reading comprehension tests than those who had access to only one form of input. Similarly, Yoshi and Flaitz's (2002) compared three annotation types: text-only, picture-only, and a combination of the two. They found that the combination of text plus picture led to better vocabulary retention. The picture only group did, however, outperform the text only group. Akbulut's (2007) group who received combined textual and visual input also performed better on the immediate and delayed tests than the group who received only textual modification. These differences were not mirrored in the reading comprehension test, however. A study based on 'noticing' (Interaction Hypothesis) by Yanguas (2009) included 4 reading groups: control, picture only, text only, and picture + text. The latter group significantly outperformed the others in terms of reading comprehension, yet no significant differences were observed between the groups in terms of vocabulary improvement. A more recent study on reading comprehension by Shalmani and Sabet (2010) yielded similar results. They also found that the picture + text group performed better on the post-test than the picture-only and text-only groups. As in Yoshi and Flaitz's (2002) study, the picture-only group also outperformed the text-only group. Shalmani and Sabet (2010) have a number of explanations for this. Firstly, they suggest that it is because visuals are more readily remembered than words in the working memory (they present evidence from other

studies to corroborate this). Secondly, they state that visuals may be better for deciphering meaning than words (Babaie, 2010). The third explanation they gave is that pictures are better at ‘grabbing attention’, and thus learners are more likely to focus on what they are reading. Shalmani and Sabet did not attribute the favourable finding of using pictures to the dual-coding effect, but this may have been an influential factor. In his study with 102 university students, Abraham (2007) found a significant correlation between looking up annotations and results on the vocabulary post-test and the reading comprehension summary task. The two experimental groups with access to annotations performed significantly better on the post-test than the control group who did not have access to annotations. The annotations in the study included L1 and L2 definition, pictures and short video clips. Although the theoretical framework for Abraham’s study was based on the GTM, it is not possible to relate his findings directly to the theory. This is because the first experimental group was given the freedom to choose any type of annotation they preferred while reading (L1 translations, pictures or videos). With the second group, the experiment was designed in a forced-lookup condition whereby students had to consult every annotated word. This in itself goes against the GTM at the core of which lies the ‘freedom’ to select, organise and integrate information. Since the dual-coding elements were not separated or controlled across groups, it cannot be said that a particular type of dual-coding allowed the experimental groups to outperform the control group on the post-test and reading comprehension. Abraham speculated that a ‘cognitive overload’ may have occurred with the forced-lookup group. Qualitative data for this shows participants stating that they faced difficulty comprehending the text because their time and energy had been devoted to consulting the annotations to understand the vocabulary. Log files from the experiment revealed that, with the choice-look up group, 81% chose to consult L1 translations. Although the two experimental groups (forced-look up and choice-lookup) outperformed the control, no significant difference was observed between the experimental groups on the vocabulary and reading comprehension post-test.

In addition to the studies on pictorial and textual annotations, researchers have examined the effects of other modes of input. Marzban (2011) wanted to test whether audio or video annotations were more beneficial for reading comprehension. His 68 participants were 13-18 years of age and were EFL learners at pre-intermediate level. They were put into two groups and read two texts of the same readability level. One group read printed

texts and had access to a regular dictionary while the other group read the electronic versions of the texts with audio and/or video annotations. Based on the results from the reading comprehension post-test, Marzban concluded that the annotations had facilitated reading comprehension. These findings are somewhat problematic though because the variables were not fully controlled. The different reading mediums (printed vs. electronic) and the two types of dictionaries are variables in themselves. Although the participants were at the same level of English, their cognitive abilities would undoubtedly be different and this may affect how they grasp the concepts in the texts and the load imposed on their working memories by manual dictionary consultation. Motivated by previous studies on GTM and annotations, Rezaee and Shoar (2011) conducted their own to examine the effect of annotation type on vocabulary learning. They set up three groups; a control group without annotations, an experimental group with picture annotations and a third group with video annotations. The results on the post-test were highest for the video annotations group and lowest for the control group. They concluded that their results were in line with the GTM since the multimedia annotations, video in particular, lead to better results on the post-test. Their conclusion that video clips are more effective than still pictures, however, is debatable. This is because the video clips for the target words were presented *after* the participants had read the text. Hence, the dual-coding information (video + text) was not presented simultaneously as in the text + picture condition. The fact that the videos were presented later (i.e. closer to the post-test) may have rendered them more retainable in the working memory. Although Mayer et al.'s (2001) experiment did not involve annotations, they found that presenting non-essential video clips before or during a task resulted in decreased performance. They speculate that the videos may have activated irrelevant prior knowledge which caused a cognitive overload and distracted their participants from focusing on the task. Al-Seghayer had conducted a similar study in 2001 and found that annotations which included text accompanied by video were more effective for vocabulary learning than annotations which presented text with still pictures. These findings were based on the results from two types of vocabulary post-tests, a recognition test and a productive one. Al-Seghayer speculated that the improvement was due to the possibility that video clips create a better mental image and increase concentration. However, unlike Mayer et al.'s (2001) study, the video clips in Al-Seghayer's study were directly related to the annotated words. The annotations included a 20-second clip which explained concrete words (i.e. a woman singing a baby to sleep to explain the word

'lullaby'). The discrepancy in the findings of the two studies indicates that the *content* of the video annotations may determine the effect they have on the learning outcome. Rezaee and Shoar's (2011) video clips were also directly related to the annotated words. Their study produced the same findings as Al-Seghayer's (2001) since the group of participants who read a text accompanied by video annotations outperformed the picture annotation and text only groups on the vocabulary post-tests. Because of these positive effects of video annotations, Rezaee and Shoar (2011) believe that using videos can lead to long-term vocabulary retention. Empirical evidence is needed though to support this claim. In fact, evidence from Abraham's (2007) study points to the contrary. A negative correlation was found between looking up video annotations and performance on post-tests. It is important to note, however, that these video annotations were not accessed frequently (4%) as the majority of participants consulted primarily L1 translations. There is also a lack of information on the content and length of the videos making it difficult to pinpoint the possible reason for the negative outcome in the study. Sakar and Ercetin (2005) also concluded that the use of annotations has an adverse effect on learning because it interrupts the reading process. They found negative correlations between reading comprehension and audio and video annotations. The fact that these annotations were 'extratextual' (i.e. providing background knowledge rather than information that is directly related to the text) may have influenced the reading process and comprehension. Ariew and Ercetin (2004) conducted an exploratory study on the effect of annotation use on reading comprehension. They examined this with two groups, advanced learners and intermediate learners. With the intermediate group, Ariew and Ercetin found a significant negative correlation between reading comprehension and time spent on video annotation and contextual graphics. Participants from the advanced group reported positively on the reading experience and perceived the annotations to be useful, but in terms of results, there was no significant contribution from annotation use to reading comprehension. After finding a significant negative relationship between reading comprehension and the time spent on video annotations, Ariew and Ercetin (2004) concluded that video annotations are a source of distraction and interfere with reading comprehension. Therefore, it appears that the length of the video extract is a variable which needs to be considered. In a study on learner preferences by Ballester and Rodriguez (2010), the majority of participants reported that they did not take to the video annotations because they were time-consuming.

The findings of the aforementioned studies, in general, fall in line with the generative theory of multimedia. Combining two types of input appears to be more beneficial for vocabulary learning and reading comprehension. In the case of a single mode of input, picture-only annotations yielded better results than text-only. The findings that are contradictory are those which come from some of the research related to video annotations. While some yielded positive results, others found a negative effect for using video annotations. Researchers attributed this to the possible occurrence of a cognitive overload which may be due to the content, purpose and length of the video extracts. The inconclusive results have prompted researchers to examine other aspects which may influence the efficacy of annotations. Some, for example, have looked at how the location of annotation affects reading comprehension and vocabulary learning (AbuSeileek, 2008; AbuSeileek, 2011; Cheng and Good, 2009; Myong, 2005), while others have looked at the role of learner proficiency and motivation (Chen and Yen, 2013; Ercetin, 2003; Yoshi and Flaitz, 2002; Ariew and Ercetin, 2004; Knight, 1994 and Myong, 2005). The findings of these studies are presented in sections 2.7.2 and 2.7.3.

2.7.1 Empirical Research with Young Learners

The studies mentioned above have focused on adult learners of English, mostly university students at the undergraduate level. While the findings provide valuable insight into the workings of the GTM and annotations, they cannot be directly transferred to young learners due to differences in working memory capacity and cognitive abilities. Research has shown that children and adults who have problems with reading comprehension also have lower performance on working memory tasks (Just and Carpenter, 1992; Oakhill, 1993; Swanson and Berninger, 1995). Thus, adding annotations to electronic texts may disfavour children since their working memory is not as developed as adults. Furthermore, some research suggests that with young learners who are beginning to read, the presence of pictures may distract from the words (Maynard and McKnight, 2001). The Focal Attention Hypothesis posits that when readers are presented with words and pictures simultaneously, their attention goes to whichever facilitates the production of a response with the least effort (Samuels, 1970). Since pictures can present objects, the claim is that learners will depend on them for meaning rather than the text and this interferes with their reading development. These claims were made in light of teaching sight words and asking learners to produce a

verbal response. In reading with annotations, the process somewhat differs as learners are not required to produce a response while reading. Annotations are also embedded within a reading context rather than isolated words and pictures as is the case with the teaching of sight words. Nevertheless, learners may resort to the pictures in annotations to comprehend the stories and neglect to focus on the words. Gyselinck and Tardieu (1999) suggest that illustrations may be easier to process than text because they are highly salient and this allows the relationships between the important elements in the text to become more transparent. To test whether the presence of images hampers the reading process, empirical evidence is essential. A review of the literature on annotations yielded only one study involving young learners and three involving teenagers. The findings of these studies are presented below. Due to the lack of research on annotations with young learners, this section draws on research with children in other areas with the aim of providing insight into the mechanisms of electronic reading. These include reading online with hypertext and reading electronic stories with other forms of modified input.

Acha (2009) tested the dual-coding theory with 135 children aged 8-9. They were students from primary schools in Spain and were learning English as part of the school curriculum. Acha used an English text of 101 words, 12 of which were annotated, and randomly assigned the students to three groups. The first group read the text with an L1 translation in the annotations, the second with a picture only annotation and the third group read with a picture + L1 translation. Acha used a vocabulary pre-test and two post-tests (immediate and delayed) to check for any improvements across the three reading conditions. Her findings were surprising because they showed that the dual-coding group (picture + translation) scored the lowest on both post-tests. In contrast the translation-only group achieved the highest scores. Acha explained that the dual-coding approach is not suitable for young learners because their working memory is limited and thus the dual-coding input forms a cognitive load. She also stated that the translation-only condition was superior to the picture-only one because the latter requires a higher working memory. Acha compared her findings to those from studies with adult learners of low verbal and visual abilities which also showed that dual-coding was ineffective. However, there are a number of factors which may have contributed to the negative findings in Acha's study. The text was presented as a whole (on one page) rather than in separate pages as in stories. This may have led to a cognitive overload, particularly since 12% of the text was annotated. De Jong and Bus (2004) concluded from their study that

the number of annotated words has a significant impact on the outcome of reading with annotations. Their findings are further discussed below. The annotations in Acha's study also appeared in pop-up windows and this may have caused a distraction as has been alluded to in Abuseileek's (2008) study on annotation location.

A study by Chang and Yang (2010) examined the cognitive load of high-school learners as they engaged in web-based activities. They found that the use of articles, notebooks and online tests led to a relatively high cognitive load. In contrast, online discussions and animations demanded a lower cognitive load from the participants and this resulted in better understanding of the concepts presented to them. Tabatabaei and Shams (2010) wanted to examine the effects of annotation type on vocabulary learning and reading comprehension. Their study included 60 older children from an Iranian junior high school for girls. The participants were placed into 4 groups and were given 3 reading passages of the same readability level to read at their own pace. The first group read the passages without any annotations (control group). The second group read the passages with text annotations, the third with picture annotations, and the fourth with picture + text annotations (dual-coding). Results showed that the annotation groups significantly outperformed the control group on both vocabulary and reading comprehension tests. With the vocabulary tests, the dual-coding group (picture + text) performed significantly better than the picture-only and text-only groups which did not differ among themselves significantly. Hence, only the dual-coding condition significantly impacted vocabulary learning. With respect to the reading comprehension test, no significant differences were observed between the experimental groups. Although participants in this study were selected according to their results on a standardised proficiency test, a pre-test for the vocabulary items would have given more validity to the findings. Additionally, the design and administration of the post-tests also jeopardised the validity of the scores. The post-tests were administered 4 days after the experiment. The vocabulary post-test presented 3 new passages that contained the same annotated words as those in the experiment; only these passages were paper-based and not annotated. Participants had to read the new passages and answer some fill-in-the-blanks questions. These were designed so that the participants would write the target words (annotated in the experiment passages) in the blanks. This means participants were exposed to the annotated words twice in different reading texts. One might argue that this set up was more favourable for the treatment groups. This is because their first exposure to the

target words was in annotated form and, therefore, they are more likely to notice these words 4 days later when they were presented in the post-test without annotation.

There is generally a lack of research on the use of annotations with young learners but there are a number of studies on children's electronic reading behaviour. These may inform the design and enhancement of hypermedia annotations. Rose et al. (2009), for example, demonstrated that the age factor makes a difference in the way children choose and process information, particularly when the input is complex. Their study tested the effects of learning cues and interactive navigational aids (maps and content lists) in Web sites for children. The findings showed that the use of maps (visual in nature) was more beneficial to participants aged 7-9, whereas those aged 10-12 performed equally well with maps and content lists. Rose et al. (2009) found that the presence of interactive learning cues on web-based materials improves the recall of information. These cues were of more benefit to the younger participants as the older ones were better able to focus on important information and more capable of ignoring the irrelevant. This is consistent with the argument put forth regarding the role of leads in reducing the cognitive load (see section 2.6.2). Although this study was not conducted in an EFL environment, there are aspects which may be of relevance to SLA and the GTM. The observation that age moderates the relationship between learning cues and learning outcomes contradicts the findings of Acha's (2009) study. Rose et al.'s (2009) study revealed that learning cues are more important for younger learners who naturally possess lower abilities than their older counterparts. The provision of these cues did not form a cognitive overload but in fact facilitated the learning process. The younger participants also benefited from a visual navigational aid with no apparent occurrence of a split attention effect.

Other research has also looked into the role of modified input and their effects on young learners' reading comprehension and vocabulary learning. De Jong and Bus (2004) also wanted to examine whether adding animations to stories would cause distraction among their participants aged 4-5. They compared and contrasted the effect of participants reading 3 electronic stories on their own and participants hearing these printed stories being read to them by adults. With the electronic stories, a picture appeared on the screen for each page and the narration was presented orally (no text appeared). The picture included interactive animations that the participants could click on while listening. De Jong and Bus found no evidence to show that the animations were a source

of distraction or interfered with comprehension of the story. What makes these findings interesting is that De Jong and Bus (2004) used animations (visual and sound effects) that were incongruous to the story. Clicking on a pot in the main character's kitchen, for example, would produce an unrelated monster accompanied by music. In theory, this is more likely to cause a cognitive overload since the readers are processing additional, irrelevant information. These positive findings on electronic reading with animations reaffirmed those of Ricci and Bea (2002) but contradicted those from other studies (Beck and McKeown, 2001; and James, 1999). De Jong and Bus (2004) speculate that it is the number of animations in a story which may affect the outcome, with an increase in animations being more likely to cause distractions and hinder comprehension. Pike et al. (2010) set out to assess the influence of different types of illustrations on children's working memory and reading comprehension. Their study included three types of reading conditions. In the first condition, the illustrations were 'consistent' in the sense that they directly support correct inferences. In the second condition, 'inconsistent' illustrations represented portions of the text that were inconsistent with the information needed to make correct inferences. The third condition was the control condition whereby no illustrations were present. Pike et al. predicted that the 73 young participants (aged 7-11) would make more correct inferences in the consistent condition and more errors in the inconsistent one. They also speculated that the pictures in the consistent condition would reduce the load on the working memory because they support correct inferences. Their predictions were confirmed as findings showed that indeed consistent illustrations had a facilitative effect whereas the inconsistent ones had an interfering effect. These effects were more pronounced with the younger participants. Hence, unlike De Jong and Bus (2004), the presence of partially irrelevant illustrations was detrimental to the reading process and a cognitive overload may have occurred. It is worth noting that Pike et al.'s experiment included text and this, coupled with the inconsistent images may have placed a heavier mental burden on the readers. In terms of annotations, the inferences that can be drawn from these experiments are that it is crucial to provide clear, representative images so as to facilitate comprehension and reduce the cognitive load. While these findings point to the role of images in establishing meaning, they do not negate the argument that learners resort to images rather than text to establish meaning and this may interfere with their reading development. For this reason, it is important that both vocabulary gains and reading comprehension be

measured side by side so as to ensure that the effects observed in one skill do not come at the cost of the other.

In addition to the studies on visual modified input, Smeets and Bus (2012) wanted to examine the effects of extra-textual vocabulary instructions in electronic stories on vocabulary learning. Their study was motivated by previous research on printed stories that showed that children learn 10 to 18% more words when texts are accompanied by extra-textual questions (Biemiller and Boote, 2006; Blewitt et al., 2009; Brabham and Lynch-Brown, 2002 and Collins, 2010). Smeets and Bus's experiment involved electronic stories in the form of videos. Like De Jong and Bus (2004), the stories used did not include any text but were embedded with multiple-choice questions for particular vocabulary items. The efficacy of these questions was tested against electronic stories that had 'hotspots' which provided short definitions of the vocabulary items. Twenty participants aged 4-5 took part in the experiment by reading 5 electronic stories in different conditions. Each story included 8 target words. The participants who read the stories with questions scored higher on the receptive and productive vocabulary post-tests than those in the control condition where no questions were present. As for the differences between reading with questions and reading with hotspots, Smeets and Bus found the questions to be more beneficial than the hotspots. They attribute this to the increased mental effort required by questions. Their findings confirm Chapelle's (2003) input enhancement principle of elaboration (see section 2.6.1) and the involvement load hypothesis (Laufer and Hulstijn, 2001). They state that the more engaging the material the more likely the learners are to retain the information. Hence, the authors recommend designing software that encourages children to 'make meaning' rather than 'take meaning'. Smeets and Bus also refer to Paivio's contiguity principle in the dual-coding theory by stating that both questions and hotspots were presented in close proximity to the content of the stories and this aided in the memorisation and learning of words. The researchers also claim that the presence of questions throughout the stories did not interrupt the flow of reading nor negatively affect the working memory and vocabulary learning. Their experiment did not, however, include a reading comprehension component which would have provided a better indication of whether the embedded questions affected the working memory and overall comprehension of the stories.

More general studies on electronic reading processes looked at the differences between reading electronic material and printed material. In a study by Greenlee-Moore and

Lawrence (1996), the grade 4 participants who read narratives from the computer achieved higher reading comprehension scores than those who read the same narratives in print. This was only the case with long narratives though as shorter ones yielded no significant differences between the two forms. This suggests that with longer texts, electronic reading may be more capable of sustaining young reader's interest and attention which in turn facilitates comprehension. Coiro and Dobler (2007) examined the difference between reading printed text, closed hypertext systems, and open-ended internet environments. Their participants were eleven 6th graders from public schools in America. Although their study does not explicitly discuss the GTM, part of their theoretical framework is closely related. They describe reading as an active, constructive, meaning-making process which requires multiple strategic cognitive processes for selecting, organizing, connecting, and evaluating the reading material. These strategies include drawing connections and making inferences. Coiro and Dobler (2007) also highlight the need for readers to use their existing knowledge to "more clearly understand new ideas encountered within texts...to make predictions about what might come next...and to reason strategically when they encounter barriers to comprehension" (Coiro and Dobler, 2007: 217). The results from their qualitative study suggest that the processes involved in reading online texts are both similar to printed-text, yet more complex. They found that the skilled readers used their previous knowledge to decode meaning and their inferential reasoning to make use of structural and contextual cues. They then used a self-regulated reading process to monitor and select relevant information. Thus, electronic texts can be more captivating than printed texts (Greenlee-Moore and Lawrence, 1996) yet require more complex processing strategies (Coiro and Dobler, 2007:217).

2.7.2 Learner proficiency

The inconclusive findings of studies on annotations have prompted some researchers to examine whether there is a relationship between learners' language proficiency and efficacy of annotations. In other words, do learners of higher or lower proficiency level benefit more or less from annotations? If so, how does this fit into our understanding of the Interaction Hypothesis and GTM?

The argument put forth earlier in section 2.4.2 centred on the notion that the use of annotations renders a vocabulary threshold less significant. If this is the case, then

learners at a lower level are more likely to make-up for their lack of vocabulary knowledge through annotations, but at a higher level the use of annotations may become less significant. Indeed, the participants at intermediate level in Ercetin's (2003) study accessed various types of annotations more frequently than their advanced counterparts. Knight (1994) found annotations to be more advantageous for the low ability participants as they were able to learn nearly as many words from an electronic text as their high ability counterparts. Their lack of vocabulary was, therefore, compensated through annotation use. It is important to highlight, however, that frequency of consultation does not always equate with a better learning outcome. While the intermediate proficiency participants in Ercetin's (2003) study consulted annotations more frequently, the advanced group performed better on the reading comprehension test. Other research shows the reverse with annotations being more advantageous for higher proficiency learners. Chen and Yen (2013) carried out a study to test whether annotations were more useful for a particular level of language proficiency. They recruited 83 participants at university undergraduate level with an age range of 18-22. The participants took a diagnostic proficiency test to determine the L2 level of reading comprehension and were placed accordingly into high, medium, or low level proficiency groups. Chen and Yen used 16 reading passages to overcome any effects of prior topic knowledge. Participants read 4 passages over a period of 4 weeks. A vocabulary-recognition post-test was given after every passage. It tested both short-term memory of words that appeared in the passage the participants had just read, and long term retention for words from the passage they had read during the previous week. The vocabulary results of the experiment showed that the presence of annotations aided both long-term and short-term retention for medium and high proficiency groups, while no effect of annotations on the low proficiency group was observed. The researchers, however, acknowledged that it would have been useful to include a tracking tool in their research (see section 2.7.4) to gather more information about their participants' use of annotations. Myong's (2005) study investigated the effects of annotations on reading comprehension and found a correlation between learner proficiency and learning outcomes. Participants from the annotation group achieved a significantly higher score on the reading comprehension post-test than the control group. This achievement, however, was only relative to the experimental group that had access to L2 annotations, rather than L1. Qualitative data also revealed that 62% of participants preferred L2 annotations. Myong speculated that the success of L2 annotations is dependent on the proficiency level of the participants.

Those who preferred L2 glosses had higher mean scores on the reading comprehension test than those who favoured L1 glosses. Although Myong did not place her study within a GTM theoretical framework, her findings are, nevertheless, interesting because they point to the success of a ‘mono-code’ form of computer-based learning (L2 text-only annotations). Her results cannot be regarded as conclusive evidence, however, because the reading comprehension post-test had a low reliability level.

Similar to Myong (2005), Ballester and Rodriguez (2010) conducted a series of four pilot studies to examine whether learners’ proficiency level plays a part in their preferences for annotation type. The texts in the studies gave participants the option to choose the type of annotation they preferred. For each annotated word a picture appeared by default with four options for learners to choose from: L1 translation, synonym, pronunciation or video. Participants’ preferences differed across the studies and Ballester and Rodriguez speculated that this was due to learners’ different proficiency levels. Participants in Study 2 were of a higher proficiency and preferred the synonym option, whereas those in Study 3 of a lower proficiency preferred the L1 translation. These findings echo those of Myong (2005). Similarly, the intermediate and advanced groups in Ercetin’s (2003) study both preferred textual annotations over visual or audio ones. In terms of location, Chen and Yen (2013) found that participants of low and medium proficiency levels preferred in-text annotations, whereas high proficiency participants preferred glossary ones. Their reasons for this were not stated, however.

Not all research on annotations and proficiency level, however, has arrived at the same conclusion. An earlier study by Yoshi and Flaitz (2002) involved learners at beginner and intermediate level. The findings revealed that there was no significant interaction between annotation type and proficiency level. In an exploratory study by Ariew and Ercetin (2004), participants were placed in two groups according to their proficiency level. The aim was to test whether the efficacy of annotations for reading comprehension is dependent on proficiency level. A tracking tool was used which recorded the annotation type, amount of time participants spent viewing the annotations and the number of times each annotation was accessed. The findings showed that consulting annotations did not have an effect on reading comprehension. Advanced learners, nevertheless, performed significantly better on the comprehension test but they did not rely on annotations for this. This indicates that participants may have employed

successful inference strategies, thus implicating the presence of a relationship between reading skills and a vocabulary threshold.

In summary, research on proficiency level and annotation use has produced mixed findings. Some research suggests that annotations are more successful with medium and high proficiency level learners (Chen and Yen, 2013 and Myong, 2005) while others have found annotations to be more favourable for low proficiency learners (Ercetin, 2003 and Knight, 1994). In some cases, no correlation was observed between annotations and proficiency levels (Yoshi and Flaitz, 2002; Ariew and Ercetin, 2004). To understand the cause of this discrepancy it may be of benefit to relate the findings to the GTM. In theory, high proficiency learners are more capable of understanding a text by using inferencing strategies and this would free up their working memory to process the information, if need be, from annotations effectively. Although low proficiency learners may consult annotations more frequently, due to their low vocabulary threshold, this may not result in the correct uptake of information. Lower proficiency learners may be more likely to experience a cognitive overload from trying to understand the text and simultaneously process the information in the annotations. This is more likely to occur with video annotations as they may be a bigger source of distraction. Hence, the proficiency variable may be related to the annotation type variable.

The studies above on proficiency were carried out exclusively with adult learners. It is worth researching whether the variables that come into play also affect young learners. Is the outcome of reading with annotations dependent on their proficiency level? If so, is this governed by the annotation type and its consequent effects on their working memory? Will their preferences for annotation type be similar to those of adults (L2 and textual annotations over audio and visual ones)?

2.7.3 Learner interest and prior knowledge

It is often thought that technology is attractive and will engage learners, but research into learning and motivation in CALL has not always concurred (Moos and Marroquin, 2010). Learners' attitude and preferences must be taken into account in order to effectively develop CALL material and enhance language learning. Mckenzie (2009) points out that advancement in reading technologies are of particular importance to young learners nowadays as they are growing up in an electronic era. They may find

traditional approaches boring and prefer electronic reading that is both attractive and engaging. This section looks at ‘interest’ and ‘background knowledge’ as two factors that can influence learners’ attitudes towards CALL material and subsequently the learning outcome. These factors were chosen for discussion because their influence also extends to the working memory. The findings from research with annotations are presented alongside those from other CALL-related research.

Interest

Renninger et al.’ (1992) classify ‘interest’ as either individual interest (personal preferences), or situational interest (an interest evoked by text features, visual and/or auditory stimuli). Both types of interest are deemed influential for deep processing of a text and the recall of information. Researchers have examined whether the use of electronic material increases interest. After observing the success of e-books in the L1 context, Huang (2013) decided to conduct an exploratory study to look at students’ perceptions of using an e-book for L2 reading. She found a previous study by Anuradha and Usha (2006) showing e-books to be popular with the majority of teachers and students, both new and accustomed users. Her one-year study (2013) comprised 67 EFL first year university students who were at an intermediate level of proficiency. They read 1 book per week and turned in reading logs after every book. Participants also sat for semi-structured interviews at the end of the study. The overall findings of the study were positive, especially for the functions that were available to the participants. They liked the online dictionary function and the read-aloud option which, according to students, facilitated comprehension. It appears that the use of electronic-specific features positively impacts interest. L2 learners from a different study reported negatively on their e-reading experience stating that more functions were needed for improving reading such as highlighting and attaching notes (Lam et al., 2009). In a study by Zhu (1999), participants who read a hypertext with fewer links rated the reading experience more positively than those who had access to more links. Although Grimshaw et al. (2007) found no differences in enjoyment between reading e-books and printed ones. Nevertheless, certain electronic features had a positive influence on the young participants’ (aged 9-11) reading experience. These include the provision of an online dictionary which was accessed greater than in printed form, and a verbal narration which resulted in significantly higher reading comprehension scores. A synthesis of 20 studies looking into the effect of the ‘interest’ factor on motivation, revealed mixed results but

some research suggested that the use of hypermedia is related to an increase in interest (Moos and Marroquin, 2010).

While evidence suggests that aspects of CALL material can increase interest, this should not automatically be construed as an improvement in learning. Huang's study above yielded positive results, but the impact of e-reading on learning in this case cannot be verified due to the lack of a pre-test and post-test procedure for the features available. A tracking device (see section 2.7.4) would have also provided information as to whether students had actually used the functions available to them. This type of data is crucial as often there is a discrepancy between students' perceptions and their actual use of technology. Huang's study was also conducted in a social setting (an online discussion forum) and this may have contributed to the positivity that the participants felt. Another issue is that, in many cases, participants' opinions on hypermedia reading are gathered after a single exposure to a text. This calls into question the 'novelty effect' on interest which may diminish over time (Moos and Marroquin, 2010).

Prior knowledge

While individual interest and situational interest (that which is induced by text features and stimuli) can evoke readers' interest in a text, it is also necessary to examine the reverse, i.e. the role of prior domain knowledge in increasing motivation and interaction with the CALL material. It is argued that since hyper-reading is not as structured as printed text, there is a greater need for prior knowledge (Moos and Marroquin, 2010). This knowledge works hand-in-hand with topic interest to facilitate learning. The overall findings of 8 synthesised studies in Moos and Marroquin's (2010) review demonstrated that learner interest is vital because it facilitates the navigation decisions that need to be made in hypermedia reading but once again this is dependent on learners' prior domain knowledge. A learner with high interest but low domain knowledge will struggle to make decisions and the learning outcome will, therefore, suffer (Moos and Marroquin, 2010). Ercetin (2010) focused on the role of prior knowledge and topic interest in facilitating reading comprehension. She enrolled 54 participants who were undergraduate students at university with a proficient level of English. A hypermedia text with a tracking tool was used and annotations were available to help with words and to provide background information on the topic. The control text included a navigation map that organized the pages in a hierarchical manner. Immediately after reading,

participants had to write down everything they remembered from the passage. This was scored by two independent raters with high inter-reliability. The prior knowledge test was administered in English two days prior to the experiment and contained five open-ended questions about the topic. The ‘topic interest’ test was also given two days before the experiment. Semi-structured interviews with 10 participants were conducted immediately after the reading task. The results of this experiment were that interest and prior knowledge are two independent variables with a weak relationship. Prior knowledge did not have an effect on recall, whereas topic interest did. The presence of a navigation map in the control condition, however, may have affected the results. Having this type of lead may have made the text more organised and accessible to participants and thus increasing the situational interest. Additionally, the experiment involved one reading text only which raises the question of the novelty affect. Ercetin and Ariew (2004) found a positive relationship between prior knowledge and reading comprehension for their intermediate group participants. As for the advanced group, the relationship was significant and positive but weak.

With young learners, it is important to examine how background knowledge may manifest in the non-expository texts they read such as fictional stories. Background knowledge may involve simply being accustomed to reading, being familiar with a story layout (genre), or understanding the concept of fiction. One can assume that the higher the learners’ interest and background knowledge, the less likely they are to be affected by a cognitive overload. Background knowledge can also be built up from repeated exposure to hypermedia texts. Over time, it will facilitate navigation and allow learners to consult annotations effectively. Establishing whether there is a correlation between learners’ attitudes towards hyper-reading and their performance, nevertheless, requires empirical evidence. The introduction and use of software cannot simply be based on learners’ preferences without the support of results. The following section discusses the importance of using tracking devices as a method of collecting data that can substantiate learners’ and teachers’ claims about the efficacy of software.

2.7.4 CALL tracking

“The use of the computer to deliver instruction and collect data opens the possibility for more precision and control in data collection” (Doughty, 1987 in Chapelle, 2009: 743).

Using tracking for CALL material is not only useful for checking what learners are doing, but is also valuable for investigating SLA principles such as input enhancement and negotiation of meaning (Fischer, 2007). Claims may be made about the instructional value of software material and those claims must be empirically supported (ibid). Learner preferences and perceptions about the benefit of particular software can also be checked using CALL tracking. This is important because research has shown that learner preferences may not always be consistent with their performance results (Abuseileek, 2008; Ariew and Ercetin, 2004 and Cheng and Good, 2009). Learners may mistake convenience for effectiveness (Chen and Yen, 2013). As a result, it is necessary to check if, and how, students are using the technology available to them. There are often discrepancies between what learners think they do and their actual behaviours and this renders CALL tracking more objective than student surveys or feedback (Fischer, 2007).

Computer-based tracking is regarded as a type of ethnographic research (Fischer, 2007). Researchers can use audio or video recording of students at work, but computer-based trackers are less obtrusive. This makes them a more suitable instrument for collecting data from children and may be more appropriate for some cultures. If used without the participants' knowledge, these trackers can reduce the chances of altered behaviour that may come from participants' awareness of being monitored. The ethical issues surrounding such use of trackers are discussed in section 3.11 as they are pertinent to this research. Fischer (2007) also points out that students may skip chunks of CALL material for the sake of "getting it over and done with". The probability of this occurring is likely to increase if the CALL task is not assessed as participants may not feel the need to perform well.

Summary

The outcomes of research on annotations are somewhat inconclusive but section 2.7 has shown that this is largely attributed to the diverse research approaches that were employed. Not all the studies discussed were carried out within an interactionist framework or in accordance with Mayer's GTM. The many different variables across these studies made it initially difficult to pin-point the specific elements responsible for the mixed outcomes. Nevertheless, it was through the SLA theories that interpretation of some mixed findings was possible. Mayer's (1997) generative theory of multimedia may

explain why particular annotations were more successful than others. Some researchers presented a single mode of information in the annotations while others used a combination of modes. Annotations that include more than one mode of input (i.e. text + picture) appear to be more beneficial for text comprehension and vocabulary learning. This falls in line with the dual-coding aspect of the GTM. In contrast, presenting too much information may cause a ‘cognitive overload’. This may have been the case with video annotations posing a strain on the participants’ working memory which distracted them from reading, or resulted in a cognitive overload for some. Some studies were designed so that the annotations appear in the margins, some as pop-up windows and others presented them in-text. Findings corroborated The Contiguity Principle of the GTM by showing that the nearer the annotation to the word the less distraction it is likely to cause. The literature has also demonstrated that there are other factors that play a part in determining the learning outcome of hypermedia annotations. These include learner proficiency, interest and prior knowledge. These two factors were discussed because of their possible effects on working memory and annotation use. From a theoretical perspective, learners of a higher proficiency level will have more ‘free’ working memory to process annotations while those at a lower level are more susceptible to distraction. Empirical research showed that learner proficiency may indeed govern the success of reading with annotations (Chen and Yen, 2013) and the type of annotations preferred by learners (Myong, 2005 and Ballester and Rodriguez, 2010). Intermediate and advanced learners may benefit more from annotations and are more likely, if given the choice, to select L2 annotations. Having an interest in a particular topic and/or background information can reduce the cognitive load which results in better utilisation and benefit from reading with annotations.

A review of the literature revealed a lack of information on how reading with annotations affects young learners. A study which directly examined the dual coding theory with young learners (Acha, 2009) conflicted with those involving teenage participants. Other research on electronic reading, though not conducted within a GTM framework, suggests that young learners benefit from interactive features such as those they can call up or click on in electronic texts. The working memory and cognitive load, in particular, differ considerably between adults and children. For these reasons, we cannot say with confidence that annotations will benefit or hinder young learners’ reading comprehension and vocabulary learning. Section 2.7 concluded with the importance of using computer tracking devices to provide a more objective view of how students use

CALL software. The differences in learners' perception and performance make it unreasonable to use qualitative data alone to inform the design of software, determine its success, or substantiate theoretical claims.

2.8 The Rationale and Focus of this Thesis

This PhD thesis is motivated by previous studies on hypermedia annotations, particularly those centred on Mayer's generative theory of Multimedia and Paivio's dual coding theory. An overview of the literature, however, revealed that the primary focus of these studies has been placed on adult participants, mostly those at university level. The scant information on annotation use with young learners means it is difficult to establish with any certainty whether this learning tool is beneficial for them. The disparity in cognition, level of English and reading behaviour make it unreasonable to generalize the findings of studies with adults to children. By conducting research with children, one can obtain more robust results and contribute to understanding the workings of Mayer's theory and the Interaction Hypothesis within a CALL environment. Since children possess a lower working memory than adults, this research aims to provide a new dimension to the relationship between annotation consultation and cognitive overload. Although the research does not explicitly test the learner proficiency and motivation variables, the findings from the research on this have informed the design of this study. The same applies to the use of a CALL tracking device. As for the practical reasons for this research, the findings aim to provide a deeper understanding of how to enhance the electronic reading experience for young learners. This will ultimately shape teachers' and software writers' design of TESOL reading material for young learners.

Chapter 3. Methods

The previous chapter discussed the theoretical framework and literature related to using hypermedia annotations for L2 reading comprehension and vocabulary learning. This chapter focuses on the research questions, data collection instruments and experiment procedures. Since this research uses a deductive approach, the literature and theories in Chapter 2 were used to generate hypotheses. The rationale for the design of the experiment's instruments and procedures also stems from the findings presented in Chapter 2 and this is highlighted throughout.

This chapter also discusses the pilot study and describes how the research instruments were developed and modified for the research. Since this research is experimental, a number of variables are present in the study design, the instruments, participants, and procedures. Hence, it was important to devote a section to explaining the measures taken to control these variables along with how validity and reliability were maintained. This chapter concludes with the ethical issues surrounding the research and the steps taken to ensure that guidelines were upheld. Details of data analysis and the statistical tests will be provided in Chapter 4.

3.1 Research Questions

Do hypermedia annotations facilitate reading comprehension and vocabulary learning for young EFL learners?

In order to answer this overarching research question, a number of questions and hypotheses were posed. The justification for using hypotheses and their formulation are outlined in section 3.3.

R-Q 1: Do hypermedia annotations facilitate vocabulary learning for young Kuwaiti learners?

Hypothesis 1: Consulting annotations has no effect on immediate vocabulary recall

R-Q 2: Do hypermedia annotations facilitate reading comprehension for young Kuwaiti learners?

Hypothesis 2: Consulting annotations has no effect on reading comprehension

R-Q 3: How do young Kuwaiti learners use annotated texts and how do they feel about this type of reading?

The effects of hypermedia annotations are sometimes examined in light of only one language skill, i.e. vocabulary learning or reading. A review of the literature, however, brought to light two arguments that can only be addressed by examining both skills simultaneously. Firstly, it was argued that young learners are likely to become dependent on images, rather than text, to create meaning because images are more salient and require less mental effort. Dependency on images may consequently stunt reading development. It was, therefore, necessary to include a textual vocabulary test in this experiment to establish whether learners pay attention to written forms of words rather than simply looking at the annotations to understand the story. The second argument from the literature pertains to the working memory. Although consulting annotations is brief, in contrast to regular dictionary consultation, it was argued that the disruption to the flow of reading may, nevertheless, influence overall comprehension. Thus, focusing exclusively on any vocabulary gains that are made may come at the cost of checking whether consulting the annotations has jeopardised the overall comprehension of the text. For these reasons, it was seen fit to test both skills in order to reach a better understanding of how annotations may influence young learners.

In addition to examining the general effects of hypermedia annotations on the two skills, the study also uses a within-subjects design to determine whether a certain *type* of annotation yields better results. The ‘type’ of annotation here refers to the nature of the information it includes. The efficacy of the dual-coding approach is tested by comparing two types of annotations. One annotation type includes a combination of picture + L1 translation and the other annotation type combines picture + L2 pronunciation. Hence, two sub-questions were posed for research questions 1 and 2:

Sub-Q 1: Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for L2 vocabulary recall?

Hypothesis 1.1: There is no effect of annotation type on immediate vocabulary recall

Sub-Q 2: Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for reading comprehension?

Hypothesis 2.1 There is no effect of annotation type on reading comprehension results

The third research question aims to bring in the voice of young learners to the existing literature on vocabulary learning and reading comprehension with annotations. This is vital because increasing numbers of children are using technology nowadays and learning tools are more likely to be successful if the target audience find it appealing and motivating. The quantitative data (log files) and qualitative data (focus groups) gathered for this third research question are also used to aid with the interpretation of quantitative data from Question 1 and 2. Details of this are discussed in section 3.7 under research instruments.

3.2 Research Paradigm: Positivism

The term ‘epistemology’ is simply defined as the philosophy of obtaining knowledge (Trochim, 2000). It is closely related to ontology and methodology. While ontology is concerned with the philosophy of reality, epistemology looks at how that reality is formed (Krauss, 2005). Methodology then identifies the specific practices used to attain knowledge of that reality. ‘Positivism’ is a branch of epistemology which calls for the use of “...observation and reason as means of understanding behaviour” (Cohen et al, 2011: 7). It uses a deductive approach to obtaining knowledge by generating hypotheses which can be tested objectively (Bryman, 2012). Hence, knowledge is gained through empirical means of observation and experimentation (Cohen et al., 2011).

Since objectivity is central to a positivist approach, quantitative analysis is seen as fit for capturing reality (Cohen et al., 2011). This is not to say, however, that qualitative methods are entirely dispensable. In fact, researchers advocate the use of them in

conjunction with quantitative analysis (Bryman, 2012), particularly because positivism cannot fully reflect the complexity of human nature and social phenomena (Cohen et al., 2011).

This study operates within a positivist paradigm by using a deductive approach to testing the GTM. Hypotheses were first posed for the effects of annotations on immediate vocabulary recall and reading comprehension for young EFL learners. Both quantitative and qualitative research instruments (mixed method approach) were then designed and procedures set in place in order to test these hypotheses.

3.3 Experimental Research

The key feature of experimental research is the control and manipulation of conditions by the researcher in order to determine whether the introduction of a particular intervention has an effect (Cohen et al., 2011). An experiment examines the effect of one variable – the independent variable – on another variable known as the dependent or outcome variable.

There are many other variables that may affect the outcome of an experiment and every effort must be made to control them. This study devotes a section to explaining how the variables of the experiment were controlled (section 3.6). It is impossible, however, for a researcher to control all the variables and *randomization* is, therefore, critical. This involves the random sampling and random allocation of participants to either a control or experimental group (Morrison, 2009). Since random allocation encompasses both observed and unobserved factors, the control of the latter is rendered unnecessary (Schneider et al., 2007). Randomization and other means of controlling the variables were undertaken in this study and these are discussed throughout the following sections. Experimental research also involves the formulation and testing of hypotheses. When a researcher predicts that an effect is present, this is called an ‘alternative hypothesis’ (H_1) (Field, 2009). There is another type of hypothesis, however, known as the ‘null hypothesis’ (H_0) where the researcher assumes no effect is taking place. The reason the latter is used with statistical analysis is because a researcher cannot prove an effect by using statistics alone, and thus it is sounder to reject the null hypothesis. That is not to say, however, that rejecting the null hypothesis automatically entails the presence of an effect, rather it supports the possibility of one (Field, 2009).

3.4 Participants of the Study

The selection of individuals for research is imperative to its quality (Cohen et al., 2011). Yet, it is not feasible for the researcher to gather data from an entire population and thus careful selection of a smaller, representative group is required. There are two main sampling strategies: probability sampling and non-probability sampling (Bryman, 2012). Probability sampling involves a random selection approach whereby individuals have an equal chance of being selected. This approach allows the researcher to make generalisations about a population (Cohen et al., 2011). Non-probability sampling, on the other hand, refers to all other types of sampling that do not conform to the principles of probability sampling (Bryman, 2012) and therefore, aims to only represent a specific group (Cohen et al., 2011). After deciding on the appropriate sampling method, the researcher then needs to determine the sample size. While there are no hard and fast rules about sample size, Cohen et al. (2011) suggest a bare minimum of 30 participants per variable if the researcher intends to use any statistical analysis on the data. Sample size is also governed by time, cost and heterogeneity of population (Bryman, 2012). The more heterogeneous the population, the larger the required sample will be in order to adequately reflect the population.

This study used a probability sampling method in the selection of schools and participants. There are six districts in Kuwait and one district was randomly selected during the process of gaining permission for school entry from the Ministry of Education. The schools within the district were notified of the research and two public schools were randomly selected from those willing to participate. From the two schools, 112 young Kuwaiti EFL learners, 52 girls and 60 boys, from six classes chosen at random, took part in the experiment. They were all students at year 4. Students study five years in primary school and English is taught as a compulsory school subject from year 1. Students have English classes every school day (5 times a week) with each lesson lasting a period of 45 minutes. The school books used at schools is the Longman's Series 'Fun with English'. Students in years 1 and 2 are taught basic listening skills and word recognition. They begin reading and writing in year 3. Appendix V includes an excerpt from the curriculum objectives of year 3 to clarify what the participants at Year 4 in this study have previously completed. The 112 participants came from six Year 4 classes, three classes from the girls' school and three from the

boys'. The six classes were randomly assigned to the order in which they read the stories over a period of 3 weeks (see table 3.7).

Table 3.4: Random assignment of participants

Classes	Week 1	Week 2	Week 3
Class 1 (16 girls)	Treatment A Electronic text with picture+L2 pronunciation	Treatment B Electronic text with picture+L1 Translation	Control Electronic text <u>without</u> annotations
Class 2 (18 boys)			
Class 3 (19 girls)	Control Electronic text <u>without</u> annotations	Treatment A Electronic text with picture+L2 pronunciation	Treatment B Electronic text with picture+L1 Translation
Class 4 (19 boys)			
Class 5 (17 girls)	Treatment B Electronic text with picture+L1 Translation	Control Electronic text <u>without</u> annotations	Treatment A Electronic text with picture+L2 pronunciation
Class 6 (23 boys)			

(Table 3.4 The left column shows the distribution of the 112 participants from 6 classes. The consecutive columns show the reading conditions they were exposed to every week. For purposes of analysis, classes 1 and 2 were labelled group 1. This is because both classes were exposed to the same reading conditions and in the same order. Similarly, classes 3 and 4 were labelled group 2 and classes 5 and 6 were labelled group 3)

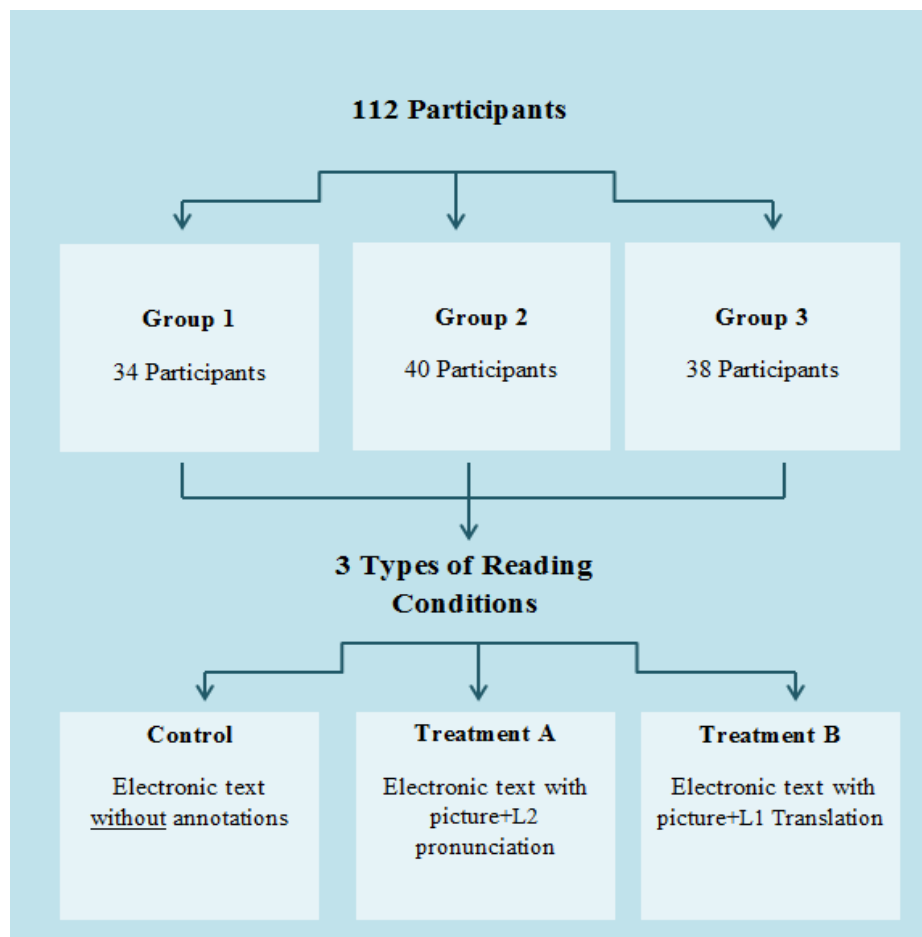
A sample size of 100+ was deemed fit for the experiment, particularly since the group of participants was largely homogenous. All participants were at year 4, they were all Kuwaiti nationals and all shared a mother tongue of Arabic. In addition to homogeneity, there were additional elements which needed to be considered when determining sample size. The participants were of mixed abilities because classes in public schools are not classified according to student performance during the previous year. Each class, therefore, included students who passed the previous year with an award of excellent, very good, good or pass. By choosing a sample size of over 100 participants, differences in level of English which may affect performances on the post-tests are better controlled. Students in Year 4 range between 9-10 years of age. Data from the male and female participants were collected separately from two schools. This is due to gender segregation of public schools in Kuwait. Each class has a maximum of 30 students of mixed abilities. Students at public schools also have a Computer Science subject which

is taught to all grades once a week. The computer labs are equipped with enough PCs for students to be able to work individually. In addition to the Computer Science subject, classes are taken to the lab for content-based lessons such as science, math, English... etc. Hence, participants are accustomed to using the computer lab to develop their computer skills and subject knowledge.

3.5 Study Design

This study used a repeated measures experimental design in which participants were randomly allocated to the order of intervention they were exposed to. This design was chosen as it offers considerable control potential (Cohen et al., 2011). Figure 3.5a below outlines the groups of participants and the reading conditions.

Figure 3.5a: Outline of the research design.



A ‘counterbalancing’ method was used whereby each group of participants was exposed to three types of reading conditions, in an alternating fashion, over a period of 3 weeks.

This was done with the purpose of reducing the bias that could result from the order of condition in which participants were exposed to. For example, during the first week, each group read the same electronic story but in a different condition. All groups read the story ‘Platypus’ by Chris Riddell but Group 1 read it in the Treatment A condition, Group 2 in the control, and Group 3 in the Treatment B condition. In the second week, the groups read a new story but in a different condition to that of the previous week and the same applied to the third week. By the end of the experiment each group had read 3 stories in 3 different conditions (see table 3.4a).

Table 3.5a: Repeated measures experimental design

Stories → Groups ↓	Week 1 ‘Platypus’ by Chris Riddell	Week 2 ‘The Very Hungry’ Caterpillar by Eric Carle	Week 3 ‘The Tiger who Came to Tea’ by Judith Kerr
Group 1	Treatment A	Treatment B	Control
Group 2	Control	Treatment A	Treatment B
Group 3	Treatment B	Control	Treatment A

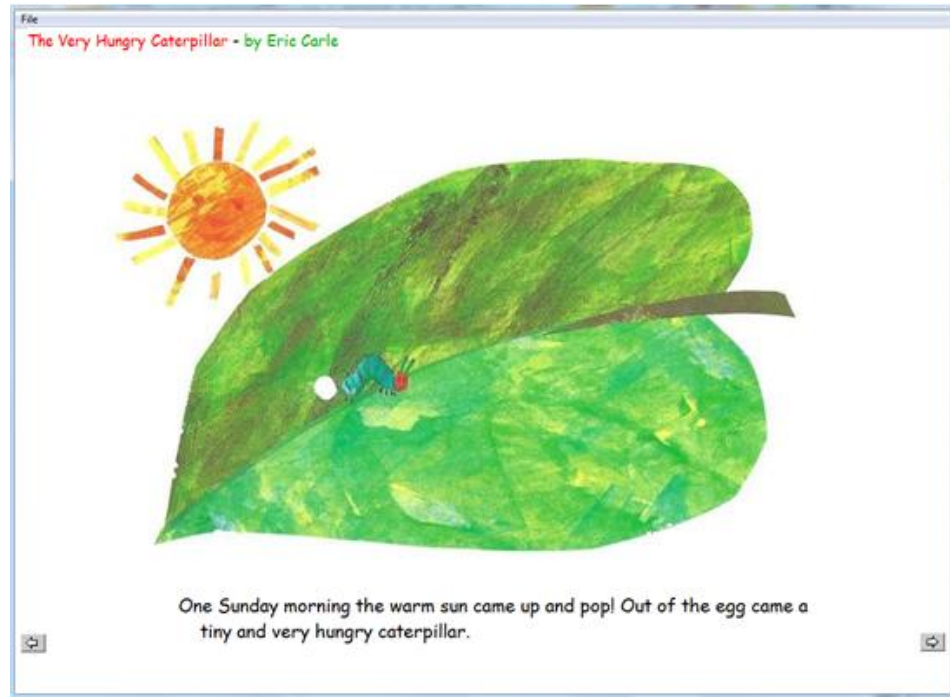
A within-subjects design was set up for comparing performances on the vocabulary and reading comprehension post-tests. The two language skills, vocabulary and reading, were placed as the dependent variables while the reading condition formed the independent variable (table 3.5b).

Table 3.5b: The independent and dependent variables of the experiment

	Dependent variables → Independent variables ↓	Vocabulary post- test scores	Reading comprehension scores
Reading Condition	Control		
	L1 translation +picture		
	L2 audio +picture		

In the control condition, the groups read an electronic story without the presence of any annotations (figure 3.5b).

Figure 3.5b: Sample of the control reading condition



(Figure 3.5b shows a sample page from the control reading condition for the story 'The Very Hungry Caterpillar'. The text appears in black with no highlighted words or annotations)

In the Treatment A condition, however, the electronic text included highlighted words (in blue) that were linked to hypermedia annotations. When a participant clicks on a highlighted word, an annotation appears containing a picture to illustrate the meaning of the word and an audio L2 pronunciation of the word (Picture + L2 Pronunciation) (figure 3.5c).

Figure 3.5c: Sample page of the Treatment A reading condition



(Figure 3.5c shows the Treatment A condition for the story 'The Very Hungry Caterpillar'. The target words are highlighted in blue and annotations appear to the right of the screen. When placing the cursor on the word 'egg', for example, an annotation appears that includes a picture + audio pronunciation. This annotation disappears once the cursor is moved away from the highlighted word)

Similarly, the text in Treatment B included highlighted words with hypermedia annotations. Clicking on a highlighted word also produces a picture to illustrate the meaning of the word, as in Treatment A, but does not offer an audio pronunciation. Instead, the L1 written translation of the word is presented (Picture + L1 Translation) (figure 3.5d).

Figure 3.5d: Sample page of the Treatment B reading condition



(Figure 3.5d demonstrates how the Treatment B condition provides a picture + L1 translation for the annotated word 'egg')

3.5.1 Texts and target words

The process of choosing stories for this experiment was carried out on the basis of 'authenticity' and shared level of 'difficulty'. Authenticity involves choosing stories that are not written specifically for EFL learners. Authentic materials may be problematic if the themes are specific to a particular culture, but the stories selected for this experiment represent universal themes such as the cycle life, love of exploration, greed, hunger...etc. The aim was to create a more natural, realistic reading environment for the young learners and contribute to the Kuwaiti government's aim of encouraging students to read authentic English stories (see section 1.2). In terms of difficulty, the Powers-Sumner-Kearl Readability Formula was used to determine text difficulty. Unlike the more commonly used scales, such as Flesch Kincaid and Fry Graph, this scale was chosen because it is designed specifically to measure the level of texts for young learners aged 7-10. After examining a number of stories, the ones that measured 'Grade 4' on the readability formula were selected. Two EFL teachers with experience of teaching at the participants' level then confirmed that the stories were suitable for the age group. The average length of the stories is 249 words per story. Copyright permission was then

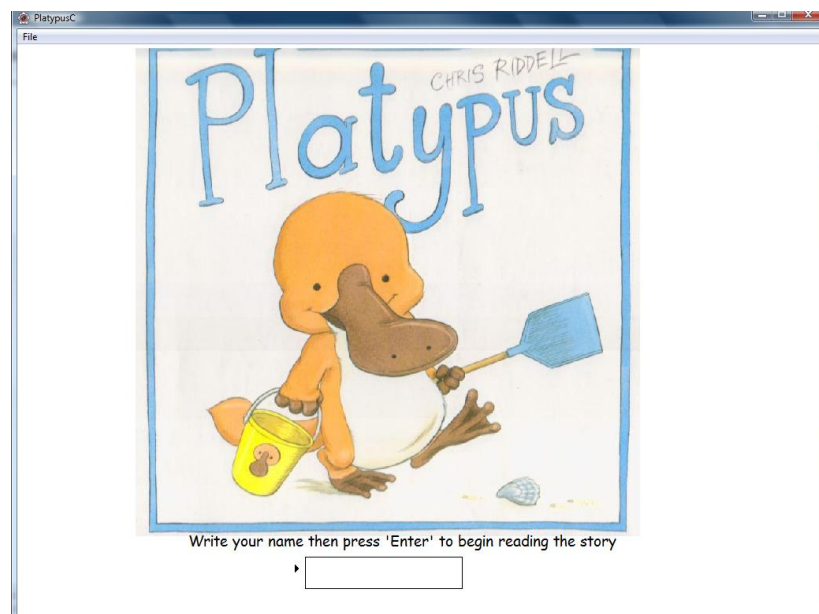
obtained from the authors or publishers to use three stories in this study. The EFL teachers marked the words in the stories that they thought the students would not know. These words were then chosen to be the target words for which annotations were later attached. Only concrete words were chosen, however, in order to reduce the cognitive load on the working memory (see section 2.6.2). In addition, annotations were created for a number of words which the students are most likely to know such as ‘sun’, ‘moon’ ... etc. This was done for purposes of gathering data about the reading behaviour of the participants. It provides information about participants’ use of annotations and whether they consult them for words which they already know.

The first story used was ‘Platypus’ by Chris Riddell. It is a story about a platypus that loves to collect things and goes to the beach to see what he can find. The text contains a total of 195 words, 10 of which were selected as the target words constituting 5% of the text) (See Appendix I). Of the three stories, this one is the shortest and has the fewest annotated words. The second story selected was ‘The Very Hungry Caterpillar’ by Eric Carle. The story is about the life cycle of a hungry caterpillar and the food it eats from coming out of an egg to turning into a butterfly. The story contains a total of 224 words with an average of 16 words per page. 14 target words (nouns) were annotated constituting 6% of the total text. A list of these words is presented in Appendix I. The third and final text in the experiment was ‘The Tiger who Came to Tea’ by Judith Kerr. A number of pages were omitted from the story in order to keep the length similar to the others. This did not affect the sequence of events, however, as two EFL teachers who had not read the story before found it logical and sequential despite the omitted pages. An example of this is at the beginning of the story. A girl called Sophie and her mother are sitting in the kitchen when they hear a ring at the door. The mother then makes a number of guesses as to who it may be before opening the door. The two pages that include these guesses were omitted and the following page where Sophie gets up to answer the door followed. Despite the removal of some pages, this story remained the longest of the three. Of the 329 total words in the story (after omission), 20 target words were annotated, constituting 6% of the entire text. The target words are also listed in Appendix I.

3.5.2 Electronic story design

Electronic versions of the stories were created using a programme called ‘Authorware’ (Authorware, 2007). This software was chosen because it allows for the inclusion of all the features necessary for the experiment. The title page of the story first appears with a box in which the participants type their names and press the ‘enter’ button on the keyboard to gain access to the story (see figure 3.5e). This creates a record that the researcher can then access to gather quantitative data about annotation use (See section 3.7.3).

Figure 3.5e: Sample title page of one of the electronic stories



The consecutive pages of the story were displayed individually, each consisting of an image in the centre with annotated text underneath (see figures 3.5c and 3.5d above). The target words were highlighted and marginal annotations appeared in the form of text, pictures and sound. The option to go back and forth between pages was available through navigation buttons at the bottom of the screen. The final page of the story also included a ‘read again’ button which took the readers back to page 1. Unlike some studies discussed in 2.6.2, navigation options in this design were kept to a minimum in order to avoid a potential occurrence of ‘split-attention’ for participants (see ‘The cognitive load’ section 2.6.2, p.27). The annotated words appeared in blue and the corresponding visual and verbal input appeared in the right margin. In the Treatment A condition, placing the cursor on a highlighted word produced a picture in the right margin and an audio pronunciation of the word (See figure 3.5c). These audio

annotations were recorded by a native speaker of English in the RP accent. Participants used headphones when reading in this condition. In the Treatment B condition, a picture appeared in the right margin with the L1 (Arabic) written translation of the highlighted word (see figure 3.5d). The pictures used in the annotations for both Treatment A and B were obtained from Clipart or via websites granting permission for non-profit usage. In the control condition, the whole text is mono-tone (black) and no annotations were available (See figure 3.5b).

Previous studies have shown that the location of annotations on the screen is influential. Annotations that appear in the margins have yielded better results than pop-up or end of page annotations (AbuSeileek, 2008). They are also faster to access (Jacobs et al., 1994), and are generally favoured by learners (AbuSeileek, 2008 and Myong, 2005). The right margin, as opposed to the left, was chosen because it is more in-line with the direction of eye movement in reading from left to right. The annotations only appear if the cursor is placed over a highlighted word. Once the cursor is moved away from the word, the annotation disappears. The aim of this was to create a temporary presence for the annotations so as to minimize any distraction caused by manual clicking to remove annotations. This was crucial because a review of the literature showed that pop-up annotations (AbuSeileek, 2008) and permanent ones (Chen and Yen, 2013) can lead to split attention and increased cognitive load.

3.6 Controlling the Variables

This section discusses the rationale behind the study design and the steps that were taken to control the variables. According to Alessi and Dwyer (2008), there are three factors that need to be considered when conducting an experimental study:

- 1) Realism – this refers to the authenticity of the task in the experiment. A task ought to resemble one that participants would normally perform outside of the experiment. If a task and its procedures are totally foreign to the participants, the experiment is seen as ‘unrealistic’. In relation to this study, the fact that participants go to the computer lab regularly and use different software for various subjects meant that a level of realism was maintained in the experiment. The use of annotations was novel but the effect of this was, to a degree, controlled by repeated exposure.

- 2) Control – this factor involves making sure that the treatments for all groups are identical except for the key experimental factors. Failure to control the treatments would make it unfeasible to suggest that the experimental factors were responsible for any observed changes in performance. Since the experimental factor in this study is the annotations, it was necessary to control all the other elements consistently across the different reading conditions. This was done by ensuring that the experiment’s setting, instructions, and procedures were identical. With the exception of annotations (experimental factor), the design and layout of the electronic pages were also all identical.

- 3) Power – ensuring that treatments are as effective as possible in order to maximize the learning benefits and produce statistically significant results. By doing so, an experiment can have practical implications for teaching and learning. For these reasons, it was necessary to create the software in a user-friendly manner by simplifying navigation and presenting clear text and images.

It is often the case that aiming to strictly meet the conditions of one factor of the above may come at the cost of sacrificing some aspects of the other two. Hence, Alessi and Dwyer (2008) stress the importance of striking a balance between the three. With respect to this study, one might argue that to maintain the highest level of ‘control’ it would be better to eliminate the main story pictures and present the text and annotations alone. This is because the pictures may give additional clues which can aid vocabulary learning and reading comprehension. Removing the pictures, however, would render the task less ‘realistic’ as students at this age normally read stories with pictures. It would also mean that ‘power’ of the task to improve learning is significantly reduced.

A number of measures were taken in order to control the variables in this study. Although there are three reading conditions (Control, Treatment A and Treatment B) and three stories, the experiment was designed so that every story is presented in the three different conditions. Participants may find the concept, vocabulary, or grammar of a particular story easier to comprehend than that of another story and this in turn would affect their results on the post-tests. Therefore, presenting three conditions for each story ensures that any linguistic gains that are made in a particular condition are attributable to the treatment itself rather than an intrinsic factor in the story. Additionally, the three reading conditions were rotated across the groups. Rather than allocating one reading

condition per group, each group was exposed to the three reading conditions. Again, this was to ensure that any improvements in post-tests results are contributed to the reading condition itself and not to the language abilities of a particular group. Random assignment of groups also aided in controlling the mixed ability classes and any possible differences in familiarity with computers. Participation was optional and some students did not take part in the experiment while others withdrew after the first or second session. The data from participants who did not sit for the three reading conditions was excluded from the analysis. More information on the participants and the selection process is available in section 3.4. Not only is optional participation more ethically sound (see section 3.11), it also ensures that those taking part are motivated and thus creates better control of the ‘interest’ variable discussed earlier (See section 2.7.3).

Controlling the variables is not only restricted to a study’s design but is also essential for devising its research instruments and procedures. The following sections present the research instruments and procedures along with the measures taken to control the variables. Additionally, section 3.9 is devoted to elaborating on the concepts of validity and reliability and the steps taken to establish them in this study.

3.7 Research Instruments

3.7.1 *Pre-test*

The presence of a pre-test is necessary in a ‘true’ experimental design as it ensures parity amongst groups (Cohen et al., 2011). The pre-tests for the stories were all paper-based and were created in the format of a Word Recognition Test (WRT) using a multiple-choice task. A picture is presented along with three word options (1 correct answer and 2 distractors). Participants have to circle the correct word for each picture (See Appendix II). To establish construct validity, instructions for the test were printed in the participants’ L1 to ensure task comprehension and a verbal explanation in L1 was also given by the researcher. The pictures in the pre-tests represent the target words to be learned in the stories (annotations). The pre-tests were all printed in colour to strengthen the clarity of the images. The tests were given to participants immediately prior to reading the story.

In cases where ‘incidental’ vocabulary learning is examined, administering a pre-test immediately before reading is problematic because of carry-over effects. However, since

the vocabulary learning condition in this study is not incidental, there were no concerns about participants being pre-exposed to the vocabulary in the story during the pre-test. In fact, this pre-exposure to vocabulary in a story falls in line with recommended classroom practice. It is meant to encourage prediction of content which leads to better interaction with the text (Harmer, 1998). It is also a means of activating learners' prior knowledge in order to facilitate top-down processing and reducing the cognitive load. A level of realism is, therefore, maintained in the experiment in addition to increasing its power by maximizing the learning potential.

The pre-test was scored using a 1-2-point scale. In this scale, '1' represents a correct response whereas '2' represents an incorrect response or a missing response. Two raters were assigned to score the test, one being the researcher and the other an experienced EFL teacher.

3.7.2 Post-test

The post-test, also paper-based, followed the same format as the pre-test (WRT) and tested the same target words as in the pre-test (See Appendix III). The post-tests were administered immediately after reading to check vocabulary recall and reading comprehension. Since learners in this study are only encountering the target words briefly, the post-tests were designed to test receptive vocabulary knowledge. It tests the learners' ability to identify the written form of a word for a given meaning (picture). It is important to point out that receptive knowledge is not acquired on the basis of a single encounter of a word. A learner may be able to identify a word at a particular point in time but unable to do so at another. The likelihood of retaining meaning increases with multiple encounters of a word in different contexts. It is, therefore, more accurate to refer to the post-tests in this study as a means to test 'immediate recall' of vocabulary in the stories.

Although the post-test vocabulary items were identical to the pre-test, the pictures used in the post-test were different from those in the pre-test and those that appeared in the annotations throughout the stories. By changing the images, the researcher can ensure that the concept of the target word is established as opposed to a single word to picture association. The pre-tests and post-tests were presented to two EFL teachers to check face validity.

In addition to vocabulary, the post-test also included a reading comprehension section comprised of 3 questions. Each comprehension question had 3 options (1 correct answer and 2 distractors). The aim of the test is to measure ‘gist’ reading comprehension as opposed to detailed or information-specific comprehension. The questions, therefore, addressed the main themes in the story and did not include any of the target vocabulary items so as not to jeopardise the construct validity of the test. The entire reading comprehension section was printed in the participants’ L1 and verbal instructions in Arabic were also given in order to avoid any issues that may arise from misunderstanding questions had they been presented in English only. This, in turn, further strengthens the construct validity of the test by making sure it measures comprehension of the story rather than the questions themselves. The reading comprehension tests can be found at the end of each vocabulary post-test in Appendix III. The scoring procedure for the post-test was identical to that of the pre-test.

3.7.3 Log files

The literature review in the previous chapter demonstrated the important role of computer-tracking devices. The Authorware software used in this experiment enabled the use of a tracking tool which kept records of the annotations accessed. Each record contained the name of the participant entered on the title page (figure 3.4e) along with a time stamp for each annotation accessed. The record also showed if a participant clicked on the same annotated words more than once. A sample of how an Authorware record is stored is available in Appendix IV. These records serve as a research instrument to gather data on annotation use which can later be correlated with pre-tests and post-tests results. They also provide information on whether a particular word or treatment yields more clicking and supplement the qualitative data obtained from the focus groups.

3.7.4 Focus groups

A focus group is when a small number of participants (usually 6 to 8) are brought together for the purpose of discussing a specific topic. The discussion is most often guided by a researcher who may invite an observer or moderator to manage the session.

In this study, the main objective of using focus groups as a research instrument was to gather qualitative data that would reflect the nuances of participants' attitudes and beliefs about reading with annotations. The choice to use focus groups was based on the importance of creating a relaxed environment that is more suitable for the young participants. They may feel intimidated in individual interviews or feel obliged to answer in a certain way in group interviews. The interaction that is typically generated from focus groups was deemed more favourable for exploring the participants' experiences and opinions about electronic reading. The discussion was managed by the researcher whom participants had become acquainted with over the course of the experiment. It was thought that introducing an observer who is unfamiliar to the participants may affect the discussion. The decision made against inviting one of the teachers from the school to the focus groups was due to concern that it might be a potential source of discomfort to participants who may fear that their responses are being assessed. The focus group in this study is classified as a 'pre-existing' focus group (Barbour and Kitzinger, 1999) because it is made up of participants who are already acquainted.

The qualitative data from the focus groups was used to shed light on two aspects of the research. Firstly, as a means to establish how the participants felt about reading with annotations. It is reasonable to say that the success of an approach is largely dependent on how the recipients feel about it and whether they are motivated to interact with the design. The participants' input was needed not only to evaluate the execution of hypermedia annotations in this study, but to also inform future designs of electronic texts. It was also thought that the discussion in the focus groups may raise other issues with electronic reading that may not relate to hypermedia annotations per se but are nevertheless pertinent. Secondly, the focus group aimed to provide insight into the participants' reading behaviour and their use of the technology. The following questions were posed as a starting point for examining reading behaviour:

- a) Do participants attempt to infer meaning of words before they consult the annotations?
- b) Do participants use the annotations to confirm any guesses they might have about meanings of highlighted words?
- c) Do participants click on annotations for words they already know? If so, why?

With questions ‘a’ and ‘b’, the qualitative data from the focus group has a prominent role in examining aspects of behaviour that cannot be attained through statistical measures. These aspects are vital because they are undoubtedly the key distinguishing factors between reading with annotations and reading without. It is worth examining whether adding annotation alters the mental reading processes and strategies either positively or negatively.

With question ‘c’, the data from the focus group plays a different type of role. There are questions that quantitative data alone cannot fully answer and, therefore, supplementation with qualitative data is needed. The Authorware log files (quantitative data) will show which annotations were accessed and these can be correlated with the results from the pre-test to establish whether participants seek clarification for words they already know. Nevertheless, qualitative data may provide answers as to why participants are doing this or if they are even aware of this behaviour. This is once again crucial for examining whether annotations are aiding or distracting the young learners. Further examples of how data from the focus group function as a supplement to quantitative data are when particularly significant or unexpected findings emerge from the statistical analyses. The following scenarios were posed as further justification for gathering qualitative data from focus groups:

- d) If certain annotations were accessed significantly more or less than others
- e) If certain annotations were accessed more by a particular treatment group

Hence, the rationale behind using focus groups is to provide qualitative data that is both self-standing and supplementary to quantitative data. Details regarding the preparation and execution of the focus groups are discussed in the procedures section (3.10) and the measures taken to strengthen the validity and reliability of the focus groups are discussed in section 3.9.

3.8 Piloting the Study

It was necessary to carry out a pilot study for a number of reasons, the most important being to validate the research instruments and check procedures. It is through pilot studies that faults and weaknesses in instruments and procedures are revealed and can thus be modified and finalised for the research. A pilot study also provides the researcher with the timescale needed to complete tests or procedures.

3.8.1 Pre-piloting the tests

Pre-piloting procedures were carried out for the preparation of pre-tests and post-tests. The pictures in these tests were isolated and presented to a non-participating group of the same age and mother tongue of the target participants. The group was asked to name the items they saw in Arabic. This was done to ensure that the pictures themselves were clear enough to stand alone. The group successfully identified the majority of pictures but some proved problematic. For example, in ‘The Very Hungry Caterpillar’ test, they struggled with the ‘cherry pie’ picture but because of the concept rather than the clarity of image. This was discovered after a number of pictures of pies were presented with unsuccessful identification. Even when the Arabic translation was offered, some members struggled to grasp the concept. Although pies (sweet and savoury) can be found in Kuwait, they are not at all as common as they are in the west. Another problematic word was ‘plums’ and this was also due to unfamiliarity, albeit to a much lesser degree than ‘pie’. When given a clue that the picture shows a type of fruit, most members were able to identify it. The picture of ‘salami’ also proved difficult to identify but owing to clarity rather than concept as participants responded positively to the Arabic translation. Based on these findings, it was concluded that a verbal Arabic translation and/or description of certain pictures is needed when participants take the pre-test. The decision against omitting the potentially problematic items from the test would damage the ‘realism’ of the task since learners inevitably encounter new words and concepts when reading. Learners will also have the opportunity to construct the meaning of items from the context of the story. This will create a more natural reading environment and allow the researcher to examine the participants’ reading strategies and behaviours.

3.8.2 Piloting the tests

The pre-tests and post-tests as a whole were piloted with three non-participating groups of the same age, nationality and school year. This was done two months prior to carrying out the research (September 2012). The three groups were randomly assigned to the three stories (Platypus: $N = 10$, The Very Hungry Caterpillar: $N = 22$, The Tiger Who Came to Tea: $N = 11$) and were given the relevant pre-tests and post-tests. A couple of discrepancies were found between items in the tests and the annotations that appeared in the stories. The word ‘strawberry’ was presented in both tests but the

annotated item in the story was in the plural form ‘strawberries’. The same applied to the word ‘pan’ in tests which appeared as ‘saucepans’ in the story. The pilot study brought to light these errors and they were subsequently amended. The piloting stage provided the basis for the instructions and time frame allocated to the test-taking parts of the experiment. Finally, the data from the piloted tests were coded into SPSS to measure the internal consistency. The results showed that the tests were internally consistent (see section 3.9.1) and there was no need to omit any items. Thus, the tests were deemed suitable for data collection.

3.8.3 Piloting the electronic stories

Using the same non-participating groups, the electronic stories were piloted to check the clarity of instructions, allocation of time, storage of Authorware records, and to address any potential issues which may arise. One such issue emerged from the first page of ‘Platypus’ which does not contain any annotated words. This caused confusion among students who had sat for a demonstration on reading with annotations as they thought there was a technical problem that prevented the annotations from appearing. Instructions were, therefore, modified for treatment groups to specify that not every page in the stories included annotated words. It was first thought that the stories may be too simplistic for EFL learners aged 9-10. The results, however, revealed that the learners, in fact, knew fewer words than their teachers had predicted. It was also a concern that the stories may be too ‘childish’ in terms of concept. An oral feedback session was conducted after the reading and post-tests to gauge the learners’ reaction. None of them expressed discontent with concept but some felt the stories were difficult because of the new words.

3.9 Validity, Reliability and Trustworthiness of the Study

It is essential for researchers to take measures to ensure the validity and reliability of a study as failing to do so may render any findings inauthentic and, therefore, unusable (Cohen et al., 2011). The quantitative part of this study is assessed in terms of the criteria for validity and reliability whereas the qualitative part is assessed in terms of trustworthiness. The following sections present the steps taken to meet these evaluative criteria in this research. The concept of triangulation is also discussed.

3.9.1 Validity

Validity is so central to effective research that without it, findings are rendered worthless (Cohen et al., 2011). The validity of a research instrument relates to its accuracy in measuring what it was designed to measure (Weir, 2005). Ensuring validity gives integrity to the outcomes of a piece of research (Bryman, 2012). Cohen et al., (2011) explain that there are several kinds of validity, but this research will focus on *face validity*, *content validity*, and *construct validity*.

Face Validity

Face validity is the most basic type of validity and is concerned with whether, on the face of it, there is a correspondence between the research measure and the content it was designed to measure (Bryman, 2012). Face validity can be achieved by asking people, particularly those with expertise in the research area, to simply look at the research instrument and decide whether it is suitable for measuring the concept under investigation (ibid). Face validity in this study was established by presenting the stories and tests to experienced EFL teachers. They confirmed that they appeared to be valid for measuring immediate vocabulary recall and reading comprehension.

Content Validity

To attain content validity, a research instrument must demonstrate that it fairly and comprehensively covers the topic or item(s) it was designed to cover (Mackey and Gass, 2005). It is impractical for researchers to examine all facets of a concept and thus it is imperative that the items selected for measurement represent the concept as closely as possible (Cohen et al., 2011). Content validity is particularly important when comparing two or more teaching strategies and can be established by ensuring that all participants are exposed to identical testing conditions (Mertens, 1998). Since this study examines vocabulary and reading under different conditions, two steps were taken to strengthen the content validity. Firstly, rather than using one story, vocabulary items from three different stories were tested. By doing so, the selected test items can better represent the type of concrete words that appear in authentic children's stories. Secondly, the study was designed so that all the participants were subjected to the same conditions with the exception of the experimental factor (treatment). Hence, equivalence was maintained throughout the experiment's setting, procedures, instructions, and the pre-tests and post-tests.

Construct Validity

Also known as *measurement* validity, construct validity entails looking at whether a measurement tool is indeed measuring the aspect it was intended to measure (Mackey & Gass, 2005). A construct is an abstract and the researcher needs to therefore check that his/her understanding of a particular construct matches the general understanding of that construct. The tests used in this experiment were designed specifically to answer the research questions. Since the study is concerned with immediate, receptive vocabulary recall, the tests were administered immediately after participants read the texts and involved a receptive word recognition exercise rather than a productive one. Additionally, both the experiment and test instructions were administered in the participants' L1 in order to avoid misunderstanding and strengthen construct validity. Similarly, the focus groups were conducted in the participants' L1 (Arabic).

3.9.2 Reliability

Reliability is of particular importance to researchers using quantitative measures (Bryman, 2012). Simply defined, reliability is concerned with consistency and repeatability of measures (Trochim, 2006). According to Bryman (2012), there are three types of reliability: *stability*, *internal reliability*, and *inter-observer consistency*. A measurement is considered 'stable' if it is able to yield similar results if re-administered at another time. Internal reliability, however, relates to the consistency of indicators, i.e. participants' responses to the indicators are all related. As for inter-observer consistency, this involves consistency between the scores awarded by two or more raters for a particular instrument.

Instrument consistency was maintained in this research through equivalence of pre-tests and post-tests. Equivalence of both tests is crucial because if one test is more difficult than the other, the effect of the treatment cannot be measured accurately (Mackey & Gass, 2005). Thus, with the exception of re-shuffling the post-test items to avoid any carry over effects, the tests in this study were identical. Furthermore, *internal consistency* was checked during the piloting stage whereby data from the tests were entered into SPSS to obtain the Kuder-Richardson (KR-20 measure of internal consistency for dichotomous items. This statistical test was selected because it is widely used and recognised as a measure of internal reliability (Bryman, 2012). Results showed

that the tests were internally reliable (see table 3.9) with an alpha coefficient ranging from .7 to .8 across the three stories. While the cut-off point for reliability is debatable, and in some cases misleading (Lance et al. (2006), Nunnally and Bernstein (1994) suggest that .70 is an acceptable minimum for a scale that is newly developed.

Table 3.9: Internal consistency of tests (KR-20)

Reliability Statistics			Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.738	.740	10	.850	.851	22

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.815	.812	19

Inter-rater reliability was not a point of concern in this study since all the tests used were objective (multiple-choice) and thus an answer could only be marked as correct or incorrect. Nevertheless, an experienced EFL teacher reviewed the marked test papers and assisted with the data input into SPSS to help minimise human error.

3.9.3 Trustworthiness

Qualitative data is subject to multiple interpretations since there is no one or single correct way of analysis (Cohen et al. 2011). The validity and reliability of qualitative data is, therefore, examined in light of trustworthiness. This concept covers the terms *credibility* and *transferability* which respectively mirror the internal and external validity in quantitative data (Bryman, 2012 and Mackey & Gass, 2005). In order for data to be credible, they have to be collected over a sufficient period of time so as to ensure a true representation of the area of examination. The data in this study were collected over a period of 7 weeks. Rather than a one-off reading experience, participants were exposed to three electronic texts. This was done to enable the researcher to gather sufficient information about improvements in participants' performance and changes in reading strategies. It is argued that special features may have an initial attraction but will

promptly wear off (Chu, 1995). This ‘novelty effect’ that comes from an initial exposure to an experiment can render the findings implausible. Repeated exposure minimises this effect and gives credibility to the findings. Transferability of qualitative data is achieved through the provision of a thorough, detailed account of the data and sufficient interpretations of their findings (Bryman, 2012). With respect to this study, these are presented in Chapter IV (results) and Chapter V (discussion). Using ‘data triangulation’ also promotes trustworthiness and this technique is discussed in section 3.9.3.

In terms of the reliability of qualitative data, this is manifested in the concept of *confirmability*. All the data collected upon which the researcher’s arguments and claims are based must be included in the research in order to achieve confirmability. The data from this experiment are presented in the following chapter. To strengthen the reliability of data from the focus groups, the discussions were audio recorded. This enabled the researcher to refer to the recordings at any time for data coding and checking. Transcriptions of the focus groups were also made for ease of reference and are included in Appendix VI. The validity of qualitative research can be jeopardised by researcher bias as researchers may influence the findings of a study either intentionally or unintentionally. A researcher may omit data that does not match his/her expectations or theories. A researcher may also influence the way in which data is collected by, for example, asking participants leading questions in interviews or focus groups in order to obtain specific information. To overcome researcher bias in this study, all the collected data are included either in the results chapter or the appendices.

3.9.4 Triangulation

Triangulation of data is crucial for reducing bias and increasing the validity and reliability of the research findings. It involves the use of two or more methods (multi-method approach) for data collection (Cohen et al., 2011) to provide a more adequate conclusion of a study (Mackey and Gass, 2005). Triangulation in this study is achieved through the use of both quantitative and qualitative data. The quantitative data comprised pre-tests and post-tests in addition to log files that were created when participants read the texts. The qualitative data comes from the focus groups and works hand in hand with the quantitative data in order to provide a well-rounded answer to the research questions.

3.10 Experiment Procedures

The experiment was carried out over a period of 7 weeks. The first 3 weeks were spent at the girls' school. The three classes were taken to the computer lab separately, once a week, for three weeks. They were taken to the lab mostly during 'free lessons' (when a teacher of a class is absent) and occasionally in place of 'leisure lessons' (lessons that are not graded such as music, art ...etc).

The following three weeks were spent at the boys' school. Since the Computer Science subject is not a graded one, the participating school saw fit to conduct the experiment during this lesson. Hence, as with the girls, the boys were taken to the lab once a week over a period of three weeks. During the seventh week, the focus groups were conducted at both schools.

Each participating class was taken to the computer room separately. The nature of the experiment was explained to the participants and those who did not want to participate sat out (see section 3.11 for a detailed account of ethical issues for participation). Participants were then given the pre-test and asked to complete it within 8 minutes. They were encouraged to seek clarification from the researcher for any pictures that were not clear. A translation or description in this case was given for the pictures. Participants were reassured that the tests were not-graded and that their teachers and parents would not see them. They were also encouraged to leave the questions they did not know blank. This was done to limit any guessing which might skew the results.

In order to reduce the extraneous load of the reading task (see p.25), a whole-class demonstration on how to use the software then followed. An interactive white board displayed a sample page and instructions were given in the learners' L1. Three sample pages were created for purposes of demonstration. These pages were not from the electronic stories in order to avoid any pre-exposure to the story and annotations in the experiment. Each group was shown the sample page that corresponded to the reading condition they were exposed to on that day. Thus, there was a sample page with no annotations for the control condition and two sample pages for the two treatment conditions. In addition to reducing the extraneous load, the demonstration also aimed to provoke interest in the task in order to maximise the learning potential (see section 2.7.3). Following the demonstration, participants were told that they were going to read a story about a platypus/caterpillar/tiger. They were also informed that on completion

they were going to be given a test on the words that appeared in the story and some questions on the events that took place. While some researchers may argue that the purpose of reading can influence participants' reading strategies and look-up behaviour (Linderholm and van den Broek, 2002), maintaining task authenticity was of importance in this experiment. If there are implications to be made from this experiment, there needs to be a close correspondence between the task type and procedures in the experiment and that of real classroom tasks. It was vital, therefore, to encourage the participants to look at both words and events in the story. It is also worth noting that alerting students to the purpose of a reading task is more problematic when it comes to examining 'incidental' learning. Since this experiment deals with intentional learning, it can be argued that raising awareness prior to reading is a form of schema activation and is common practice that precedes EFL reading. Furthermore, some studies have found that learners tend to process meaning only without attending to form, unless explicitly instructed to do so (VanPattern, 2004). The task instructions or demands may, therefore, encourage learners to focus on form.

Every participant had individual access to a computer. They entered their names into the front page of the story and were asked to read at their own pace. They had the option to read the story as many times as they liked in order to keep the reading experience as realistic as possible (see section 3.6 on 'realism'). Most participants finished in under 20 minutes. The reading time allocated was the same for all groups regardless of the reading condition. Students who finished reading had to close the software and raise their hands to receive the vocabulary and reading comprehension post-test. They were not allowed to re-open the story while answering the test. The average time it took for participants to complete the post-test was approximately 10 minutes. At the end of each reading session, the Authorware records detailing the access to annotations were collected from the computers.

3.10.1 *Focus groups procedures*

It was important to conduct the focus group in a location that is familiar to the participants. This is to increase their sense of security, encourage interaction and avoid any distractions from new surroundings. The focus group took place in the computer lab, while vacant, during the final week of the experiment.

Purposive sampling was used in order to make the groups as closely representative of the population as possible. The groups included participants who scored highly on both tests, those with a low score on both tests and those who scored low on the pre-test but high on the post-test. This heterogeneous design of the focus group was also developed in order to encourage diversity in the discussion. With respect to the group composition, the girls' focus group consisted of 4 participants and the boys' of 6. The literature indicates that with young participants it is recommended to construct smaller groups of members (4-6 members) (Kennedy, et al., 2001) in order to encourage equal participation. The group size also makes it more feasible for the researcher to control the interaction.

The participants sat in a circular seating arrangement with the moderator embedded. They were reminded of how the focus group will be conducted (participants were debriefed during recruitment; see section 3.11 on ethical considerations) and the recording equipment was tested. The focus groups were held over a period of 35 minutes with a halfway break of 8 minutes. Morgan et al. (2002) found that the quality of their young participants' contributions deteriorated progressively after 45 minutes. They found that holding two twenty-minute sessions separated by a break was more beneficial for the quality of interaction. The break in this study was also used to ensure that participants were comfortable and happy to continue. They also used this time to have a drink and eat a snack if desired.

The focus groups were all conducted in the participants' mother tongue (Arabic). The beginning of the focus group was dedicated to reflecting on the stories and types of reading the participants had done over the last three weeks. It was important to refresh their memories since the focus group was not conducted immediately after reading the stories. The moderator then posed a number of questions and encouraged participants to discuss and build on each other's contributions. The questions revolved around their experiences and preferences for reading in different conditions, how they approached the texts in different conditions, and suggestions they had for improving electronic reading.

3.11 Ethical Considerations

Ethical concerns can stem from the topic of investigation, the methods used to gain reliable and valid data, the type of data collected, the participants involved, and the

experiment procedures (Cohen et al., 2011). Ethical issues may, therefore, surface at any stage of an experiment. Participants' rights and privacy can be protected through obtaining informed consent and exercising confidentiality in the handling of data. The researcher must ensure that information given by participants will not be disclosed to other parties in any way that would reveal the participants' identity.

Due to the vulnerable age of the participants in this study, ensuring that ethical standards are met was of particular importance. Ethical approval was first obtained from Newcastle University prior to conducting the experiment. Once this was obtained, it was taken to 'The Research Institute' in Kuwait. The experiment procedures were explained to the institute and the research materials were presented for approval. The institute then issued a letter of approval which was sent to the Ministry of Education in Kuwait who gave the researcher the final approval to access public schools. The next step was to gain approval from the gatekeepers. The principals and heads of departments at various schools were debriefed about the experiment. Some schools declined because they were either too overcrowded or had problems with their computer labs. Of the schools that accepted, two took part in the experiment.

Informed written consent was obtained from parents of the participants prior to participation. They were made aware of the purpose of the experiment and that participation is optional and does not affect their child's school marks in any way. It was also made clear that participation is anonymous and that the participants have the right to withdraw from the experiment at any point in time without repercussion. This information was similarly reiterated to the young learners in person. They were debriefed about the overall procedures of the experiment and were reassured that they could opt out at any point.

It was also necessary to reduce any anxiety that may arise from taking the tests in the experiment. Participants were reassured that their tests results are not part of their school grades and will not be viewed by their parents or anyone in the school. Hence, their right to confidentiality was maintained. Some students decided not to take part in the experiment at all and the researcher respected their decision by not pressing for a justification. These students were present in the lab but occupied themselves with activities of their choice such as drawing, doing homework, or using the computer for other activities. A few of these students later changed their minds and decided to take part in week 2 or 3. In contrast, other learners took part in the first week but decided not

to continue after that. The participants who did not sit for a total of 3 stories were eliminated from the data analysis. The same consent procedures were applied to the execution of the focus groups. Although some parents declined, the participants from those who had agreed did not later withdraw.

An ethical dilemma arose with regards to the Authorware records that stored participants' access of annotations (see section 3.7.3). Although participants need to be aware of the data collected from them, informing them of these records would likely alter their reading behaviour and in turn jeopardise the validity of the experiment. In such situations, the researcher needs to examine the costs/benefits ratio (Frankfort-Nachmias and Nachmias, 1992). Simply put, this involves weighing the cost to participant rights against the possible impact of the research. Since the records in this study do not contain sensitive information and are treated with confidentiality, a decision against informing the participants was made. The records were collected immediately after the class left the lab and were then erased.

It was also ethically necessary to meet the conditions of the publishers and authors who had granted permission to use their stories in the experiment. In order to do so, the stories were only accessible locally, i.e. not via the Internet. The material was only accessed by the participants and researcher. Once the experiment was completed, the stories were erased from the computers in the school.

Chapter 4. Results

This chapter analyses the quantitative and qualitative data gathered to answer the research questions on the effects of annotations on immediate vocabulary recall, reading comprehension and reading behaviour. The analysis includes quantitative data from pre-tests, post-tests and log files, in addition to qualitative data from focus groups. The data from participants who did not sit for the three reading sessions were omitted from the analyses.

The statistical software SPSS was used to examine differences in performance between pre-tests and post-tests across the three reading conditions and whether these differences are significant. The statistical analysis also included the log files to examine participants' clicking behaviour in terms of whether a particular condition yielded more clicking and whether participants clicked on familiar words (these are words that received high scores in the pre-tests). Additionally, the log files were used to test for correlations between results on the tests and frequency of accessing annotations. With respect to the focus groups, a thematic approach is used to tabulate the qualitative data.

This chapter is divided into three sections that correspond to the three main research questions. Each section begins with a brief description of the data followed by the statistical tests that were employed for the analyses and the rationale behind them. A hypothesis for each research question, and sub-question, is given for ease of analysis and reporting of results. The chapter concludes with a summary of the findings in order to prepare the reader for the discussion chapter that follows.

4.1 Vocabulary Outcomes

This section analyses quantitative data pertaining to research question 1 and its sub-question:

R-Q 1: Do hypermedia annotations support vocabulary learning for young Kuwaiti learners?

Sub-Q 1: Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for L2 vocabulary recall?

Given that this study uses pre-tests and post-tests from three stories in three reading conditions, it was necessary to clarify the steps taken to prepare the data to answer each research question. The subsections below describe the processes of inputting data, the hypotheses and the statistical tests used.

4.1.1 Preparing the data for R-Q 1

R-Q 1 is concerned with examining whether there is a difference in immediate vocabulary recall (dependent variable) between two reading conditions (independent variables):

Reading with annotations Vs reading without annotations

To form the ‘reading with annotations’ condition, it was necessary to combine data from the two treatment conditions (Treatment A: Picture + L2 Pronunciation, Treatment B: Picture + L1 Translation) and compare these to the control (no annotations):

Table 4.1a Preparing data to answer R-Q1

		Dependent variables	
		Vocabulary Recall	
		Pre-test results	Post-test results
Independent variables	Reading with annotations	Treatment A + Treatment B	Treatment A + Treatment B
	Reading without annotations	Control	Control

The reading conditions all included pre-test and post-test results from 3 stories (see table 3.4a). It was crucial to provide an equal basis for comparing these results since the stories differed in the number of target words they included (See Appendix I). Story 1 tested 10 items, story 2 tested 22 items and the third story 19 items. The stories were read in different conditions and in different orders (see counterbalancing method, section 3.5). The following table 4.1b is a sample of the calculation method followed for combining the treatments to form a pre-test

score for participants. The same procedures were applied when combining participants' post-test results in the two treatments.

Table 4.1b combining treatments to obtain a 'reading with annotations' score

Participant	Treatment A pre-test result	Treatment B Pre-test result	Total score	Score divided by maximum possible total	Final pre-test score (percentage for reading with annotations)
1	6 (out of 10)	15 (out of 22)	21	$21/32^* = .65$	65%
2	9 (out of 22)	4 (out of 19)	13	$13/41 = .31$	31%

*32 is the maximum possible score participant 1 could have achieved from reading in treatment A (10 points) and treatment B (22 points), $10+22 = 32$. Note the differences in maximum score achievable which was dependent on the stories the participants read.

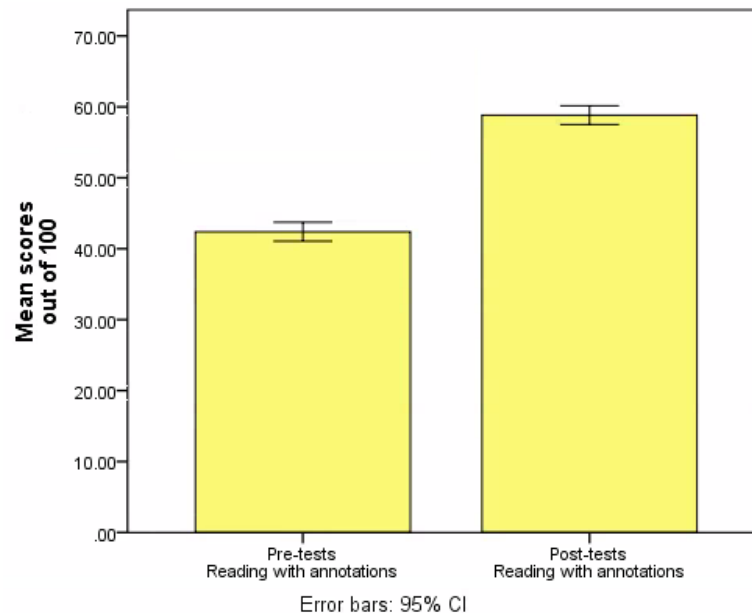
4.1.2 Testing Hypothesis 1

Hypothesis 1 Consulting annotations has no effect on immediate vocabulary recall

To test the hypothesis that means of the pre-reading tests ($M = 42.39$, $SD = 18.61$) and post-reading tests ($M = 58.83$, $SD = 20.11$) for reading with annotations were equal, a Paired Samples T-Test was performed using the statistical software SPSS. Prior to this, it was necessary to examine the assumption of normality in the distribution of the difference scores. This is to ensure that there are no scores that are abnormally higher or lower than the average as this would skew the data. The assumption was considered satisfied since the skew and kurtosis levels results were $-.033$ and $-.427$ respectively. These figures fall in line with Posten's (1984) maximum allowable values for a t-test (i.e., skew $< |2.0|$ and kurtosis $< |9.0|$). The use of the Paired Samples T-Test to examine this hypothesis was deemed suitable as the correlation between the two conditions was estimated at $r = .73$, $p < .001$. The correlation in this case refers to the participants' performances on both tests. If a participant performed lower than average on the pre-test, his/her performance on the post-test, even if improved, will remain lower than the average of improvements by other participants. Based on the findings of the T-test, the null hypothesis of equal vocabulary scores means for both tests was rejected, $t(111) = -12.20$, $p < .001$ because the post-tests mean was statistically significantly higher than the pre-tests mean. It was also necessary to examine the size of the difference between the

two tests (effect size). Cohen's d was estimated at 0.84 which is considered large according to Cohen's (1998) and Cohen's (1992) guidelines for effect size in the social sciences (small 0.10, medium 0.30, large 0.50). Hence, with regards to reading with annotations, the data points to a significant difference between the pre-tests and post-tests with the latter yielding higher scores. Figure 4.1a displays a graphical representation of the means and adjusted 95% confidence intervals based on the Loftus and Masson (1994) procedure for within-subjects design. This procedure was followed because it takes into account the correlations between the pre-test and post-test for each participant and thus reduces the standard error of the difference between means.

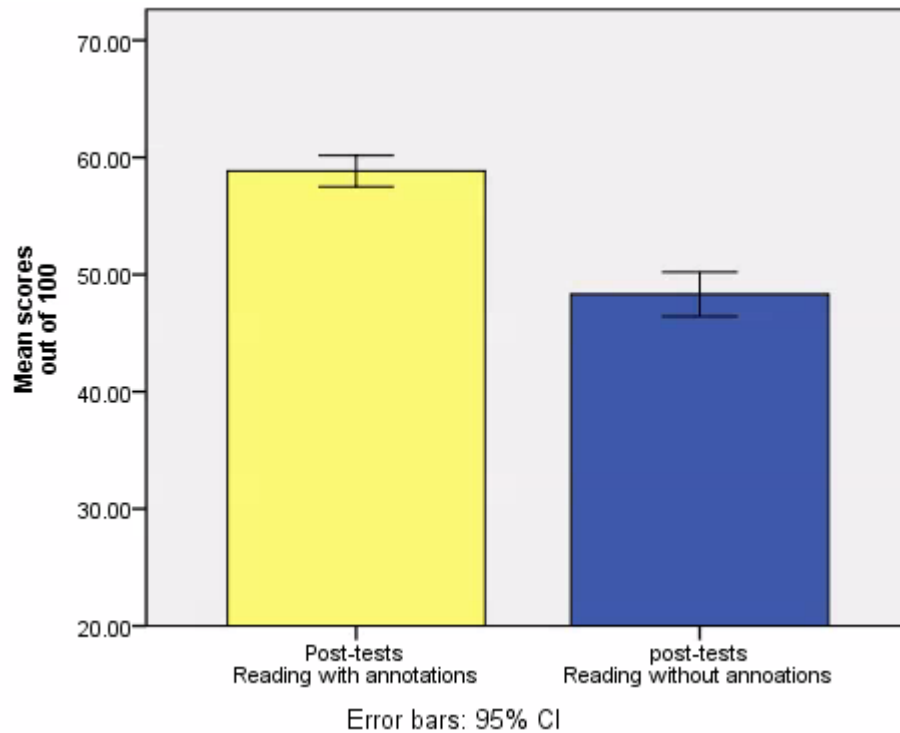
Figure 4.1a: Vocabulary pre-tests and post-tests means for reading with annotations



After establishing that providing annotations improved participants' results on the vocabulary post-tests, it was necessary to compare this to their performance when no annotations were present. A Paired Samples T-Test was first used to establish whether participants' pre-test scores were, more or less, equal in both conditions. This step is important because it gives the researcher confirmation that the intervention (i.e. annotations) was responsible for any improvement in performance. The assumption of normality was confirmed with a skew level of 0.18 and a Kurtosis level of 0.20, and the T-Test result then showed that there was no significant difference between pre-test results in both conditions $t(111) = .334, p > .73$. The next step was to perform another Paired Samples T-Test to compare the post test results from reading with annotations ($M = 58.83, SD = 20.11$) to those from the control condition ($M = 48.33, SD = 22.84$) where no annotations were available. Again, the assumption of normality was tenable with a

skew and kurtosis level of 0.03 and 0.51 respectively. Results of the T-Test show a significant difference between the post-test results for reading with annotations and reading without annotations $t(111) = -5.71, p < .001$. These findings suggests that when participants read with annotations, their results on the post-tests were significantly higher than when they read without annotations (effect size = 1.02). A graphical representation of the means and adjusted 95% confidence intervals is presented in figure 4.1b.

Figure 4.1b: Vocabulary post-tests means for reading with and without annotations



After establishing the effect of annotations, the next step in the analysis was to determine whether the two types of annotations in the experiment (Treatment A and Treatment B) differed in their effects on immediate vocabulary recall. This constitutes Sub-Q 1 below.

4.1.3 Preparing the data for Sub-Q 1

Sub-Q 1 looks at the effects of three reading conditions on immediate vocabulary recall.

Treatment A - Treatment B – Control

Unlike R-Q1 where data from the two treatments were combined, Sub-Q1 is concerned with examining the differences between the two treatments and they were, thus, kept separate. The vocabulary pre-tests and post-tests data from three stories were converted so that a maximum score of 100 (i.e. a percentage) could be awarded on all pre-tests and post-tests as follows:

Story 1: 10 (items) X 10 = 100

Story 2: 22 (items) X 4.5455 = 100

Story 3: 19 (items) X 5.2362 = 100

If a participant, for example, scored a result of 15/22 on the test for story 2, this result would be multiplied by 4.5455 and entered into the statistical programme as 68.18/100.

4.1.4 Testing Hypothesis 2

Hypothesis 2 There is no effect of annotation type on immediate vocabulary recall

Table 4.1c displays the descriptive statistics for the pre-tests and post-tests results across the three reading conditions (Treatment A = Audio+ Picture annotations, Treatment B = Translation + Picture annotations, and Control = no annotations).

Table 4.1c: Descriptive statistics for the pre-tests and post-tests across the reading condition.

Reading Condition*		Mean	N	Std. Deviation	Std. Error Mean
Treatment A	Pretests	42.3744	112	23.13262	2.18583
	Posttests	62.0151	112	25.23758	2.38473
Treatment B	Pretests	44.8475	112	20.81609	1.96694
	Posttests	59.4220	112	23.15926	2.18834
Control	Pretests	42.9409	112	21.91525	2.07080
	Posttests	48.3271	112	22.84487	2.15864

(*Treatment A = Picture + L2 pronunciation, Treatment B = Picture + L1 Translation, Control = no annotations)

It is evident that the post-test mean for each reading condition is higher than the pre-test mean, indicating an improvement in the participants' results. This holds true for the control reading condition as well (Pre-test: $M = 42.9$, $SD = 21.9$, Post-test: $M = 48.3$, $SD = 22.8$). However, it was necessary to test whether the apparent improvements in mean scores were statistically significant and thus the Paired Samples T-Test was used for this

purpose. The results in table 4.1.4b below reveal that the difference between the means of pre-tests and post-tests is indeed significant at ($p < .05$) for all reading conditions.

Table 4.1d: Paired samples t Test for pre-tests and post-tests in three reading conditions

		Paired Differences	T	Df	Sig. (2-tailed)
		95% Confidence Interval of the Difference			
		Upper			
Pair 1	PretestsTA – PosttestsTA	-15.82149	-10.190	111	.000
Pair 2	PretestsTB – PosttestsTB	-11.36230	-8.991	111	.000
Pair 3	PretestsC – PosttestsC	-2.16453	-3.313	111	.001

Table 4.1.4b shows a significant improvement in results from the pre-test to post-test for all reading conditions (Pair 1 = Treatment A, Pair 2 = Treatment B, Pair 3 = Control).

These findings suggest that the Treatment A condition and the Treatment B condition both yielded significant improvements in vocabulary scores on the post-tests ($p < .001$). Further tests were needed, however, in order to establish which treatment is superior. A Paired Samples T-Test was thus carried out on the vocabulary post-test results for the two treatments. The findings revealed that there was no significant difference in the post-tests scores for Treatment A ($M = 62.02$, $SD = 25.23$) and Treatment B ($M = 59.42$, $SD = 23.15$) conditions; $t(111) = 1.05$, $p = .292$. Since neither annotation combination proved superior to the other, the null hypothesis that they have an equal effect cannot be rejected. Hence, it cannot be said that the combination of Picture + Audio in the annotations is better for immediate vocabulary recall than Picture + Text annotations or vice versa.

Since the control condition also yielded a statistically significant improvement on the post-test results, a One-way Repeated Measures ANOVA was carried out to see whether the difference between the treatment conditions and the control condition on the post-test results is significant. This type of ANOVA test is used to examine the performance of a group of individuals who all undergo three or more conditions (i.e. the same group of

participants in this research sat for three reading conditions). The Mauchly's Test of Sphericity was first used to test the assumption of homogeneity of variance and covariances. This test assumes that variances of the differences between all possible pairs of groups (i.e., levels of the independent variable) are equal. The result was $p > .05$ demonstrating that the assumption of equality was not violated. The ANOVA test revealed that there was a significant difference between the treatments on the vocabulary post-tests, Wilks' Lambda = 0.72, $F(2,110) = 20.98$, $p = .000$. Follow up comparisons indicated that two pairwise differences were significant ($p < .01$). The post-test scores for the Treatment A condition were significantly higher than those of the Control, and the vocabulary scores for the Treatment B condition were significantly higher than those of the control. Figure 4.1c below displays the means graphically.

Figure 4.1c Vocabulary post-test means for Treatment A, Treatment B and Control

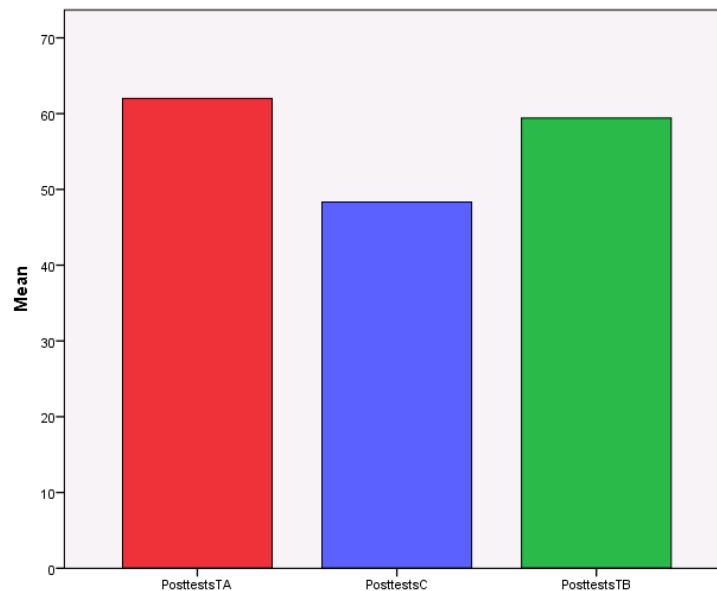


Figure 4.1c illustrates the means for the post-tests in the three conditions. The means of Treatment A (TA) and Treatment B (TB) were significantly higher than that of the control.

The findings in this section have shown that after reading with annotations, participants' performance on the vocabulary post-test improved significantly. While the post-test results for the control condition were a significant improvement on the pre-test results, the improvements observed in the two treatment conditions (Picture + text and picture + audio) were significantly higher than for the control condition. No significant difference was found, however, between the two treatments.

4.2 Reading Comprehension

The analysis in this section is carried out to answer research question 2 and its sub-question:

R-Q 2: Do hypermedia annotations facilitate reading comprehension for young Kuwaiti learners?

Sub-Q2: Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for reading comprehension?

Unlike the vocabulary component of the tests which differed in the number of items tested, the maximum score achievable on all the reading comprehension tests was 3. Therefore, there was no need to convert these scores to 100. The reading comprehension test was only present in the post-tests since reading the stories, in this experiment, was a prerequisite. A similar procedure to the vocabulary analysis was adopted for comparing results of the reading comprehension tests. Results from Treatment A and Treatment B were combined and then divided by two to obtain an average (see table 4.2), thereby rendering the data fit for testing the hypothesis in the following section.

Table 4.2a Data preparation for reading comprehension post-tests

Participant	Treatment A Reading post-test result	Treatment B Reading Post-test result	Total score	Total score divided by 2 (average)	Final score for reading post-test*
1	2 (out of 3)	3 (out of 3)	5	$5/2 = 2.5$	2.5
2	1 (out of 3)	3 (out of 3)	4	$4/2 = 2$	2

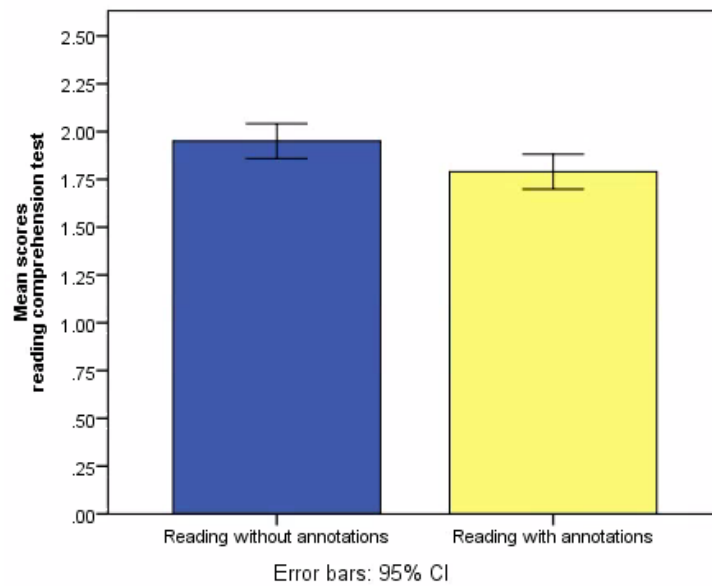
*Maximum final score from combining two treatments and dividing by two is 3. This is then compared to the maximum score from reading post-test in the control condition which is also 3.

4.2.1 Testing Hypothesis 3 Consulting annotations has no effect on reading comprehension

The assumption of normality in the distribution of the difference scores was first tested. The assumption was considered satisfied since the skew level was .078 and the kurtosis .813. These figures fall below Posten's (1984) maximum allowable values for a t-test (i.e., skew < |2.0| and kurtosis < |9.0|). To test whether the mean results on the reading

comprehension test for reading with annotations ($M = 1.79, SD = 0.82$) and without annotations ($M = 1.96, SD = 0.96$) were equal, a Paired Samples T-Test was performed. The use of this test for this hypothesis was appropriate as the correlation between the two conditions was estimated at $r = .41, p < .001$. Figure 4.2 gives a graphical representation of the means and adjusted 95% confidence intervals based on the Loftus and Masson (1994) procedure (see section 4.1.2).

Figure 4.2: Reading comprehension post-test means (with and without annotations)



Based on the T-Test results, the null hypothesis of equal reading comprehension scores means for both conditions was not rejected, $t(111) = 1.744, p = .08$. Hence, the reading with annotations condition did not result in statistically significantly higher results than reading without annotations. Although the reading comprehension scores were higher in the control condition, the difference between the conditions was not statistically significant. Therefore, it is also not possible to conclude that the presence of annotations had an adverse effect on reading comprehension.

4.2.2 Testing Hypothesis 4 There is no effect of annotation type on reading comprehension results

In this analysis, the two treatments, previously combined, were kept separate in order to determine the effect of each one on reading comprehension. Table 4.2b displays the means for the reading comprehension scores in each reading condition along with the standard deviation and total number of participants ($n = 112$).

Table 4.2b Descriptive statistics for the reading comprehension scores in the post-tests

	Mean	Std. Deviation	N
Treatment A	1.79	1.052	112
Treatment B	1.80	1.021	112
Control	1.96	.962	112

The highest mean was that of the control group ($M = 1.96$, $SD = .962$), whereas the lowest mean was that of the Treatment A condition ($M = 1.79$, $SD = 1.05$). This indicates that the control group outperformed the treatment groups on the tests. However, to test whether the differences between the above means is significant, a One-way Within-Subjects (Repeated Measures) ANOVA test was conducted. The homogeneity of variance and covariances assumption was tested using Mauchly's Test of Sphericity which yielded a result of $p > .05$, indicating that the assumption is tenable. The ANOVA test revealed that there was not a significant effect of the treatments on reading comprehension, Wilks' Lambda = 0.97, $F(2,110) = 1.52$, $p = .222$. Since the significance value for reading scores across the three groups is $p > .05$, the null hypothesis that the means of scores are equal cannot be rejected. Thus, it can be concluded that the type of annotation does not appear to affect reading comprehension either positively or negatively.

4.3 Reading Behaviour

The above statistical tests provided results pertaining to this thesis' first and second research questions and their sub-questions. This section reports on both the qualitative and quantitative data that were gathered to answer research question 3:

R-Q 3: How do young Kuwaiti learners use annotated texts and how do they feel about this type of reading?

The two types of data aim to provide insight into participants' reading strategy (approach to different reading conditions) and use of annotations (clicking behaviour). The data also addresses participants' preferences for reading conditions and whether this correlates to their performance. Qualitative data from

the focus groups is first presented followed by the statistical analysis of the quantitative data.

4.3.1 Focus group data

The focus groups were conducted at the two participating schools during the last week of the experiment (December 2012). The first focus group consisted of 4 girls who read the electronic stories in the order of: Treatment A – Treatment B – Control. The second focus group consisted of 6 boys who read the electronic stories in the order of: Control Treatment A – Treatment B. The aim of these focus groups was to gather information relating to two key themes in this study: participants' reading behaviour (strategies and clicking), and their perceptions and preferences for reading in different conditions. The focus groups were also intended as an opportunity to investigate any issues that the researcher did not account for but may, nevertheless, be of relevance to the investigation. A data tabulation method is used below to organise and present the findings (table 4.3.1). These are referenced to the focus groups transcripts that were translated into English and are available in Appendix VI.

Table 4.3 Tabulation of focus group data

Area of Investigation	Aim	Findings	Reference*
How participants deduce meaning of unknown words in regular printed texts (un-annotated)	To examine whether any existing strategies are carried over to electronic reading or are altered with the presence of annotations.	use dictionaries obtain an online translation ask an adult or parent	T1, 40 T2, 73 T1, 31; T2, 69
How participants deduce meaning of unknown words in the experiment (un-annotated texts)	To examine learners' inferencing strategies in electronic texts.	looked at the main picture from the text itself remember clues from the pre-test	T1, 136; T2, 176 T2, 179 T2, 177
Whether participants use annotations exclusively for new words or for known words as well	To gather information about how learners' use annotated texts and their clicking behaviour	New words only New and known words	T1, 81; T2, 130, 134 T2, 136, 146
Learners' preferences and perceptions	To examine whether a particular reading condition is more popular and may thus be more successful for learning	Audio + picture annotations L1 + picture annotations No annotations	T1, 62; T2, 93 T1, 56, 87; T2, 99 T1, 54, 64; T2, 119
Emerging themes	To bring to light any pertinent issues that were unaccounted for	Reading behaviour influenced by pre-tests and post-tests	T1, 149; T1, 159; T1, 234; T1, 147; T2, 185

* T= Transcript (T1= Transcript 1, T2= Transcript 2), T1, 40 = Transcript 1, line 40. All transcripts are in Appendix VI

Participants' responses were varied in terms of their use of strategies and clicking behaviour. No clear consensus was reached as to which reading condition is optimum. The findings from the above table will be discussed further in Chapter IV in conjunction with the quantitative data that is analysed below.

4.3.2 Weekly performance data

The statistical tests that were carried out earlier in this chapter compared the three reading conditions using the collective data from three stories over a period of three weeks. The analysis in this section looks at participants' performances on a weekly basis. This is done to gather information on any trends or changes which may have occurred from one reading session to the next. During the focus groups, some participants reported that they had consciously adapted their reading strategies according to the condition they were exposed to. This section, therefore, looks for any quantitative evidence to confirm that the order of the conditions is an influencing factor in participants' reading behaviour. The statistical findings from each week are first presented. This is followed by some further statistical tests on the log files to establish participants' clicking behaviour in the treatment conditions. The section concludes with a brief summary of the findings.

Week 1

Participants read the story 'Platypus' in the first week. In terms of vocabulary, the maximum achievable score on the pre-tests and post-tests was 10. The pre-test mean scores across the three conditions ranged from $M = 4.97, SD = 2.07$ to $M = 5.15, SD = 2.1$ and the post-test means from $M = 5.02, SD = 2.3$ to $M = 7.5, SD = 2.4$. A Paired Samples T-Test was used to establish whether the improvement in the post-test mean scores is significant. The results show a statistically significant improvement in the post-test for participants who read in the treatment conditions ($p < .000$) but not for those who read in the control condition ($t(37) = -.183, p = 0.85$). Hence, participants who read in the control condition in week 1 did not achieve a significant improvement on the vocabulary post-test.

As for the reading comprehension test, a One-way Between-Subjects Anova test was used to determine any differences in mean scores between the three reading conditions. Results showed that there was no significant effect of reading condition on comprehension scores with $p > .05$ [$F(2,109) = .05, p = 0.949$]. Figure 4.3a gives a visual illustration of participants' performance during week 1.

Figure 4.3a: Participants' performance during week 1

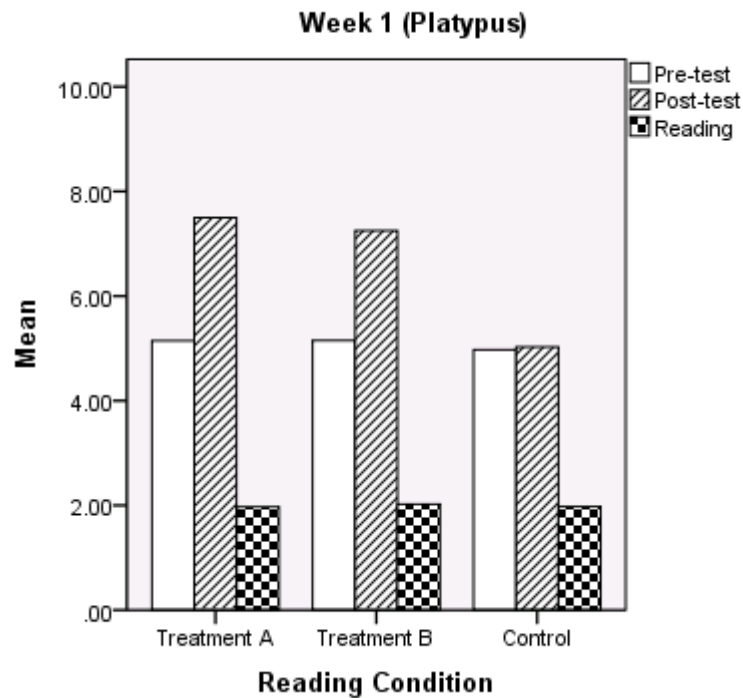


Figure 4.3a presents three bars which correspond to the mean scores for the vocabulary pre-test, vocabulary post-test and reading comprehension test. The X axis shows the three reading conditions and the Y axis shows the mean scores (maximum score of 10 for vocabulary and 3 for reading).

Week 2

During the second week, participants read the story 'The Very Hungry Caterpillar'. Participants in all reading conditions made statistically significant improvements on the vocabulary post-test with the Paired Samples T-Test yielding a p value of $< .05$. Unlike the control condition in Week 1, participants who read in the control condition in week 2 achieved a significant improvement on their vocabulary post-tests ($M = 11.28$, $SD = 5.53$). The means of the reading comprehension test were compared using a One-way Between-Subjects Anova test. No significant effect was found for reading condition on comprehension scores, $p > .05$ [$F(2,109) = .74$, $p = 0.477$]. Figure 4.3b displays these findings graphically.

Figure 4.3b: Participants' performance during week 2

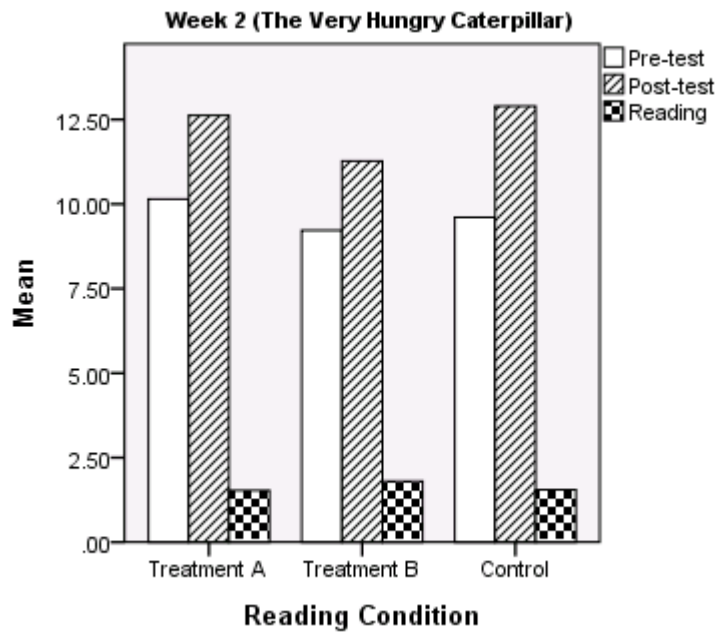


Figure 4.3b presents three bars which correspond to the mean scores for the vocabulary pre-test, vocabulary post-test and reading comprehension test. The X axis shows the three reading conditions and along the Y axis is the mean scores (maximum score of 22 for vocabulary and 3 for reading).

Week 3

In the third and final week, participants read the story 'The Tiger Who Came to Tea'. The Paired Samples T-Test was also used to compare the means of the pre-test and post-test for each condition. As with week 2, a statistically significant difference in all conditions at $p < .05$ was observed. A One-way Between-Subjects Anova test was used to compare the effect of the reading conditions on the reading comprehension scores in the post-test. Results were consistent with the previous weeks as the reading condition did not significantly affect comprehension scores, $p > .05$ [$F(2,109) = .81, p = .448$]. These findings are illustrated in figure 4.3c below.

Figure 4.3c: Participants' performance during week 3

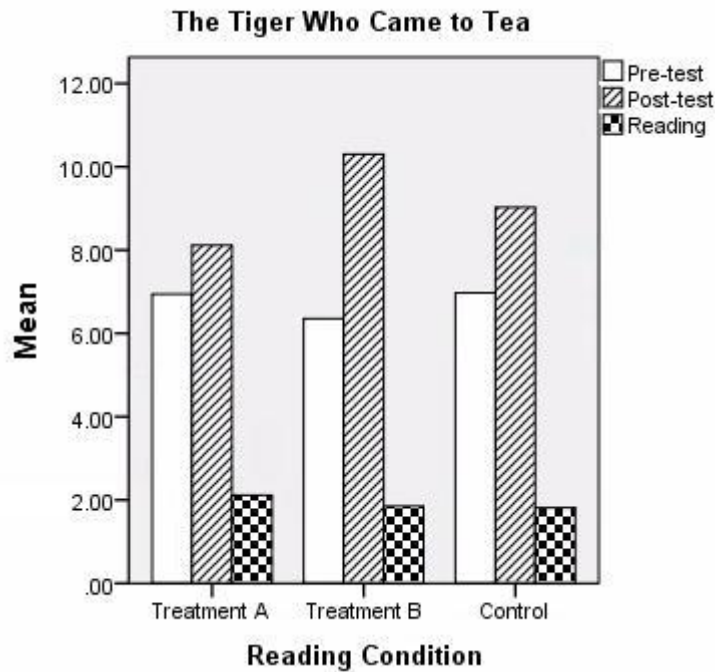


Figure 4.3c presents three bars which correspond to the mean scores for the vocabulary pre-test, vocabulary post-test and reading comprehension test. The X axis shows the three reading conditions and along the Y axis is the mean scores (maximum score of 19 for vocabulary and 3 for reading).

Summary of weekly performance

The most prominent observation from the quantitative data above is that improvements in the control condition only occurred after participants had read in a treatment condition. In other words, participants who read in the control condition during the first week did not exhibit significant improvements in vocabulary. This corroborates reports from the focus groups that participants had compensated for the removal of annotations by adopting inferencing strategies. It was thought that the stories may be a factor in determining the participants' performance. It is important to note, however, that the story in week 1 (Platypus) was the shortest of the three stories and contained the fewest number of target words. Thus, in theory, this story ought to be the easiest for participants, yet their vocabulary performance was superior in the subsequent two stories. No differences were observed in reading comprehension scores from one week to the next.

4.3.3 Clicking behaviour

Using the collected log files (see section 3.7.3 and Appendix IV), additional analyses were carried out to examine the participants' clicking behaviour. This involved looking at whether participants clicked on familiar words, whether a particular treatment results in more clicking, and whether there is a correlation between clicking on annotations and results on the tests. Even though some participants in the focus groups maintained that they had only used the annotations for words they knew, the familiar words that were inserted such as 'sun', 'moon' 'egg'...etc. were all accessed by all the participants. This shows a disparity between participants' perceived behaviour and their actual behaviour (see section 5.3.1 for participants' comments). The provision of the tracking tool (log files) in this study, therefore, proved beneficial for providing insight into this matter. It was also hypothesised that participants who had scored low on the pre-tests may consult annotations more frequently in order to improve their vocabulary. Similarly, it was hypothesised that an increase in clicking on annotations would result in a higher mark on the post-tests. The following sections discuss both the preparation of data and the statistical tests used to examine these hypotheses.

Testing clicking frequency

To test whether participants clicked more in a particular condition, the average number of clicks per word for each participant in both conditions was inputted into SPSS for analysis (see table). It was decided that determining the average was the best choice as it would control the different number of words in the stories.

Participant	Treatment A	Treatment B
1	7*	5
7	9	3

*This represents the average number of times participant '1' clicked on each word in the story that he/she read in the Treatment A condition.

After determining that the skew and kurtosis levels for Treatment A and B log files were satisfied (.308 and .548 respectively), a Paired Samples t Test was used to compare students' use of annotations in the two treatment conditions. A significant difference was found between the effect of Treatment A ($M = 5.07$, $SD = 2.68$) and Treatment B ($M = 4.25$, $SD = 1.92$) on annotation consultation; $t(111) = 3.02$, $p = 0.003$. This suggests

that participants consulted annotations more frequently in the Treatment A condition (Picture + Audio).

Testing correlations

A scatter plot was first created to check whether a linear relationship exists between the variables. A visual inspection of the scatter plot graph indicated that the distribution of the log files data was nonlinear. Further numerical tests revealed that the data were indeed non-normally distributed with a skewness of 0.622 ($SE = 0.228$) and kurtosis of 0.514 ($SE = 0.453$). The Shapiro-Wilk test also confirmed that the normality assumption was violated ($p < .05$). Based on these findings, the Pearson's correlation test was ruled out as a valid test for correlation. The next step was to determine whether the Pearson's non-parametric equivalent (Spearman's Rank correlation) could be used. A lack of monotonicity in the data, however, meant that this test could not be used. It was necessary, therefore, to perform a non-linear transformation of the data. The Log10 transformation was performed and a normality test was re-administered to check the data. Numerical results reflected a normality in the distribution of data with a skewness of $-.635$ ($SE = .228$) and kurtosis of $.708$ ($SE = .453$). Since the data transformation was successful and there were no significant outliers, the Pearson correlation test became fit for use. This enabled the researcher to test for correlations between two variables (frequency of clicking and results on the tests). Two hypotheses were posed as follows:

Hypothesis 5: Performance on the pre-test has no effect on frequency of consulting annotations

A Pearson's product-moment correlation was run to assess the relationship between pre-test scores and frequency of consulting annotations. No correlation was found between the two variables $r(-.056)$, $n = 112$, $p = .558$ and the null hypothesis, therefore, cannot be rejected. In other words, there is no evidence to suggest that a low score in the pre-tests results in more clicking on annotations when reading or vice versa.

Hypothesis 6: Frequency of consulting annotations has no effect on post-test scores

To examine the relationship between frequency of consulting annotations and post-test scores, the Pearson's product-moment correlation was used. Results showed no correlation between the two variables $r(-.038)$, $n = 112$, $p = .690$. The null hypothesis

cannot be rejected as the findings do not show that more clicking leads to higher results on the post-test, or vice versa.

Summary

This section has presented the findings from quantitative and qualitative data collected to answer the research questions. When comparing the vocabulary pre-tests and post-tests results, participants' performance on the post-tests improved significantly after reading with annotations. A significant improvement on the post-tests was also observed in the control condition but reading with annotations yielded significantly higher results. No particular annotation type (Treatment A or Treatment B) was more effective for immediate vocabulary recall but both were significantly more effective than the control reading condition. With respect to reading comprehension, no significant differences were found between the three reading conditions. The analysis of participants' weekly performance revealed that, with the exception of the control condition in week 1, vocabulary post-test results were significantly higher than the pre-test, confirming an improvement in performance. No significant difference was witnessed, however, on the reading comprehension test. The log files were also analysed and results demonstrated that participants accessed annotations more frequently in the Treatment A condition. The possible reasons for this are discussed in the next chapter. No correlation was found between results on the tests and access frequency, i.e. the pre-test results do not determine the number of clicks, nor do the number of clicks determine the post-test results. The qualitative data was tabulated and referenced to the focus group transcripts in Appendix VI. The findings varied and, in most cases, no consensus amongst participants was reached in terms of reading behaviour and preferences. The following chapter, however, will elaborate on these findings and the quantitative ones.

Chapter 5. Discussion

This penultimate chapter presents the interpretations of the results and relates them to the theories and literature discussed previously in Chapter II. The discussion is divided into three main sections which correspond to the study's three research questions. Section 5.1 analyses the effect of annotations on vocabulary recall, Section 5.2 examines their effect on reading comprehension and Section 5.3 looks at participants' reading behaviour and perceptions. Each section answers the research questions through an elaboration on both the quantitative and qualitative findings presented in the previous chapter. Where applicable, verbatim conversation from the focus groups is inserted and referenced to the transcripts in Appendix VI using the same approach as in table 4.3.1.

5.1 Effect of Annotations on Immediate Vocabulary Recall

This section reports on the findings pertaining to the thesis' first research question and sub-question. It looks at the effect of annotations and annotation type on immediate vocabulary recall.

R-Q 1: Do hypermedia annotations support vocabulary learning for young Kuwaiti learners?

The Paired Samples T-Test was first used to compare the results of pre-tests and post-tests from reading with annotations. The test revealed that participants' results were significantly higher on the post-test for immediate vocabulary recall ($p < .001$). This suggests that the annotations facilitated the comprehension and retention of unknown vocabulary items. The results from the post-tests ($M = 60.71$, $SD = 20.45$) were also compared to those from the control condition ($M = 48.33$, $SD = 22.84$) through the Paired Samples T-Test. A significant difference was found between the two with the annotation condition yielding a higher average of scores ($p < .001$, effect size = 0.57). These result mirror those of other studies, such as Abraham (2007) and Rezaee and Shoar (2011), which show a positive effect of annotations on immediate vocabulary recall.

An interesting finding, however, was that an improvement in post-test results also took place after participants had read in the control condition (pre: $M = 42.9$, $SD = 21.9$, post: $M = 48.33$, $SD = 22.84$). The Paired Samples T-Test revealed that this improvement in mean scores was also statistically significant. However, when comparing this improvement with that made in the annotation condition, the latter was significantly higher.

The positive effect of annotations can be explained in light of the theories discussed in Chapter II. The interaction hypothesis principles are manifested in the presence of highlighted words throughout the text. These may have attracted learners' attention and initiated the process of 'negotiation of meaning' (Long, 1996) whereby the annotations were used to build meaning. The annotations also provided participants with multiple modes of input (visual, textual and audio) which, according to the GTM, simplifies the input. The provision of multiple modes of input may have also activated more networks in the brain and allowed the learners to make stronger connections and build meaning (see section 2.1 on connectionism). Additionally, participants also had the freedom to select and organise information when reading with annotations and this may have also aided comprehension.

With regard to the witnessed improvement in the control condition, qualitative data shows that participants may have resorted to other means of deducing meaning. Participants reported that they used the text itself and/or the story images on each page to help decode meaning. When asked about deducing meaning in the control condition, some participants stated that they turned to the images in the story:

"H: Teacher I looked at the picture and read the word and then I knew it" (T1, 136)

"L: Mmm I see the pictures" (T2, 176)

"L: Teacher if there are pictures it helps you focus on what you're reading"
(T2, 304)

"Sg: You need pictures for the difficult words and the easy ones you don't" (T2,
306)

The effect of using illustrations was discussed in the literature with Pike et al. (2010) suggesting that illustrations provide important contextual cues and facilitate comprehension. However, it was also argued that young learners are likely to become dependent on images to establish meaning because they are more salient than text.

Below is a comment from a participant which exemplifies this. It is made in reference to the main images in the story, rather than the annotations:

“Teacher if there’re pictures like you can know the story quicker if there aren’t you have to read to know.” (T2, 302)

It can, therefore, be speculated that in the treatment conditions, participants negotiated meaning through annotations, whereas in the control condition, the process was carried out through the text and main image. The differences in reading strategies between the two conditions can also be explained in terms of top-down and bottom-up processing discussed in section 2.2. Participants are likely to have used a predominantly bottom-up approach when reading with annotations but a top-down one in their absence. In both reading conditions, participants demonstrated the ability to tackle texts that contain a large number of unknown words, i.e. higher than their vocabulary threshold. The findings of this research thus concur with the view that annotations may reduce the need for a threshold (Akbulut, 2007; Nation, 2001; Jacobs et al., 1994 and Long, 2007) while also illustrating that, given the correct input, young learners may be able to read at a higher level by decoding meaning from other sources.

The Involvement Load Hypothesis (Laufer and Hulstijn, 2001) predicts that annotations may not be an effective word learning tool because it requires less mental effort than deciphering meaning from context. Despite these suggestions, the findings of this study show that while learners may have indeed depended on images, particularly in the control condition, this did not distract them from paying attention to the written forms of the words and improving their results on the post-tests. This is crucial since the literature regarding the cognitive load also suggested that annotations are more effective with adults or learners with higher cognitive abilities. It was, thus, thought that their presence might cause a cognitive overload or a ‘split-attention’ effect. It is also worth noting that the vocabulary uptake in this experiment was not confined by the number of annotated words per story which ranged from 10-22 words. The findings, therefore, match those of Madrid et al. (2009) who observed no negative effect of the number of links in a text on cognitive load. The findings also favour Marcus et al.’s (1996) suggestion that the presence of images can reduce the demands on the working memory when processing text.

5.1.1 *Effect of annotation type on immediate vocabulary recall*

This thesis also conducted a more in-depth examination of annotations by testing whether a particular combination of information is superior for vocabulary recall. The basis for this examination was Paivio's dual coding theory and its subsidiary, the Modality Principle. The two treatments used for this were 'Picture + Audio' annotations (Treatment A) and 'Picture + Text' annotations (Treatment B).

Sub-Q 1: Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for L2 vocabulary recall?

A statistical analysis was carried out on the pre-tests and post-tests results in the two treatment conditions and the control. The highest improvement in post-test scores (mean difference) was for the Treatment A condition (19.64), followed by Treatment B (14.57) and the Control (5.38). The Paired Samples T-Test revealed that, in all three reading conditions, participants performed better on the post-tests and that this improvement was at a statistically significant level ($p < .001$). While the post-test results from the control condition were significant, those from the two treatment conditions remained significantly higher. In other words, participants improved in all reading conditions, but with annotations their improvement was statistically greater. Although the post-test scores were higher for the Treatment A condition, the difference was not statistically significant. Hence, it cannot be said that annotations in the form of picture + pronunciation are more successful than picture + text.

The vast majority of studies in the literature review established that dual-coded annotations result in better vocabulary retention than single-coded annotations (Abraham, 2007; Akbulut, 2007; Chun and Plass, 1996; Turk and Ercetin, 2014; Yoshi and Flaitz's, 2002). Acha's study (2009) with young learners, however, concluded with a negative effect of the dual-coded annotations. She argued that dual-coding places a heavy cognitive load on young learners who have a limited working memory. Acha's study was critiqued in the literature review and the explanation of the negative findings was centred mainly on the presence of all the annotations on a single page and in the form of pop-up windows. No evidence was found in this study to suggest that the young learners' performance was negatively impacted by the dual-coded annotations.

The literature on the dual-coding theory suggested that adhering to the 'modality principle' by combining audio and pictorial information yields more favourable results

than the combination of text + picture. The findings of this study, to a degree, concur with the modality principle as the Treatment A condition (picture + audio) resulted in the highest improvement in post-test results. The theory behind the success of combining audio + picture is that it is less demanding on the visual system than text + picture. Although text is processed in the verbal system, it initially passes through the visual system and when coupled with pictures, the two compete for visual attention and may cause a split-attention effect (Mayer and Moreno, 2002). The findings of this study, however, pose an additional explanation as to why picture + audio can be more successful. Due to the inherent nature of audio materials, learners reading with audio input may feel the need to listen more than once to grasp the input. The log files from this study confirmed that participants consulted the annotations more frequently in the Treatment A condition (see section 4.3.3). The fact that the pronunciation of the annotated words was in the L2 may have magnified the need to repeatedly click in order to hear the input. This repetition may have, in turn, facilitated the retention of the annotated words and resulted in higher post-test scores. While the connectionist view does not concur with the compartmentalisation of the mind into audio and visual systems, it can offer an explanation as to why the picture + audio combination was more favourable in this experiment. The repeated clicking may have aided the short-term memory through the process of rehearsal (see section 2.6.2). Learners may have subvocally articulated the pronunciation and this in turn aided with their retention. There may also be other variables that made reading with audio more favourable, such as the novelty effect or even personal preference, but further evidence is needed to confirm whether this is the case.

5.2 Effect of Annotations on Reading Comprehension

The second research question aimed to examine whether the presence of annotations facilitates reading comprehension for young learners. Examining the effects of annotations on reading was deemed crucial since it was speculated that their presence may unwittingly hinder the reading process by creating a distraction and a cognitive overload.

R-Q 2: Do hypermedia annotations facilitate reading comprehension for young Kuwaiti learners?

Analysis of the results revealed that when participants read the electronic stories with annotations they did not perform significantly better on the reading comprehension post-test compared to when they had no access to annotations. The presence of annotations did not, therefore, improve participants' reading comprehension. In fact, reading in the control condition resulted in a higher mean score ($M = 1.96$, $SD = 0.96$) than the annotation condition ($M = 1.79$, $SD = 0.82$). The T-Test, however, revealed that the slightly higher mean score of the control condition was not statistically significant. Thus, it also cannot be concluded that reading without annotations is better for general comprehension.

Unlike Sakar and Ercetin (2005) who found annotations to be detrimental to the reading process, participants with access to annotations in this experiment did not appear to have been affected negatively by annotations. Their performance was similar to the experimental group in Cheng and Good's study (2009) who achieved comparable scores to the control group. Similarly, the performance of advanced learners who read with annotations in Ariew and Ercetin's (2004) study was equal to the control on the reading comprehension test.

From a theoretical standpoint, the similar performances in both conditions may be attributable to a number of external factors. Participants may have been able to follow the events of the stories by simply looking at the pictures. Focus group data confirm that participants did indeed resort to the main images in the story to derive meaning. This was particularly the case in the control condition (see table 4.3.1). The stories themselves may have also been simple in terms of their concepts and thus there was no need to rely on annotations for constructing overall meaning. If this indeed was the case, it brings to light a potential issue with using authentic stories with EFL learners. The stories that are at an appropriate vocabulary level for these learners may be too simple in concept and the ones with more complex concepts are likely to contain more advanced vocabulary. It was initially thought that the stories in this experiment may be too easy for the participants, but the pilot study demonstrated otherwise. Participants in the pilot study reported positively on the themes of the stories and some stated that the words were difficult. In addition to the concepts of the stories, another factor which may have aided reading comprehension is the layout of the electronic texts. The stories in the experiment were presented in a book format (i.e. page by page) rather than a full text presentation. This meant that the participants were not bombarded with too much

information and they had the ability to navigate from one page to the next at their own pace. This type of self-paced learning is believed to be of particular benefit to beginners (Ariew and Ercetin 2004 and AbuSeileek, 2011) and may have thus facilitated the comprehension of texts in the experiment. Participants were also made aware of a pending post-test that included a reading comprehension component. This may have also ensured that participants do not neglect the overall meaning of the stories. Focus group data confirm that the presence of the pre-tests and post-tests impacted the way in which participants approached the texts. Participants reported paying extra attention to the story when no help was available in order to perform well on the post-tests. A detailed discussion on the effects of these tests is presented in section 5.3.3.

5.2.1 Effect of annotation type on reading comprehension

This research also set out to examine whether the type of dual-coded information has an impact on the overall comprehension of an annotated text either positively or negatively.

Sub-Q 2: Which annotation combination (picture +L1 translation or picture + L2 pronunciation) is better for reading comprehension?

It was thought that the picture + L2 pronunciation annotations may be more favourable for reading comprehension since they are based on the Modality Principle (Combining picture + audio is more effective than picture + text. See section 2.6.2). Furthermore, Myong's (2005) research suggested that L2 annotations are more successful for reading comprehension outcomes than L1 annotations. The analysis of this study, however, showed no significant differences in reading comprehension scores between the two treatments (Treatment A: $M = 1.79$, $SD = 1.05$, Treatment B: 1.80 , $SD = 1.02$). These findings replicate those of Tabatabaei and Shams (2010) and Chen (2006) who also found no significant differences in reading comprehension scores amongst participants who were subjected to three reading conditions. While the dual-coded annotations did not positively affect reading comprehension, it is also important to note that they did not have an adverse effect on reading. Thus, it can be argued that the vocabulary gains made from reading with annotations do not come at the cost of overall reading comprehension. This finding is crucial because the literature stated that young learners' working memory capacity is lower than adults and that this difference is likely to cause problems or hinder reading and comprehension.

5.3 Learners' Reading Behaviour and Preferences

The third and final research question looked at participants' reading behaviour and use of annotations. This was done by gathering quantitative data from log files that recorded participants' access of annotations and qualitative data from focus groups to examine participants' attitudes towards reading with annotations. Verbatim conversation is inserted throughout the discussion and referenced using the same approach as in table 4.3.1.

R-Q 3: How do young Kuwaiti learners use annotated texts and how do they feel about this type of reading?

The combination of qualitative and quantitative data gathered from this research question not only enriched the findings of the two sections above but also provided further insight into how young learners use annotated texts. This section includes a detailed discussion covering three areas of investigation. The first looks at participants' reading strategies and use of annotations. This covers the strategies used to deduce meaning in both annotated and un-annotated texts. The second sub-section reports on participants' preferences for reading condition and whether these match their performances. The third subsection sheds light on additional themes that were not accounted for but emerged from the focus group discussion.

5.3.1 Learners' reading strategies and use of annotations

Part of the focus group was dedicated to exploring the participants' reading behaviour. This involved examining whether inferencing strategies were employed and whether these were subject to change according to the reading condition. It was first necessary to establish how participants go about deducing meaning of unknown words from regular texts. This was to provide a starting point for the investigation into how annotations may alter learners' current behaviour. When questioned, participants stated that they establish meaning by asking an adult or parent (T1, 31; T2, 69), obtaining an online translation (T2, 73), and using dictionaries (T1, 40). It was interesting to discover that the young learners consult adults or dictionaries because the literature promoted the use of annotations to encourage learner independency and as a replacement for dictionaries. One can assume that long term exposure to annotations may well change the way in

which learners extrapolate meaning. Participants were also asked about the way they approached the un-annotated texts in the experiment. The purpose of this was to probe further into the type of strategies they use, particularly when no external sources of input are available (adults/dictionaries). Participants reported using a number of techniques to understand both the vocabulary and events of the stories. Some participants looked at the main picture on each page to decode meaning (T1, 136; T2, 176), while others tried to derive meaning from the text itself (T2, 179) or remember clues from the pre-test (T2, 177).

One participant who had first read in the two treatment conditions said that she had exerted extra effort in the control condition since no aid was available. Her comment is presented below along with the reaction of other participants' to reading un-annotated texts (control condition).

“Ko: Teacher those others [annotations] helped me in the two stories but in this one now, I mean I had to read it all properly because it might come in the test and I don't know it or haven't memorised it, so I have to remember it” (T1, 246)

“M: The hardest thing was the third story, the tiger, because it was without words [annotated words] and it was difficult to remember all the words because they might come in the test and we won't know them. So we had to think hard and focus on it so we know and get it right” (T1, 251)

“L: Teacher I focused more when there weren't blue because it doesn't show that this word is there and that word is there” (T2, 268)

These comments demonstrate that when no peripheral help was available in the experiment, participants made a conscious effort to decode meaning through focusing on the text. This gives credence to the assumption laid out in section 5.2 regarding the reason for this similarity in reading comprehension performance either with or without annotations. It appears that participants actively compensated for the lack of annotations by paying extra attention to the text and words. In contrast, one participant reported using a different approach:

*“Sg: Teacher like the words I don’t know or understand I just skip them...
I don’t have to read them” (T2, 230)*

This participant sacrificed word meaning for general understanding. This in itself is considered a strategy since it is common for L2 readers to focus too closely on single words in a text. Martinez-Lage (1997) describes such learners as those who "perceive the text as a series of isolated words, each of which has to be deciphered individually in order to move on with reading" (Martinez-Lage 1997: 122). The issue with this type of reading is that it can lead to a cognitive overload and loss of overall sense of meaning (Abraham, 2007).

The comments above show that participants adapted their strategies according to the reading condition. It is important to note, however, that these participants had first read in a treatment condition. The counterbalancing method used in the experiment meant that the order of the three reading conditions was alternated over the three week period (see section 3.5). It was thus thought that participants’ resourcefulness may have been sparked by the removal of annotations. This speculation prompted a statistical investigation into participants’ weekly performances (see section 4.3.2) to establish whether the order of reading is indeed a determining factor in strategy use. The statistical analysis revealed that with the first story (Platypus) in week 1, the group who read in the control condition did not perform significantly better on the post-test. This finding was crucial because it does not match that of the groups who read in the control condition in weeks 2 or 3, i.e. after being exposed to one or two treatment conditions. The later groups produced significantly improved performances on the post-tests, indicating that the order of reading may indeed have altered the participants’ reading behaviour. In addition to provoking alternative strategies, the order of reading may have simply familiarised participants with the reading procedure and the pre and post-tests. This possibly reduced the extraneous load of the task and ‘freed up’ some working memory capacity allowing learners to employ other strategies during the consecutive weeks.

In addition to examining participants’ strategies in un-annotated texts, it was also important to check whether any strategies were used when help was available via annotations. This is because the readily available help may cause participants to abandon their inferencing strategies and rely solely on annotations for constructing meaning. The focus group discussion was, therefore, designed to explore when and how

participants consult annotations. One of the questions posed was whether participants consulted annotations for words they already knew. In theory, clicking on unnecessary information (annotations for familiar words) may have a distracting effect and it would, therefore, be ideal if learners used annotations selectively. A number of participants reported that they had only clicked on unknown words. Examples of their responses are as follows:

“If I know it [the annotated word] I don’t click on it” (T2, 130)

“aaah I clicked on the words I didn’t know and didn’t click on the ones I knew” (T2, 134)

“L: teacher me only the ones I don’t know

L: teacher me too

L: the same” (T2, 143-145)

Although participants thought they had only used annotations for unknown words, their log files showed that all annotations were accessed. Familiar words such as ‘sun’ and ‘moon’ were annotated intentionally for purposes of examining this type of clicking behaviour. Log files showed that these words were indeed accessed. Participants’ perception in this case was at odds with their actual behaviour. Other participants acknowledged that they had consulted annotations for words they already knew. Their explanation for this is as follows:

“L: Teacher I mean I might know a word before I click on it because I know the words that come before it and after it.

T: Aha so you clicked on words which you knew?

L: Yes and we clicked on the words we didn’t know

T: Ok if you knew a word why did you click on it?

L: To see if it’s right or wrong

T: Ah Ok.

L: But Teacher if there is a word I know and sure of, I don’t click on it” (T1, 75-81)

“L: Teacher I clicked on the ones I know and the ones I don’t know

T: and why did you do that?

L: Teacher to find out if I’d understood it correctly or not”

(T2, 136-139)

“Teacher I clicked the ones I knew and the ones I didn’t know to be sure” (T2, 146)

The statements above indicate that clicking on familiar words can be a purposeful act of providing reassurance about word meaning. This reiterates the advantage of using annotations as a means of reducing incorrect guessing (Lenders, 2008). From a theoretical perspective, the Involvement Load Hypothesis emphasises that the increased mental effort from inferring meaning is more beneficial for retention (Laufer and Hulstijn, 2001) and development (Ellis, 1995). It was thus debated that readily available meaning in annotations may reduce these desirable effects. The findings here, however, demonstrate that the presence of annotations does not necessarily equate with reduced mental effort. The findings agree with the argument that annotations are needed, particularly for beginners, to minimise incorrect inferences (Paribakht and Wesche, 1997; Peters et al., 2009 and Lenders, 2008). Further evidence of this is apparent from the following excerpts:

“Teacher I kinda thought about it then clicked” (T2, 124)

“Teacher I tried to read it I couldn’t then I clicked” (T2, 126)

The above statements offer some insight into how annotations were used in the experiment. To expand on this, participants were also asked about the number of times they clicked on a single annotation. This question, however, did not yield a wealth of responses but two participants explained their clicking behaviour as follows:

“If it’s difficult I click repetitively and if not just once” (T2, 150)

“Teacher but sometimes you click on it because you don’t know it you click on it a lot”

(T2, 156)

It was thought that participants might compensate for a low mark on the pre-tests by consulting annotations more frequently. Likewise, it was speculated that a higher rate of

consulting may lead to a higher score on the post-tests. Using quantitative data, an attempt was thus made to determine whether a correlation exists between frequency of clicking on annotations and results on the pre-tests and post-test (see section 4.3.3). No correlations were observed between these variables. These findings are in accordance with other research showing no significant relationship between number of look-up times per word and post-test results (Abraham, 2007). One possible explanation for this behaviour is the novelty effect of seeing annotations appear and re-appear with repetitive clicking. A longitudinal study would provide insight into whether young learners' clicking behaviour changes over time.

5.3.2 Learner preferences for reading condition

This research took into account the role of learner preferences and attitudes regarding them as key to the success of CALL material. The discussion in section 2.7.3 showed that when learners are interested in the material, they are more likely to reap its benefits. The focus groups were thus constructed to encourage the learners to reflect on the types of reading they were exposed to and to express their views.

In terms of the three reading conditions, participants' responses differed considerably and no consensus was reached regarding the optimum reading condition. Those who preferred the audio + picture annotations reported that the pronunciation had helped them read the difficult words (T1, 62) and that using the headphones was more beneficial than the Arabic translation (T2, 93). As for the participants who preferred the text + picture treatment, their comments centred on the belief that the translation was better at getting the meaning across:

“Teacher the Arabic is easier because teacher the word in English I don't understand what it is” (T2, 99).

“Because I mean it [the Arabic] helps me with this word and that word” (T1, 56)

Mh: Because the words I don't know I'll know them in Arabic, I'll know how to say them in English (T1, 87).

The variety of responses demonstrates that no particular annotation type is more popular. It may be that in some cases, where the images in the annotations are not clear, the L1 translation is preferred to reinforce meaning. In contrast, when the images are clearly

representative of meaning, the L2 pronunciation will aid in reading the new word. Since all participants were at the same proficiency level (beginners), these findings do not match those of research showing proficiency as a determining factor in preference for annotation type (Ballester and Rodriguez, 2010; Ercetin, 2003; Myong, 2005). The differences in preferences may, therefore, stem from participants' individual learning styles (verbal, visual ...etc.). It is also worth noting that there were participants who expressed their preference for the control condition stating that it is easier (T2, 119) because it had words which they knew (T1, 54; T1, 64).

It was necessary to probe further into why some participants preferred reading in the control condition. One participant explained that he was able to focus more in the control condition because the highlighted words were distracting. His classmates contributed to the discussion to explain to the researcher the situation. The conversation that took place is presented below:

“F: when there are no words [annotated words]

T: You felt you focused more?

F: Yes

T: Why?

F: Because when there are blue words you stop there and then forget

T: How do you mean?

F: I mean if someone stops at a word next to the blue he might forget his place

L: Teacher he means..

Sg: there are too many blue words

L: He is reading for a while then stops, by the time he reads, by the time he understands the word he forgets what he is reading

T: Is that what you mean?

Aa: when there are many blue words next to each other and someone stops at a blue place like later when he returns he forgets his place

T: aha ok so you forget your place on the page

Sg: teacher he means there are too many blues” (T2, 249-262)

This short conversation opens the door to more questions regarding the core issues of cognitive overload that were debated in the literature review. In terms of vocabulary gains, the quantitative data produced no evidence of cognitive overload as participants'

vocabulary performance after reading with annotations was significantly higher than with the control. In terms of general reading comprehension, participants' performance was equal in both conditions and thus no proof of distraction was evident. Research has shown that it is not unusual to find a mismatch between participants' preferences and actual performances (Abuseileek, 2008; Ariew and Ercetin, 2004 and Cheng and Good, 2009) and for these reasons it is necessary to collect both quantitative and qualitative data. The statements above, however, cannot be dismissed particularly since a number of participants contributed to the discussion which indicates a possible underlying issue. The expression "lose his place" suggests that returning to the text is problematic after looking at the marginal annotations. Other comments such as "forgets what he is reading" indicate difficulty in remembering the context or event in which the annotated word had occurred. Given that participants' actual performances on the post-tests did not match their perceptions, one can speculate that participants may indeed have felt an extra mental imposition but were able to compensate for this by spending extra time going back and forth between the annotations and text.

5.3.3 Emerging themes

One of the justifications for using focus groups in this research was to provide a basis from which unexpected, yet important, issues can emerge. Early into the focus group discussion, the most prominent observation was that the presence of the pre-tests and post-tests influenced the participants' reading behaviour. Despite the knowledge that the experiment would not affect their school grades, the tests heightened participants' awareness to the electronic texts and the task requirements. This was particularly the case in the second and third weeks as participants' familiarity with the procedures grew. The following excerpts from the focus groups exemplify the effect of the tests:

"L: I like know the word like If I what's it called get it wrong in the first test I'll know it in the second because I read the whole story and know it and understand it"(T1, 149)

"L: Teacher I paid attention while reading so I wouldn't leave anything unanswered in the test" (T1, 159)

“Ko: I had to memorise the words so if they came in the test I’d know how to answer” (T1, 234).

“L: I mean for example you give us the test before... before we see the computer... then I see because if I have mistakes the second test after the story I’ll know the words” (T1, 147)

“L: like I know it like I understand it I remember it for the test” (T2, 185)

*“teacher any page.. any difficult word is in blue but the ones in black are all easy”
(T2, 239)*

These comments show participants’ drive to focus on the texts in order to check their answers on the pre-tests and achieve a good score on the post-tests. The last comment also reflects the participant’s process of identifying and predicting the target words. The fact that the target vocabulary became explicit is not a problem in itself since the experiment is based on an intentional, rather than incidental, learning condition. Furthermore, it was thought that a pre-test may have a desirable effect by activating learners’ prior knowledge and encouraging them to make connections (see schema mapping in section 2.6.2, p.21). The literature review also highlighted the possibility of a positive relationship between prior knowledge and learning outcomes (Ariew, 2004; Moos and Marroquin, 2010). The post-tests, on the other hand, provided purpose to the reading task and this mimics the natural reading practice in EFL classrooms. Adding a reading comprehension component to the post-tests also encouraged participants to pay attention to the stories as a whole rather than simply scanning through the texts for highlighted words.

All in all, the findings from this research do not concur with those carried out with adults of lower working memory capacity or those at beginner level of English. While the latter may struggle to cope with annotations and experience a cognitive overload, the children in this study were able to use the annotations and improve their immediate vocabulary recall without jeopardising their overall comprehension of the text.

Chapter 6. Conclusion

This thesis has examined the effect of annotations on reading comprehension and vocabulary learning. After introducing the topic and presenting the research problem in Chapter I, Chapter II gave a detailed review of the literature on annotations and the key underpinning theories. Chapter III then described the study design, research instruments and data collection methods, Chapter IV presented the results of the experiment and Chapter V discussed these in detail. This final chapter gives a brief summary of the study, highlights the pedagogical implications and offers recommendations for future research.

6.1 Brief Summary of the Study

In many cases, teachers may feel obliged to incorporate a CALL element into their teaching practice. The decision to introduce a CALL tool, however, ought to be based on sound theory and empirical research. The literature on annotations shows that they are, by and large, a successful tool for vocabulary learning and reading. However, the bulk of research in this area has been conducted with adults and, due to cognitive differences, the success of annotations with young learners remained questionable. This study set out to investigate how annotated texts impact young learners. The two key theories that underpinned the research were the interaction hypothesis and the generative theory of multimedia. Both theories emphasise the importance of drawing learners' attention to input in order to initiate negotiation of meaning and promote learning. The dual-coding theory, a subsidiary of the GTM, was also tested by comparing the efficacy of combining different modes of input in the annotations.

The findings of this study demonstrate that young learners too can benefit from reading with annotations. By comparing pre-tests and post-tests results, it was evident that participants' immediate vocabulary recall improved significantly after having read with annotations. The literature suggested that while annotations can aid with vocabulary, they may inadvertently cause a cognitive overload which could impede overall reading comprehension. The outcomes of this research, however, confirm that reading comprehension was not affected by annotations. Evidence for this came from the reading comprehension component of the post-tests which were similar for reading both

with and without annotations. Hence, the presence of annotations had a positive effect on vocabulary recall and a neutral effect on reading comprehension. It was thus concluded that using annotations to boost vocabulary learning will not come at the cost of disrupting the reading process or negatively impacting comprehension.

Although reading with annotations yielded positive results, a significant improvement in vocabulary post-tests results was, nevertheless, witnessed in the control condition. The absence of annotations provoked other inferencing strategies such as using the main pictures or texts in the stories to decode meaning. It is important to note, however, that while significant vocabulary improvements were made in the control condition, the improvements made from reading with annotations were significantly higher.

The investigation into the dual-coding aspect of the GTM revealed that no particular combination of input in the annotations is more favourable for immediate vocabulary recall or reading comprehension. The highest mean score for vocabulary post-tests came from the combination of audio + picture information. This falls in line with the modality principle within the dual-coding theory. The theoretical justification for promoting audio + pictorial information stems from the assumption that presenting text alongside pictures places a heavier demand on the visual systems. This study, however, suggests an alternative view as to why audio annotations may be more successful for vocabulary recall. The transitory nature of audio material is likely to prompt learners to click on the annotated words repeatedly in order to listen to the input. Log files from the experiment confirmed that annotations were accessed more frequently in the audio condition. The positive effects of audio + picture were not carried over, however, to the reading comprehension component of the post-tests.

In terms of learners' preferences, focus group data brought to light a mismatch between perception and actual performance. Some participants maintained that the presence of annotations was distracting. Since the suggestion of a cognitive overload was not substantiated through quantitative findings, it was hypothesised that learners may have spent extra time and effort reading and re-reading but this did not jeopardise their ultimate performance. The annotations also did not appear to have mitigated the desirable mental effort involved in establishing meaning. Evidence suggests that learners made guesses about word meaning prior to using the annotations for verification.

The data from the focus groups also revealed that the presence of the pre-test and post-test was influential. The tests appear to have encouraged learners to focus on the stories in order to check their answers on the pre-test and perform well on the post-tests. The theoretical and pedagogical implications of having a pre and post reading task are discussed below along with those of the other findings.

6.2 Theoretical Implications

This section looks at how the findings of this study can contribute to the theories and pedagogical practices concerning using annotations with young learners.

The involvement load hypothesis argues that exerting extra mental effort to construct meaning is important for retention. It was, therefore, thought that using annotations may, to a degree, encourage learners to become passive, rather than active, constructors of meaning. The findings of this study, however, indicate that learners were actively navigating through the text and annotations in order to connect the two forms of input. Learners also used the annotations to make connections to the pre-tests and post-tests. Their immediate vocabulary recall after reading with annotations was significantly higher than when no annotations were available. It was thus concluded that readily available help through annotations does not necessarily reduce the desirable effects of increased mental effort.

6.3 Pedagogical Implications

The findings from this study have a number of pedagogical implications. Participants were able to read stories of different lengths and improve their vocabulary significantly without the need of the teacher to pre-teach the target items. They were able to use the annotations effectively to understand new words. This confirms the pedagogical value of using annotations to reduce teaching time and provide a more learner-centred approach to reading. Evidence also shows that when annotations were removed from subsequent texts, the young learners used contextual and pictorial clues to deduce meanings. They did not receive any type of training for this but were able to use such strategies intuitively. The effect of the tests administered before and after the reading task cannot be overlooked. The purpose of a pre-reading task is to activate learners' prior knowledge and create interest in the material. The findings suggest that the presence of a pre-test motivated learners to read and check their answers. The post-test also gave purpose to

the reading task as learners focused on the text in order to achieve an improved score. The pedagogical implications of this are that setting a goal for a multimedia reading task promotes both concentration and resourcefulness. Pedagogical implications can also be drawn from the layout of the stories and can inform future design. The way in which the stories and annotations were presented appears to have contributed significantly to the positive results from this study. Presenting the stories page by page may have reduced any cognitive overload effects found in other studies (e.g. Acha, 2009) where the full text was presented on a single page. Additionally, the young learners reported positively on the images both on the pages and in the annotations by stating that they had aided comprehension.

6.4 Evaluation of the Study

This section evaluates the underlying theories in this study, its overall design and execution.

It was important to choose a solid theoretical framework for the study. This became apparent during the literature review where studies that lacked such a framework created variables which were not accounted for nor controlled, thus producing findings which could not be explained adequately. The opening section of Chapter 2 discussed two general theories of vocabulary learning; constructivism and connectionism. Later sections then delved into more specific theories of learning and focused on those related to CALL, namely the Interaction Hypothesis and the GTM, which formed the theoretical basis of the research. The findings were, by and large, explainable through the Interaction Hypothesis and the GTM but in some cases a referral to the more general theories in Section 2.1 proved useful. For example, the findings showed that the picture + audio input yielded better results on the vocabulary post-test. The Dual-coding theory (a subsidiary of the GTM) explained the favourability of this combination of input from a verbal/visual system perspective, whereas the connectionism theory demonstrated that rehearsal is key to retention. Neither the GTM nor connectionism, however, could provide insight into why improvements were witnessed in the control condition whereby no additional help was provided. It was through the Interaction Hypothesis that a possible explanation emerged. The improvements were likely to have been a result of change in learners' strategy (interaction) with the text. The removal of annotations

encouraged them to use other means of deducing meaning through top-down and bottom-up processes. The findings of this study have shown that rather than aiming to fully prove or reject a theory, it is important to evaluate theories and examine how they may complement one another to better our understanding of learning.

With respect to design, the points of strength in this experimental research lie within the levels of control that were maintained through the use of random sampling, multiple texts and the counterbalancing method. Additionally, the study gathered both quantitative and qualitative data which worked hand in hand to consolidate the findings. Two issues that often arise in empirical research are a restricted timeframe and the novelty effect. Although this study made use of multiple texts over a period of three weeks, more time is needed to form a better picture of how participants read annotated texts. A longitudinal study would not only allow for the novelty effect to wear off but would also enable the researcher to gather information on whether there are long-term gains to be made from reading with annotations. The use of the software Adobe Authorware was effective in the sense that it created and presented the annotations correctly and stored all the log files efficiently. However, there are issues which may be of concern to those wishing to use the software. Firstly, it appears that Authorware is becoming redundant as the last update to the software took place in 2007. The software file is also large in size and needs to be installed on every PC, rather than uploaded to the internet. Bearing this in mind, it may be advisable to look into current software to create electronic texts with annotations.

In terms of the data collection, some issues had surfaced but were not deemed detrimental to the research. An effort was made to ensure that the reading sessions did not take place during the first or final lesson of the day. This is because learners' attention is likely to be at its minimum during those times. To further control this issue, it would have been ideal to conduct the reading sessions at the same time of day every week but this was not always possible. Furthermore, while the focus groups provided good data for the research, they may have been more fruitful had the groups of participants been more accustomed to the setting and procedure. This could have been achieved had the researcher had more time to organise a focus group after every reading session.

6.4 Conclusion and Recommendations for Future Research

This study has contributed to CALL research by adding a young learner perspective on the previous research into annotations with adults. The evidence shows that this learning tool is beneficial to young learners' vocabulary development, and more importantly, demonstrates that the presence of electronic features does not distract or hinder the overall focus on the task.

To further our understanding of the role of annotations in vocabulary learning and reading comprehension, it may be of benefit to present electronic stories/texts without the main pictures. The reason behind this is that participants reported using the pictures, particularly in the control condition, as contextual clues for meaning. Eliminating the pictures would, therefore, provide a more controlled condition for testing the efficacy of annotations and learners' inferencing strategies. A degree of the experiment's realism, however, would have to be sacrificed (see section 3.6). To gather more information about learners' mental effort, the reading time can be controlled so as to establish whether there are any differences in time spent between reading with and without annotations. The time factor can also be controlled to test whether learners are spending more time with audio annotations due to repetitive clicking. A longitudinal study would provide rich data showing how learners' reading and consultation of annotations can change over time. Participants may use the annotations less frequently or consult them only for new words as the novelty effect diminishes

Appendix I

Annotated Words in 'Platypus'

- | | |
|---------------|-----------|
| 1. Box | 6. Bucket |
| 2. Lunch | 7. Toe |
| 3. Seaweed | 8. Shell |
| 4. Sandcastle | 9. Spade |
| 5. Sea | 10. rock |

Annotated Words in 'The Very Hungry Caterpillar' story

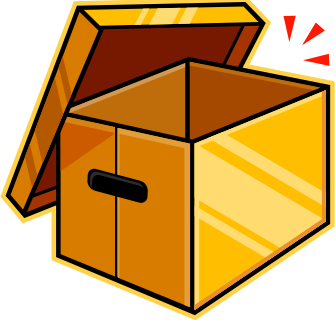


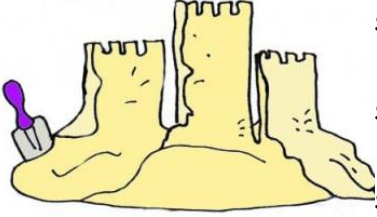



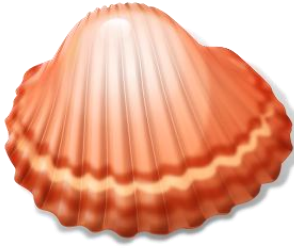


- | | |
|--------------------|------------------|
| 1. Caterpillar | 12. Pickle |
| 2. Egg | 13. Swiss cheese |
| 3. Sun | 14. Salami |
| 4. Moon | 15. Lollipop |
| 5. Leaf | 16. Cherry pie |
| 6. Pear | 17. Sausage |
| 7. Strawberries | 18. Cupcake |
| 8. Plums | 19. Watermelon |
| 9. oranges | 20. Stomach ache |
| 10. Chocolate cake | 21. Cocoon |
| 11. Ice-cream cone | 22. Butterfly |

Annotated Words in 'The Very Hungry Caterpillar' story





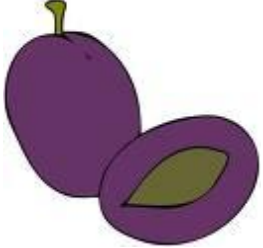
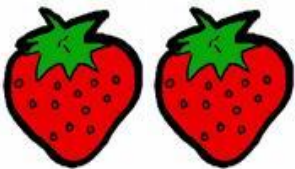




- | | |
|--------------|---------------|
| 1. Kitchen | 11. Ring |
| 2. Stripy | 12. Plate |
| 3. Sausages | 13. Furry |
| 4. Biscuits | 14. Cupboard |
| 5. Bun | 15. Tap |
| 6. Fridge | 16. Ice-cream |
| 7. Chips | 17. Tins |
| 8. Saucepans | 18. Packets |
| 9. Teapot | 19. Café |
| 10. Jug | |









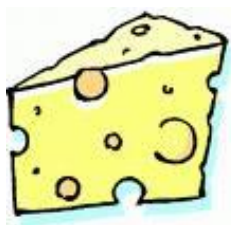


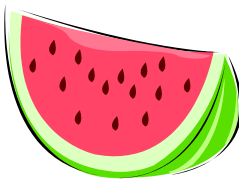
Appendix II (Pre-tests)

Pre-test for 'Platypus' Circle the correct word for each picture ضع دائرة على الكلمة المناسبة لكل صورة

<p>1</p>  <p>bite bill box</p>	<p>2</p>  <p>lunch ladder lounge</p>
<p>3</p>  <p>seaweed sword sunrise</p>	<p>4</p>  <p>seesaw sandcastle sandwich</p>
<p>5</p>  <p>seed snake sea</p>	<p>6</p>  <p>jacket racket bucket</p>
<p>7</p>  <p>toe glue key</p>	<p>8</p>  <p>sugar shell sheet</p>
<p>9</p>  <p>plane spade shoe</p>	<p>10</p>  <p>rock dock sock</p>

Pre-test 'The Very Hungry Caterpillar'
 ضع دائرة على الكلمة المناسبة لكل صورة Circle the correct word for each picture

<p>1</p>  <p>lavender lollipop lipstick</p>	<p>2</p>  <p>cupcake closet cricket</p>
<p>3</p>  <p>egg leg ear</p>	<p>4</p>  <p>biscuit building butterfly</p>
<p>5</p>  <p>drum plum pond</p>	<p>6</p>  <p>gooseberries strawberries blueberries</p>
<p>7</p>  <p>sun sea salt</p>	<p>8</p>  <p>mint moose moon</p>
<p>9</p>  <p>paint purse pear</p>	<p>10</p>  <p>coat cone club</p>

<p>11</p> 	<p>12</p> 
<p>13</p> 	<p>14</p> 
<p>15</p> 	<p>16</p> 
<p>17</p> 	<p>18</p> 
<p>19</p> 	<p>20</p> 
<p>21</p> 	<p>22</p> 

pickle
buckle
pistol

cherry pie
custard tart
meat pie

lock
leaf
land

sponge
scissors
sausage

back ache
head ache
stomach ache

sleeve
salami
slipper

caterpillar
cinnamon
courgette

carrot cake
chocolate bar
chocolate cake

soft cheese
cream cheese
swiss cheese

onion
octopus
orange







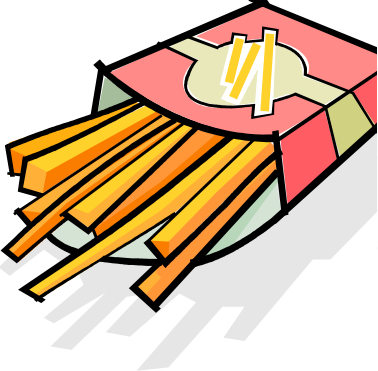
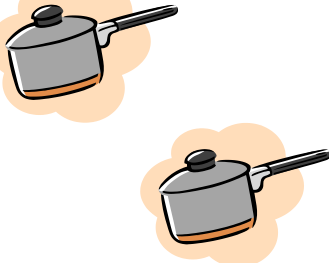


castle
cocoon
caravan

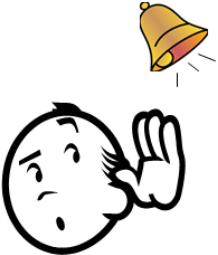








watermelon
windmill
whistle

Appendix II – Cont.

Pre-test 'The Tiger who Came to Tea'
Circle the correct word for each picture

ضع دائرة على الكلمة المناسبة لكل صورة



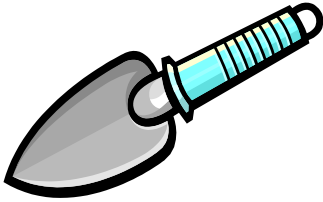

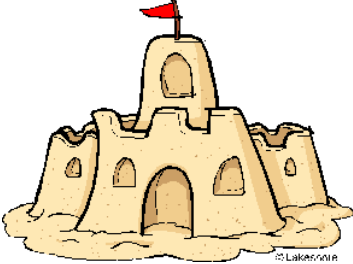





<p>1</p>  <p>kitchen karate knight</p>	<p>2</p>  <p>stars stripy sticky</p>
<p>3</p>  <p>sausages seasons sardines</p>	<p>4</p>  <p>battles berries biscuits</p>
<p>5</p>  <p>bin bib bun</p>	<p>6</p>  <p>flood fridge flies</p>
<p>7</p>  <p>chips chalk chain</p>	<p>8</p>  <p>saucepans seashells scissors</p>
<p>9</p>  <p>trumpet teapot ticket</p>	<p>10</p>  <p>jet jam jug</p>

 <p>rug ray ring</p>	<p>12</p>  <p>paper plate pocket</p>
 <p>field fence furry</p>	<p>14</p>  <p>cupboard cucumber compost</p>
 <p>mat rat tap</p>	<p>16</p>  <p>ice-cream instrument ice-skating</p>
 <p>tins tops twins</p>	<p>18</p>  <p>plasters packets pigeons</p>
 <p>card cafe cube</p>	

Appendix III (Post-tests)

Platypus post-test

ضع دائرة على الكلمة المناسبة لكل صورة Circle the correct word for each picture

<p>1</p>  <p>jacket racket bucket</p>	<p>2</p>  <p>sugar shell sheet</p>
<p>3</p>  <p>plane spade shoe</p>	<p>4</p>  <p>seed snake sea</p>
<p>5</p>  <p>seesaw sandcastle sandwich</p>	<p>6</p>  <p>bite bill box</p>
<p>7</p>  <p>rock dock sock</p>	<p>8</p>  <p>lunch ladder lounge</p>
<p>9</p>  <p>seaweed sword sunrise</p>	<p>10</p>  <p>toe glue key</p>

All reading post-tests were presented in the participants' L1 but a translation is available here.

Circle the correct answer for each question

ضع دائرة على الإجابة المناسبة لكل سؤال:

1. Where did the platypus go?

- a) to the sea
- b) to the park
- c) to the shop

1. أين ذهب البط؟

أ) إلى البحر

ب) إلى الحديقة

ت) إلى السوق

2. What did the platypus do?

- a) swam
- b) ate dinner
- c) looked for something to take

2. ماذا فعل البط؟

أ) تسبح

ب) تناول العشاء

ت) بحث عن شيء ليأخذه

3. What did the platypus put in the box?

- a) a sandcastle
- b) a big rock
- c) a shell

3. ماذا وضع البط في صندوقه؟










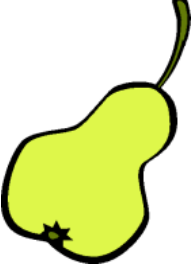
أ) قلعة رملية


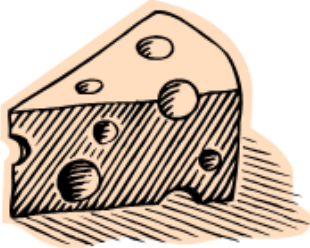





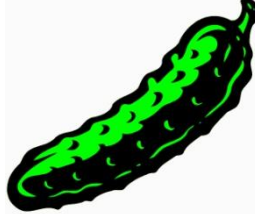

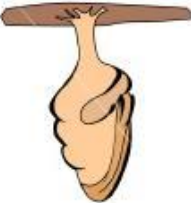


ب) صخرة كبيرة

ت) قوقعة

Appendix III – Cont.

Caterpillar post-test ضع دائرة على الكلمة المناسبة لكل صورة Circle the correct word for each picture

<p>1</p>  <p>Gooseberries Strawberries blueberries</p>	<p>2</p>  <p>egg leg ear</p>
<p>3</p>  <p>cupcake closet cricket</p>	<p>4</p>  <p>caterpillar cinnamon courgette</p>
<p>5</p>  <p>sponge scissors sausage</p>	<p>6</p>  <p>lavender lollipop lipstick</p>
<p>7</p>  <p>lock leaf land</p>	<p>8</p>  <p>watermelon windmill whistle</p>
<p>9</p>  <p>coat cone club</p>	<p>10</p>  <p>paint purse pear</p>

<p>11</p>  <p>carrot cake chocolate bar chocolate cake</p>	<p>12</p>  <p>soft cheese cream cheese swiss cheese</p>
<p>13</p>  <p>sun sea salt</p>	<p>14</p>  <p>drum plum pond</p>
<p>15</p>  <p>sleeve salami slipper</p>	<p>16</p>  <p>back ache head ache stomach ache</p>
<p>17</p>  <p>buffet building butterfly</p>	<p>18</p>  <p>pickle buckle pistol</p>
<p>19</p>  <p>cherry pie custard tart meat pie</p>	<p>20</p>  <p>castle cocoon caravan</p>
<p>21</p>  <p>onion octopus orange</p>	<p>22</p>  <p>mint moose moon</p>

Circle the correct answer for each question: ضع دائرة على الإجابة المناسبة لكل سؤال:

- 1. What did the caterpillar do when he came out of the egg?**
 - a) He sat on a leaf
 - b) He looked for something to eat
 - c) He fell asleep

- 2. Why did the caterpillar get a stomach ache?**
 - a) Because he felt hungry
 - b) Because he ate too much food
 - c) Because the food was bad




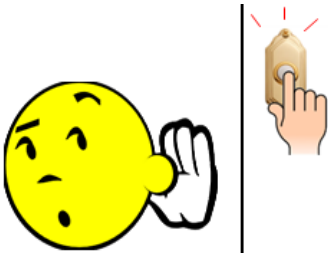

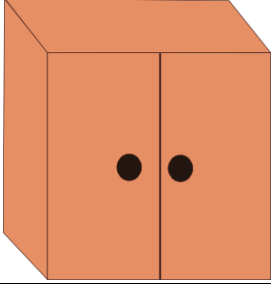




- 3. What happened to the caterpillar after he went inside the cocoon?**
 - a) He ate more food inside
 - b) He waited for the sun to come up
 - c) He started to change into a butterfly



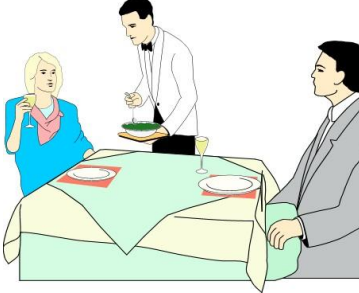
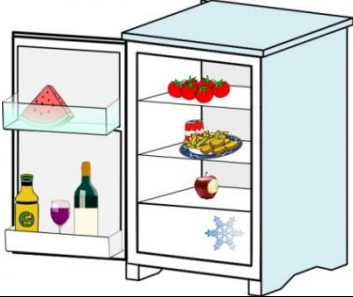





1. ماذا فعل اليرقة (الدودة) عندما خرج من البيضة؟
أ) جلس على ورقة شجر
ب) بحث عن الطعام
ت) نام

2. لماذا أصابه ألم في البطن؟
أ) لأنه شعر بالجوع
ب) لأنه أكل الكثير
ت) لأن الطعام كان فاسدا

3. ماذا حصل لليرقة بعد دخوله للشرقة؟
أ) أكل المزيد من الأكل
ب) انتظر ظهور الشمس
ت) بدأ بالتحول إلى فراشة

Circle the correct word for each picture

<p>1</p>  <p>paper plate pocket</p>	<p>2</p>  <p>mat rat tap</p>
<p>3</p>  <p>chips chalk chain</p>	<p>4</p>  <p>rug ray ring</p>
<p>5</p>  <p>tins tops twins</p>	<p>6</p>  <p>cupboard cucumber compost</p>
<p>7</p>  <p>sausages seasons sardines</p>	<p>8</p>  <p>trumpet teapot ticket</p>
<p>9</p>  <p>ice-cream instrument ice-skating</p>	<p>10</p>  <p>field fence furry</p>

<p>11</p>  <p>battles berries biscuits</p>	<p>12</p>  <p>kitchen karate knight</p>
<p>13</p>  <p>card cafe cube</p>	<p>14</p>  <p>flood fridge flies</p>
<p>15</p>  <p>stars stripy sticky</p>	<p>16</p>  <p>jet jam jug</p>
<p>17</p>  <p>bins bibs buns</p>	<p>18</p>  <p>plasters packets pigeons</p>
<p>19</p>  <p>saucepans seashells scissors</p>	

Circle the correct answer for each question: ضع دائرة على الإجابة المناسبة لكل سؤال:

1. Why did the tiger come to Sophie's house?

- a) To see Sophie's daddy
- b) Because he was hungry
- c) To play with Sophie

2. What did the tiger do in the house?

- a) He played with Sophie
- b) He helped Sophie's mummy in the kitchen
- c) He ate all the food

3. Why did Sophie and her parents go to a cafe?

- a) Because Sophie's mummy forgot to cook supper
- b) Because Sophie wanted to eat at the cafe
- c) Because the tiger had eaten all the food at home

1. لماذا ذهب النمر إلى بيت صوفي؟

- أ) ليرى والد صوفي
- ب) لأنه جائع
- ج) ليلعب مع صوفي

2. ماذا فعل النمر داخل البيت؟

- أ) جلس في كرسي والد صوفي
- ب) نظف الثلاجة في المطبخ
- ث) بدأ يأكل الطعام

3. لماذا ذهبت صوفي مع والديها إلى المقهى؟

- أ) لأن والدة صوفي نسيت إعداد الطعام
- ب) لأن صوفي أرادت أن تأكل الطعام في المقهى
- ج) لأن النمر قد أكل كل الطعام في المنزل

Appendix IV

This is a genuine Authorware record stored for one of the participants who was reading the story 'Platypus' in the Treatment A condition. The record shows the annotations the participant accessed along with a time stamp. The time stamp refers to the actual time of day that the participant clicked on the annotated word (i.e. in line one, the participants entered his/her name into the software are 11:30 am). The name of the participant has been changed to 'Name' for confidentiality.

Name 11:30

box 11:31

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:32

box 11:33

box 11:33

box 11:33

box 11:33

bucket 11:33

bucket 11:33

bucket 11:33

bucket 11:33

bucket 11:33

bucket 11:33

spade 11:33

lunch 11:33
lunch 11:33
bucket 11:34
bucket 11:34
spade 11:34
spade 11:34
lunch 11:34
spade 11:34
bucket 11:34
spade 11:34
lunch 11:34
lunch 11:34
bucket 11:34
lunch 11:34
spade 11:34
spade 11:34
spade 11:34
bucket 11:34
spade 11:34
lunch 11:34
lunch 11:34
bucket 11:34
box 11:35
box 11:35
spade 11:35
bucket 11:36
seaweed 11:36
seaweed 11:36
seaweed 11:36
rock 11:36
rock 11:37
seaweed 11:37
seaweed 11:37
seaweed 11:38
seaweed 11:38
seaweed 11:38

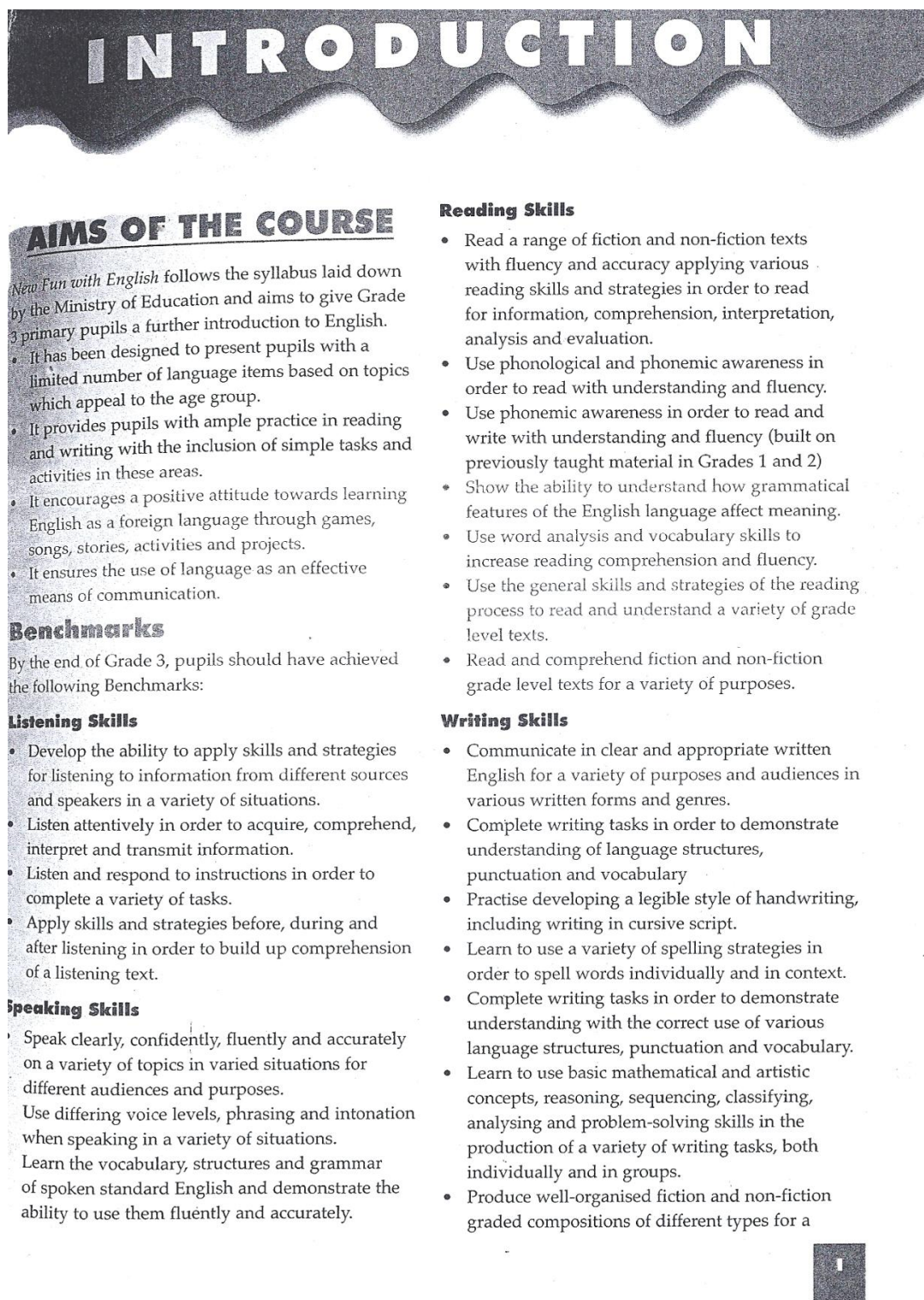
seaweed	11:38
seaweed	11:38
seaweed	11:38
rock	11:38
rock	11:38
seaweed	11:38
seaweed	11:39
seaweed	11:39
seaweed	11:39
rock	11:39
rock	11:39
rock	11:39
rock	11:39
seaweed	11:39
bucket	11:39
bucket	11:39
bucket	11:39
bucket	11:39
spade	11:39
spade	11:39
lunch	11:39
lunch	11:39
lunch	11:39
lunch	11:39
lunch	11:39
bucket	11:39
lunch	11:40
lunch	11:40
lunch	11:40
lunch	11:40
bucket	11:40
lunch	11:40
sandcastle	11:40
sandcastle	11:40
sandcastle	11:40
sandcastle	11:40

sandcastle 11:40
sandcastle 11:40
sandcastle 11:40
sandcastle 11:40
sandcastle 11:40
sea 11:41
bucket 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
shell 11:41
toe 11:41
toe 11:41
toe 11:41
toe 11:41
toe 11:41
toe 11:41
toe 11:41
shell 11:41
shell 11:41
shell 11:41
bucket 11:41
bucket 11:41
bucket 11:41
bucket 11:41
bucket 11:41
bucket 11:41
sea 11:41
sea 11:41
sea 11:41

sea 11:41
shell 11:41
shell 11:41
box 11:41
box 11:41
shell 11:42

Appendix V – Curriculum Excerpt

This is an excerpt of the aims of Grade 3. It shows the skills which participants in this study (Grade 4) had achieved in the previous school year.



(Teacher's Book, Longman's Series 'Fun with English', Grade 3, p.1)

Appendix VI Focus Groups Transcripts (1)

The following is the focus group transcript that has been translated from the participants' first language (Arabic). The participants' names have been replaced with two random letters to ensure anonymity. When learners were unidentifiable, the letter 'L' was used in place of their names. The letter 'T' denotes teacher/researcher.

[1] T: I'd like to ask you... How often do you use the computer at home?

[2] L1: A lot.

[3] T: Everyday? Twice a day?

[4] L1: we prepare topics

[5] L2: we do useful things

[6] L3: we prepare topics for school

[7] T: Yes, Ok, do you use the computer for English? Do you use it to read English?

[8] L1: Yes

[9] L2: Yes because there are stories and such we read them so later when the teacher asks about a story we will know the meanings of words.

[10]T: So you have English stories at home?

[11]L: Yes

[12]T: what did you want to say 'Ha'?

[13]Ha: Because we can gain more information in English

[14]T: Ok, you saw the stories we read in class. Have you ever seen stories with words you can click on? Or was this the first time you see a story like this?

[15]L: first time

[16]L: No

[17]L: Yes

[18]L: first time

[19]T: so for you [L] it's not the first time, so when did you see a story like this?

[20]L: we saw words which appear in English or sounds or nothing, or in Arabic and sound

[21]L: I mean they are coloured so you can click on them

[22]T: you have stories like this at home? Like the ones you used in class? Where did you get these stories?

[23]L: from school

[24]T: Aha ok

[25]L: from school and outside school.

[26]T: Ok, when you read a story in English, a regular story not one on the computer, and you come across a word in the story which you don't know, how do you go about understanding its meaning?

[27]L: we read it

[28]T: yes but if you don't know...

[29]L: We think

[30]T: yes how do you think?

[31]L: We ask our mum.

[32]T: aah ok, so you ask someone, what else do you do ?

[33]L: Yes we can.. Teacher using the computer we can..

[34]T: Yes on the computer you can but on a printed book what do you do?

[35]L: I mean if... We think about it a little.

[36]L: we get it out from the computer

[37]T: Aha ok, how from the computer?

[38]L: or from the internet, we type the name and it tells us the meaning.

[39]T: Aha ok, so you don't try to look at the words before or after or use pictures to understand the word?

[40]L: Or the dictionary

[41]T: OK

[42]L: we look up the word.

[43]T: So you don't look at the story pictures to get the meaning of a word?

[44]L: Yes, we can see

[45]T: Ok let's go to the computers so we can have a look at the stories you read because afterwards I'm going to ask you for your opinions on the stories and words. Let's go to remember Ok? Come on.

[46]T: Ok, we read three stories and each story was different. The first story you had picture and pronunciation, the second on you had

[47]L: picture and Arabic word

[48]T: that's correct Arabic. And the third story no pictures or anything.

[49]T: Which one did you find more helpful for understanding the words and story in general?

[50]L: The Arabic one

[51]T: so for you the Arabic, are all of you the same or...?

[52]L: No, for me the one without anything.

[53]T: the one without anything, why did you prefer that one?

[54]L: Because it was easy, because I know the the words in it,

[55]T: Aah ok, and why did you like the Arabic one more?

[56]L: because it helped me with this word and that word.

[57]T: so you think the meaning became easier for you?

[58]L: sort of

[59]T: What did you want to say 'Ha'?

[60]Ha: Teacher the one with sound

[61]T: the one with sound, why did you feel that was better than the Arabic or the one without?

[62]L: because it had words I couldn't read in English, so when they appeared, they were pronounced.

[63]T: aha ok.

[64]L: I liked the third story because it had words which I knew.

[65]L: Me too

[66]T: Yes, alright. The stories which had blue words, before you clicked on a blue word, a word you didn't know...Did you try to understand that word, or did you click on it straight away to get the meaning?

[67]L: No I thought.

[68]L: I thought so I know what the word is not straight away.

[69]T: you thought before you clicked?

[70]L: yes.

[71]T: Didn't you sometimes feel that you couldn't be bothered and clicked quickly to see the meaning?

[72]L: No

[73]L: No

[74]T: What did you want to say 'Ha'?

[75]Ha: Teacher like maybe I know a word before I click because I might know the words before it.

[76]T: So you clicked on words which you knew?

[77]L: and the words we didn't we clicked on

[78]T: Alright if you know a word, why did you click on it?

[79]L: to see if it's right or wrong.

[80]T: Aah ok.

[81]L: but teacher if there is a word I know and am sure of I don't click it

[82]T: you didn't click on it if you were sure?

[83]L: The one with sound there is also a picture, not just sound.

[84]T: Yes right.

[85]L: Like 'box', they put the word and under it a picture

[86]T: yes right yes, alright, did you click on a blue word more than once? Like 'box' did you click on it once, twice, three times to hear or see the Arabic?

[87]L: No

[88]L: No teacher I was memorising them for the exam time.

[89]T: Aah ok, so for every word you only clicked once or repeatedly so you could hear it more than once?

[90]L: No

[91]L: there was no sound

[92]T: No for the words which had sound

[93]L: we didn't hear sound

[94]T: the first time you used headphones

[95]L: Yes

[96]T: Alright, were there any words in blue which had sound or Arabic but you still felt their meanings were not clear?

[97]L:Yes

[98]T: Yes which ones

[99]L: Mmmm

[100] T: Let me show you..

[101] L: what do you mean I don't understand.

[102] T: For example, a word you don't know and you click on it to see the picture or the Arabic but you still don't get the meaning... were there words like this?

[103] L: No

[104] T: What 'Ha'?

[105] L: teacher there are....

[106] T: I have words here to help you remember, are there words you clicked on but still didn't understand them?

[107] L: No.

[108] T: Yes, alright, Ok. If you click on a word and still don't know its meaning what do you do?

[109] L: I tell the teacher

[110] T: OK

[111] L: I tell the teacher, she'll teach me

[112] T: No I mean from what you used, there weren't any words you clicked on and still didn't understand?

[113] L: No

- [114] L: No
- [115] T: Alright, mmm, there weren't any blue words in the tiger story, so how did you understand the words?
- [116] L: When I was young, from second year at kindergarten, I was in an English school, so I know everything.
- [117] T: you know everything.. and the rest of you girls?
- [118] L: I read stories and my mother is an English teacher so there is no need.
- [119] T: Ok but surely there were words in the tiger story which you didn't know right?
- [120] L: Yes
- [121] L: No
- [122] L: Teacher I knew everything
- [123] T: you knew all the words in the tiger story?
- [124] L: yes
- [125] T: surely there were words you didn't know.
- [126] L: No I knew them all.
- [127] T: you knew them all?
- [128] L: teacher, my mum has started teaching me year five
- [129] T: yeah alright.. 'Ha' do you want to say something?
- [130] Ha: Teacher a few words I didn't know but I knew them later
- [131] T: Yes how did you know them?
- [132] H: Teacher I tried to understand them
- [133] T: yes how did you ask someone or by yourself... how?
- [134] H: No teacher by myself.
- [135] T: yes how? What did you do to understand?
- [136] H: Teacher I looked at the picture and read the word then I knew.
- [137] T: aah ok, alright.. Now I... Before you read the story, I gave you a test.
- [138] L: yes
- [139] L: Yes
- [140] T: then you read the story
- [141] L: and then another test
- [142] T: correct.. when you read the story did you feel that the test made you look at the words..
- [143] L: correct or incorrect
- [144] T: Do you just read without being bothered about the test? Or do you feel the test affected you?

[145] L: kinda

[146] T: yes say how did it affect?

[147] L: I mean for example you give us the tests before we.. before we see the computer.. then I see because if I have mistakes the second test after the story I'll know the words.

[148] T: aha ok and the rest of you what did you do?

[149] L: I like know the word like If I what's it called get it wrong in the first test I'll know it in the second because I read the whole story and know it and understand it.

[150] T: And you [plural]... would you like to add something?

[151] L: no

[152] T: Ok, alright, when you took the second test, what made you, what helped you remember the words?

[153] Ko: the story and the computer

[154] T: Yes but for example was there a specific way that helped you?

[155] L: Yes

[156] L: Yes

[157] L: they put pictures and three words and we choose one.

[158] T: Yes that's the test, and what helped you choose?

[159] L: Teacher I paid attention while reading so I wouldn't leave anything unanswered in the test.

[160] T: Yes ok you concentrated because you knew there would be another test.

[161] T: Aha ok, alright... I'm going to show you words now, these words got the highest marks in class. I looked at the words in the first exam and then looked at the words in the second, and there are some words which nearly all of you got correct.

[162] L: (inaudible)

[163] T: Yes, I'm saying some of you, most of you didn't get them right but in the second test all of you got them right. I'm going to show you these words now, see if you can remember why you got them right.. ok? We have this word.

[164] L: "Shell"

[165] T: yes why do you think you remembered that one?

[166] L: because of the picture

[167] L: because when we read the story we clicked on "shell" and a shell appeared so we knew.

[168] T: Yes but this word also appeared.. you didn't get a high mark on this.. not all the class got this one right.. why do you think that is?

[169] L: "Spade"

[170] T: Yes why do you think? Is the picture not nice or the word is difficult?

[171] L: Teacher I say maybe some girls were not focussed on the story.

[172] L: Teacher maybe because the words were sort of similar

[173] T: These three words?

[174] L: Yes

[175] L: Like these s-s-s-

[176] L: but they added these words to them

[177] T: yes

[178] L: yes

[179] T: ok this word you also got right look

[180] L: "Sandcastle"

[181] T: Yes

[182] Sh: sandcastle, rock, shell

[183] T: yes look

[184] L: that's seaweed

[185] T: yes not many got this one right

[186] Sh: I got it right

[187] T: yes

[188] L: I got it wrong

[189] T: ok even here we have "leaf" and "cone"

[190] L: "cone"

[191] T: yes

[192] L: You didn't get anything right? This is easy

[193] T: Yes but no, this is an empty paper I'm just showing you

[194] T: careful with that chairs ha.. Ok here we have stomach ache

[195] L: "stomach"

[196] L: "stomach"

[197] L: "stomach"

[198] T: yes a lot of you got this one right.. and this one too

[199] L: that's "cake", no "chocolate cake"

[200] T: yes right, ok then.

[201] L: "sun".. no one got that right?!

[202] T: what?

[203] L: it was easy!

[204] T: No sun you knew it before, I'm talking about the words you didn't know then
you all knew. Ok then, let's take a five-minute break and then come back. I want to ask

you about reading and some other questions then we'll finish, Ok? Come on let's take a five-minute break.

- [205] L: Teacher how many lessons are you going to be with us?
- [206] T: this one and that's it.
- [207] L: Teacher take us another lesson, we have a free one.
- [208] *****
- [209] T: Ok now I'm going to ask you about reading. I just asked you about words and now the reading itself because I asked you before you read to look at the words and events in the story right?
- [210] T: OK, what I want to know is when you read the story... did you read it word for word? Or did you quickly put the cursor on the blue words?
- [211] L: No word for word
- [212] T: or some pages word for word and some pages...
- [213] L: no
- [214] L: no
- [215] L: word for word
- [216] T: you read every single word?
- [217] L: word for word, like I put the cursor to move to the next page and I read.
- [218] L: yes word for word
- [219] T: OK
- [220] L: yes me too word for word... so I can understand
- [221] T: And you 'ko'?
- [222] Ko: teacher I like I read it but if I didn't understand it.. the word I don't understand ant then like.. I read it.
- [223] T: you click on it?
- [224] Ko:yes
- [225] T: Alright...
- [226] L: and if I don't understand a word properly and it's not written on it.. not highlighted in blue.. I can ask you because you're the one who made the story, you know it like.
- [227] T: Ok, alright then mmm... you read word for word, there weren't any pages where you thought you'd just click on the words in blue?
- [228] L: No
- [229] L: No
- [230] L: no, because there might be words which I don't see and they are in blue and come in the exam so I'll be stuck then.

- [231] T: aaah so are you the same as well you were worried that words would come in the test so you read
- [232] Ko: teacher I read them later
- [233] Sh: the ones in blue come
- [234] Ko: I had to memorise the words so if they come in the test I'd know how to answer.
- [235] T: So you focused on the words in blue? Or blue and black?
- [236] My: in blue and black
- [237] T: and you?
- [238] Ko: in blue and black so I can understand the words, some of the words that come are not in blue only
- [239] T: aah Ok
- [240] H: I read word for word because you might ask me and I don't know
- [241] T: Ok alright, you feel that you read.. how did your reading differ in the story where there were no words in blue? Two stories had words in blue... in the tiger story there weren't any blue words, how did your reading differ? Did you focus more or less?
- [242] My: because they were not highlighted in blue.
- [243] Sh: In the first and second there were so they help me.. in the third story there were words which has appeared in the first and second so I remember and know them.
- [244] My: and in the test there are the same words... and some words in black come in the test so we know then.. we had focused on all the words.
- [245] T: 'ko' what did you want to say?
- [246] Ko: teacher those others helped me in the two stories but in this one now, I mean I had to read it all properly because it might come in the test and I don't know it or haven't memorised it, so I have to remember it.
- [247] T: so you focused more when there were no blue words, or was it the same in both?
- [248] My: Yes
- [249] Ko: Yes
- [250] Sh: It's the same really, the last story which...which didn't have was a little difficult but I remembered the words that were in it.
- [251] My: The hardest thing was the third story the tiger because it was without words and difficult to remember all the words because they might come in the test and we won't know it so we had to think hard and focus on it so we know it and get it right.
- [252] T: Yes
- [253] T: Do you want to say anything 'Ha'? No comment?

- [254] 'H':
- [255] Ko: Teacher I read it not only for the exam, I memorised them so if people ask me I'll remember them and know them.
- [256] Sh: because [inaudible].
- [257] T: Alright, do you think you'd like to continue with.. reading stories like this in the future?
- [258] Sh: Teacher very much so... I'd like
- [259] My: Yes I want to read stories in the future so I can become clever and the girls know me and like me
- [260] T: 'Ha' what do you think would you like to read stories like this?
- [261] Ha: Yes
- [262] Sh: I'd like to read so I know what I'm going to be when I grow up..
- [263] Ko: teacher..
- [264] Sh: Because there are stories they put a picture of a doctor or something so I see what I'm going to be.
- [265] H: Or maybe if I don't know words or not good in English I read a lot of stories so I become better in English.
- [266] Ko: Teacher I might go to many European countries and such and they speak English and I don't know English I have to know it and remember words that's why, also if I get sick and what's it called the doctor is English how would I talk to him only in English.
- [267] T: ok 'Ha' what did you want to say?
- [268] Ha: Teacher maybe when I grow up and go to a restaurant they talk in English and I can't like understand them.
- [269] T: Alright.. If I told you I'd like you to tell me what.. let's design a story together, an electronic story, I put audio and Arabic with pictures, if you had the option, what would you add? Do you have any ideas or something lacking you would like to add?
- [270] My: I would add words, add sentences, add a story like if I had a story in mind...like if a story comes to mind I'd write it..
- [271] Sh: the first thing before I put it on the computer
- [272] My: I think about it
- [273] Sh: I have to think about it a bit because maybe there's a mistake so I don't put it in the story and all the girls will see it.
- [274] T: Yes but that's not what I meant. I mean when you read a story with blue words which had picture and audio or picture and Arabic.. if you read again and want me to prepare another story or stories.. what would you like to see?

[275] H: I'd like..

[276] Ko: I'd like...

[277] My: I'd like to see pictures and new words so I know what to focus on.

[278] Sh: to see new words

[279] My: and know what not to focus on because it might be easy and I'm focussing on it but later when I remember another word I'll forget the word and remember the easy one.

[280] Sh: and I don't like it when I get a story I don't like the second to be the same meaning but in a different design, no I like to get other stories like because later in the future I know how to speak English.

[281] T: alright

[282] H: If I see a lot of pictures and a lot of words I'll become more interested in the story.

[283] T: Ok

[284] Sh: Like if I see a lot of pictures..no a lot of words after the picture how will I see the picture and what's going on.

[285] T: so you want to see more blue words?

[286] Sh: yes

[287] H: yes

[288] T: more than you already saw in the stories?

[289] Sh: Yes but.. there are stories..

[290] My: there are words which are easy like "box" and like "shell" they are easy like.

[291] Sh: but there are a lot of girls who don't get it right

[292] T: and what did you want to say 'Ko'?

[293] Ko: Teacher like when I'm at home I make a test for myself I don't put with them those words in blue and such so I know them but like if I know them like I won't click on them at all.

[294] My: I ask my mum like because my mum is an English teacher like she knows how to talk English like... like.. I ask her to make a test for me to take like and I answer it all correct.

[295] T: Yeah

[296] My: I tell her to explain the meanings of words and that

[297] T: aah Ok... alright, do you have any comments or things to say about..about the stories you read in class or anything

[298] Sh: I didn't like the first story with the platypus

[299] T: Yes why didn't you like it?

[300] Sh: because like.. I don't know I didn't like it

[301] T: Yes say why

[302] My: I didn't like it because he was taking things and putting them down making
a mess

[303] Sh: and stealing things from the ground I see things and such..

[304] My: There are things also on the ground and he takes them and without seeing
them stands on them.

[305] T: Yes

[306] Ko: Teacher I liked the caterpillar best

[307] T: alright why did you like it?

[308] Sh: Me no.. me the tiger

[309] My: Me the tiger

[310] Sh: because he eats a lot and such and I'm like him.

[311] My: I like him because he was giving orders and he was like.. first he was evil
and then they like him

[312] T: Who the tiger?

[313] Sh: Yes

[314] My: yes

[315] T: what about you what did you like or dislike?

[316] Ko: the caterpillar

[317] T: the caterpillar, any particular reason or just like that?

[318] Ko: No teacher I liked it because he kept coming out like... when he comes out
of the plume like his colour is different...when he comes out of the "strawberry" red
and green and so on

[319] T: Yes

[320] Sh: and I liked it because we took it in science

[321] T: Yes

[322] H: I like the caterpillar story

[323] T: Oh so like 'ko'?

[324] H: yes

[325] T: Yes and what did you like about it?

[326] H: because like I saw the life cycle of the butterfly and I liked it.

[327] T: Ok what.. which story didn't you like or least preferred?

[328] Sh: me the first story the duck

[329] My: I didn't like the duck

- [330] Sh: because it has a lot of things that aren't nice and easy words so I consider it childish
- [331] T: Yes yes
- [332] Sh: so I didn't like it
- [333] Ko: Teacher I like the caterpillar also because like there are nice things in it
- [334] T: Ok .. good
- [335] Sh: but the nicest one is the tiger
- [336] Ko: it was the caterpillar
- [337] Sh: the caterpillar had new information
- [338] T: Ok is there.. is there anything else you'd like to add before we leave?
- [339] Ko: Teacher the caterpillar was like a worm but how it became a butterfly
- [340] Sh: there was a lot of information
- [341] My: it was an egg
- [342] Ko: and then a caterpillar and then a cocoon
- [343] Sh: in the story.. in the story..
- [344] My: and then it grew and became a butterfly.. fully grown.. and then began to fly like
- [345] T: Yes
- [346] Sh: at the end of the story I read that there was... the tiger wanted to eat a lot and go into the kitchen... he ate all the food and the mum and dad and girl they didn't have food so what did they do.. they told the tiger to go out while they clean.. he opened the fridge and threw everything making a mess.
- [347] T: Yes
- [348] Sh: there were useful things
- [349] My: I liked the third story because the sentences were long and were like a bit difficult I felt it was for older people
- [350] T: Yes
- [351] H: I like the caterpillar story because it had lots of colours... like nice colours
- [352] T: that's right very colourful
- [353] Ko: Teacher I also liked the tiger because he was standing and throwing things and on top of the fridge as if trying to hide.
- [354] T: Yes well done, anything else you want to add? Before we leave?
- [355] My: I liked

Appendix VI Cont. Focus Groups Transcripts (2)

- [1] T: Ok boys, the first question is.. how often do you use computers at home?
- [2] Aa: ten hours
- [3] T: ten hours!!
- [4] L: about two hours
- [5] T: two hours
- [6] L: an hour and a half
- [7] L teacher three hours
- [8] T: and everyday? or...
- [9] L: Teacher me everyday
- [10]Sg: me not everyday two hours and a half
- [11]T: ok
- [12]Aa: me always Friday and Saturday ten hours
- [13]T: alright
- [14]L: an hour and a half
- [15]T: an hour and a half... alright and what do you use the computer for?
- [16]L: mm
- [17]T: go ahead say
- [18]Sg: games
- [19]T: ok
- [20]L: for school
- [21]T: ahm
- [22]L: For worksheets
- [23]T: worksheets.. aah ok
- [24]L: teacher I study
- [25]L: teacher I print out papers
- [26]T: ahm
- [27]L: from the computer to the printer
- [28]T: alright and what sort of games? Sager you said games, what other games on the computer?
- [29]L: aaaa
- [30]L: teacher I watch videos
- [31]T: aah right ok
- [32]L: I draw pictures
- [33]T: Ok... alright aah do you read in English on the computer or only in Arabic?

[34]L: me English

[35]L: me English and Arabic

[36]L: me only English

[37]L: English and Arabic

[38]L: English and Arabic

[39]T: alright, what sort of things do you read in English?

[40]L: using software and such we read in the software

[41]T: ahm

[42]L: or website

[43]T: aah the internet then

[44]Aa: Or I download an English story to read

[45]T: Ok.. alright.. emmm.. you see the stories we read in class

[46]L: yes

[47]L: yes

[48]T: have you read.. have you ever seen something similar in the computer?

[49]L: yes

[50]L: yes teacher

[51]T: yes tell me who can tell me what they've seen

[52]L: teacher like a software?

[53]T: yes in the same way, there are words, blue words you click on to get a picture which explains its meaning.

[54]L: No video with words and audio reading

[55]T: aha ok

[56]L: you don't control it just video

[57]T: aah ok so you can't click on a word

[58]L: Yes I can't

[59]L: teacher the word kitchen we knew the word kitchen in English

[60]T: from here or home?

[61]L: from here

[62]T: ok.. ha Abdullah? Do you want to say something else?

[63]Ab: no.

[64]T: alright .. mmm... alright.. I want to ask you when you read in English, not necessarily on the computer, anywhere .. if you see a word you don't know what do you do? How do you go about understanding its meaning?

[65]L: I go on google and write the word and it comes up

[66]T: come up in Arabic or English?

[67]L: anyone English and Arabic

[68]T: Yes sager what did you want to say?

[69]Sg: I ask an adult or something

[70]T: aha ok.. are there any other ways?

[71]L: I ask [inaudible]

[72]T: aha ok

[73]L: teacher like teacher I go on a website and translate

[74]T: and translate... ok.. ok don't you try to for example look at the words in the stories that come before or after it to help you understand from the story itself? Or do you always go outside the story to find the meaning like from an adult or website?

[75]L:

[76]T: do you try to look at the pictures in the story to get the meaning?

[77]L: sometimes something blue appears and I go on it and a picture appears... I don't know how to go on google and write.

[78]T: aha ok.. alright... ok.. let's now remember what we did on the computers and then come back

[79]****

[80]T: ok in the computer we had three... we said one story without blue words, and a story with blue words and English and pictures English pronunciation we used headphones and the third story...

[81]L: in Arabic and without headphones.

[82]T: in Arabic without headphones and a picture appeared.. which of these three do you feel helped you most to understand the story and the words? Mh?

[83]Mh: the third story

[84]L: the third story the tiger was.. yes with Arabic.. why did you feel the Arabic helped you more than the English?

[85]Mh: because

[86]L: I say the one in English

[87]Mh: because the words that I don't know I'll know them in Arabic... I'll know how to say them in English.

[88]T: Aha ok and the rest of you the same or do you have...

[89]L: Me in English

[90]T: you liked the English more?

[91]L: yes

[92]T: yeah why?

[93]L: because it's with headphones and sound comes out and it's better than the Arabic.

[94]T: Aha

[95]L: Teacher because he was in an English school

[96]T: aah Ok

[97]L: I sometimes struggle with Arabic

[98]T: yeah alright.. what about the rest of you ? Abdullah what do you think?

[99]Ab: Teacher the Arabic is easier because teacher the word in English I don't understand what it is.

[100] T: aha ok.. alright what about you?

[101] LLL: [silence]

[102] T: or didn't you like any of the,?

[103] L: teacher I liked them all

[104] T: you didn't feel one was better than the other?

[105] LLL: [silence]

[106] T: ha?

[107] L: I find them the same

[108] T: what do you think English or Arabic?

[109] L: Arabic

[110] T: and you Soud?

[111] S: Teacher the Arabic is the best.

[112] T: the best alright you don't think the one without words was good or not? No one mentioned that one was the best... the one without blue words.

[113] L: good good

[114] L: teacher it's difficult

[115] T: more difficult than the others?

[116] L: No easy.. I found it easy

[117] L: teacher without words or anything

[118] T: Yes Ok.

[119] L: Teacher I found it easy

[120] T: alright... now there are words in the story which you didn't know.. before you clicked on a blue word.. did you try and think what it means... or did you think it was best to click on it immediately?

[121] L: I tried

[122] L: teacher when I first saw it I knew it

[123] T: aha and the rest of you did you click straight away or thought that this word may mean this and then click?

[124] L: teacher I kinda thought about it and then clicked

- [125] T: to see whether it's right or not?
- [126] L: teacher I tried to read it I couldn't then clicked
- [127] T: aha
- [128] L: teacher if the word is long teacher like I might not know how to read it so I click on it.
- [129] T: Ok.. alright.. maybe.. surely there were words in blue.. which you knew.. easy words.. did you also click on those even when you knew their meanings?
- [130] L: if I know it I don't click on it
- [131] T: alright and the rest of you?
- [132] L: the stories which came up in blue I didn't click on them because I knew them
- [133] T: alright and you Sager are you the same or did you click on words you knew?
- [134] Sg: aah I clicked on the words I didn't know and didn't click on the ones I knew..
- [135] T: and the rest of you..
- [136] L: teacher I clicked on the ones I knew and didn't know
- [137] L: teacher teacher
- [138] T: and why did you do that?
- [139] L: teacher to find out if I'd understood it correctly or not.
- [140] T: aah to check the meaning you mean
- [141] L: teacher Sager sometimes just guesses answers in the tests.
- [142] T: never mind alrihgt.. and Soud and Saleh what did you do? Did you use the blue only for words you didn't know or also for ones you knew?
- [143] L: teacher me only the ones I don't know
- [144] L: teacher me too
- [145] L: the same
- [146] L: teacher I clicked the ones I knew and the ones I didn't know to be sure.
- [147] T: ok.. alright.. is there for example a word you clicked on more than once like the word "box" did you click it more than once "box" "box" or once and that's it?
- [148] L: teacher "box" means box
- [149] T: yes
- [150] L: if it's difficult I click on it repetitively and if not just once
- [151] T: aha ok and did you do this in both cases in Arabic and English or just the English?
- [152] L: just the English
- [153] T: so you could hear it?
- [154] L: yes
- [155] T: Ok.. what did the rest of you do?

- [156] L: Teacher but sometimes you click on it because you don't know it you click on it a lot.
- [157] T: more than once ok and what did you do? Did you click on a single word more than once?
- [158] L: teacher I click on the word I don't know I click on it so I know I know how
- [159] T: so more than once
- [160] Ab: Teacher I know it from the picture
- [161] T: from the picture ok.. ha Soud?
- [162] S: teacher I'm like Abdulla I know it from the picture
- [163] T: yes so you clicked once only?
- [164] L: yes teacher I memorise the words teacher then I answer
- [165] L: teacher teacher I don't know it from the picture.. I know like everything because I was in an English school.
- [166] T: Ok.. alright.. mmm were there words you felt.. even though a picture appears and Arabic you still felt that you didn't understand the word.
- [167] L: No
- [168] T: you found all the words clear?
- [169] L: yes
- [170] L: yes
- [171] L: yes teacher
- [172] T: so there weren't any words where you saw the Arabic or heard the English but still didn't understand?
- [173] L: no
- [174] L: no
- [175] T: ok... alright.. mmm in the platypus story.. where there were no blue words.. how did you go about understanding the meaning of the words? There were no words in blue
- [176] L: I see the pictures
- [177] L: maybe from the first test
- [178] T: alright
- [179] L: from the pictures and the words which which are inside not the blue and the picture aaa I think this word is in the picture I know this word will come in the test.
- [180] T: aah ok and the rest of you? Mh?
- [181] Mh: First thing teacher I took the test I knew the words
- [182] T: yes alright
- [183] L: teacher when I read the story I know what's in it.

- [184] T: yeah .. alright.. and over here what did you do? When there weren't any blue words no help how did you try to understand the words?
- [185] L: like I know it like I understand it I remember it for the test
- [186] L: teacher I read the story.. later when I come to write I know it.
- [187] T: yeah.. alright.. so from what you've said it seems that the first test affected the way you read.. like you so the f f first test and knew the words to look out for in the story... you feel the test affected you?
- [188] L: yes
- [189] L: yes teacher
- [190] Aa: teacher I felt the first test was easy that's why I finish quickly.
- [191] T: yes.. alright
- [192] Aa: also the second test
- [193] T: alright ok since you saw words in the first test, didn't that make you look only for those words and not focus on the whole story?
- [194] L: no
- [195] L: no
- [196] L: It made us read the whole story
- [197] T: mmm so you didn't just focus on the words because they came in the first test?
- [198] L: the words are in the first test and the story
- [199] T: yes
- [200] Sa: I I in the first lesson here.. like the first test I finished it then you said to go on the computer to read I didn't read because I know everything.
- [201] T: yes
- [202] Aa: then the second test became easier
- [203] T: alright and soud and saleh what did you do?
- [204] L: [inaudible]
- [205] L: nothing much it was easy
- [206] T: yes... ok.. alright when you took the second test what do you think helped you remember the words? For example was it the English audio or the picture or the word in Arabic .. what helped you remember in the second test? Mh
- [207] Mh: the English audio... the last story said the word
- [208] T: yes the pronunciation.
- [209] Mh: em
- [210] T: and you feel that helped you more than the picture or Arabic?
- [211] Mh: em

[212] T: and the rest of you the same or different?

[213] L: same

[214] L: same

[215] L: same thing

[216] L: same thing

[217] T: so you felt the English was better the Arabic?

[218] L: yes

[219] L: yes teacher

[220] Sg: the Arabic audio was better than the English

[221] Aa: there was no audio in Arabic

[222] T: yes there was only an Arabic word

[223] L: a word and a picture

[224] T: a word and a picture correct.. alright.. ok we'll take a five-minute break and then come back ok?

[225] ***

[226] T: Ok.. we've just.. just discussed the mmm words in the stories and how you read and what you clicked but I'd like to see how you went about reading.. knowing the events of the stories.. alright.. so when you read.. did you read the story word for word or did you for example, focus mainly on the blue words and read the rest quickly.. what did you do? Saleh?

[227] Sal: teacher I like

[228] Aah: teacher I

[229] T: saleh

[230] Sal: Teacher like the words I don't know or understand I just skip them.. I don't have to read them.

[231] T: yeah I don't mean the words the reading itself like did you read all the words in the story even the black ones? Or did you focus only on the blue? Or half and half? Mh?

[232] Mh: all the words

[233] T: all so you read word for word

[234] Mh: yes

[235] Aa: I read quickly because I'm focused on the words so quickly

[236] T: which words did you focus on?

[237] Aa: the black.. the black and blue

[238] T: Ok

- [239] L: teacher any page.. any difficult word is in blue but the ones in black are all easy?
- [240] T: aah so you focused on which one
- [241] L: the black
- [242] T: the difficult?
- [243] L: no the easy
- [244] T: the easy aah ok... and what did you do Sager?
- [245] Sg: If I don't know a word and its meaning aah the picture.. I look at the picture and then know the word
- [246] T: alright and what did you do about the words in black? Did you read them all or skip them or half and half?
- [247] Sg: half and half
- [248] T: ok.. alright.. alright do you think you focused more or less on the first story when when there weren't blue words... when did you focus more when there were or weren't blue words? Fawaz
- [249] F: when there are no words [annotated words]
- [250] T: You felt you focused more?
- [251] F: Yes
- [252] T: Why?
- [253] F: Because when there are blue words you stop there and then forget
- [254] T: How do you mean?
- [255] F: I mean if someone stops at a word next to the blue he might forget his place
- [256] L: Teacher he means..
- [257] Sg: there are too many blue words
- [258] L: He is reading for a while then stops, by the time he read, by the time he understands the word he forgets what he is reading
- [259] T: Is that what you mean?
- [260] Aa: when there are many blue words next to each other and someone stops at a blue place like later when he returns he forgets his place
- [261] T: aha ok so you forget your place on the page
- [262] Sg: teacher he means there are too many blues
- [263] T: aah .. alright.. what about you.. soud?
- [264] Sl: teacher it was better teacher when there were blue
- [265] T: yes why do you feel it's better?
- [266] S: teacher we understand
- [267] T: Ok

[268] L: teacher I focused more when there weren't blue because it doesn't show that this word is there and that word is there

[269] T: Yes

[270] L: isn't there

[271] T: so you had to read every word then

[272] Aa: teacher.. it's better without blue so you don't just guess in the test you only write what you know.

[273] T: aah..ok.. alright.. do you have any other comments about reading and how you read the stories?

[274] L: no

[275] LLLLL:

[276] T: Ok... alright do you feel you'd like to continue reading other stories similar to those you read in class?

[277] L: yes teacher

[278] L: yes

[279] L: yes

[280] L:yes

[281] T: you'd like to do that.. alright what else would you like to see in stories I mean the stories I gave you had pictures and Arabic had English audio what else would you like to see?

[282] Aa: teacher we don't want pictures .. like without pictures

[283] T: so what do you want then?

[284] Aa: I mean just the words in Arabic or English

[285] T: you mean a blue word?

[286] Aa: yes

[287] T: and what appears?

[288] Aah: the picture

[289] T: the picture only?

[290] Aa: yes

[291] Aa: but without a big picture showing what's going on.. just us reading

[292] T: aaa you mean you don't want the story pictures to be there?

[293] Aa: yes

[294] T: why would you prefer that?

[295] Aa: because...because when you read you'll know better than from the picture

[296] T: aaah you mean you feel that you are depending on the pictures to understand meaning? Is that why you'd prefer not to have pictures?

[297] Aa: yes

[298] T: and the rest of you what do you feel? What would you like to .. see more of
Mh?

[299] Mh: teacher like him because because if the picture is there I'll know the words
because everything's in the picture

[300] T: alright you mean the story's big picture or the small ones?

[301] Mh: no the big one

[302] Aa: teacher if there're pictures like you can know the story quicker if there
aren't you have to read to know.

[303] T: Ok

[304] L: teacher if there are pictures it helps you focus on what you're reading

[305] T: yes

[306] Sg: you need pictures for the difficult words and the easy ones you don't

[307] T: alright and what do you think of the big picture in the story? I mean the
picture of the platypus when he goes to the see.

[308] Sg: that's fine

[309] T: so you don't want it removed?

[310] Sg: no it's fine

[311] T: ok... anything else?

[312] L: no

[313] T: you wouldn't like to see anything other than the Arabic and English audio?

[314] L: teacher I want to see something else

[315] T: mm?

[316] L: we'd like to see something else.

[317] T: like what? Do you have ideas?

[318] L: teacher I'd like a zoo or something like that

[319] Aa: why don't you do....

[320] L: or at the see

[321] T: aah so other topics?

[322] L: yes teacher

[323] L: teacher we want about the see... about..

[324] L: teacher about whales

[325] T: about whatles ok... so other topics and what else? Any other features you'd
like?

[326] Aa: maybe do a story about a lost girl

[327] T: ok

[328] L: teacher I want to change the accent and make it Japanese ... haha

[329] T: you want to study Japanese ha? Hahaha

[330] Aa: teacher maybe you can do a story like.. like a poor boy.. looking for a house
and then someone rich gives him money and buys a house and I don't know

[331] T: ok... nice ideas.. alright.. was there anything in the software you didn't like?

[332] L: no

[333] L: no

[334] L: no

[335] T: nothing nothing?

[336] L: nothing

[337] L: nothing.. nice

[338] T: you felt the software was nice

[339] L: yes teacher

[340] T: was there anything you felt that... if this were removed or that added.. I don't
like this

[341] Aa: if the pictures were removed.. the big pictures

[342] T: yes you mentioned the big pictures yes.. but not all of you say this right?

[343] L: yes teacher

[344] Sg: if there were fewer words

[345] T: the blue ones? Or the story itself?

[346] Aa: no

[347] Sg: the story itself

[348] L?: where's Ms [teacher's name]?

[349] T: alright.. which words?

[350] Sg: the words of the whole story

[351] T: aa ok so a shorter story

[352] Sg: yes

[353] Aa: either remove the pictures or put difficult words so we can read and then
understand

[354] T: ahmm

[355] L: Teacher whenever we read it says the word we read

[356] L: yes teacher whenever we read.. teacher like the last story the word kitchen
was repeated a lot.

[357] T: yeah do you think that's a good thing or not?

[358] L: teacher a good thing

[359] T: and what do you think? Ha Abdullah you said you think it isn't good?

[360] L: Yes why is it repeated so much?

[361] T: mm...ok... do you have any other comments things to add about the software
because this is our last day now we read three stories

[362] Aa: teacher

[363] T: hm?

[364] Aa: why don't you do a story about a boy who buys a play station and plays...
and is happy

[365] T: yes ok we said other topics.. nice.. anything else you'd like to ass?

[366] LLLLLL:.....

[367] T: do you now feel you'd like to read more stories?

[368] L: yes

[369] L: yes

[370] L: yes teacher

[371] T: or no change

[372] L: no teacher we want more stories

[373] T: mmm.. alright

[374] Aah: we want without pictures

[375] T: ok

[376] L: no why are you telling her without pictures?

[377] T: or maybe we have software we you can choose before you read whether you
want pictures or not

[378] Aa: teacher to download a software that has lots of stories

[379] T: yes

[380] L: teacher like teacher you leave only the blue words for the test

[381] Aa: or download stories like on your iphone and then connect a wire to the
iphone and the recharge wire to the computer to download the stories

[382] T: yes so the story is from the phone to the computer

[383] Aa: yes but not stories with pictures.. like ones for grownups

[384] T: yes

[385] SagL there's an app aaa in the iphone.. aaa a man.. aa aa he talks aah you say
for example spider man

[386] Aa: it repeats after you

[387] Sg: yes that

[388] T: aah it copies you then

[389] Aa: it's a cat

[390] T: alright how can we out the in... benefit from it to learn English?

[391] Aa: teacher

[392] T: how do you think we can make use of it?

[393] L: I read or talk to it

[394] T: ok so it talks English

[395] Aa: I mean when you want to memorise something you repeat it

[396] Sg: or they ask ask you and you say the meaning of the word

[397] T: verbally or in writing?

[398] Sg: verbally or in writing

[399] T: ok

[400] L: teacher a blue word teacher teacher I memorise it so when I come to answer
I write it

[401] T: alright if I told you that's it there'll be no pre-test and post-test.. no first test
or second test.. would you focus more on the story or not? Be honest

[402] Aa: teacher give us like tests

[403] L: teacher no we won't focus

[404] T: Mh do you think you'd focus?

[405] Mh: yes teacher I would.

[406] T: you don't think why should I focus if there isn't a test?

[407] Aa: teacher teacher give us the same test as the caterpillar... that story

[408] T: alright.. that's it ok anything else before we finish?

[409] L: no

[410] T: okaaay that's it thanks a lot

[411] L: teacher we want to hear our voices.

References

- Abraham, L. (2007). Second-language Reading Comprehension and Vocabulary Learning with Multimedia. *Hispania*, 90, (1), pp. 98-108.
- AbuSeileek, A. (2008). Hypermedia Annotation Presentation: Learners' Preferences and Effect on EFL Reading Comprehension and Vocabulary Acquisition. *CALICO Journal*, 25 (2), pp. 260-275.
- AbuSeileek, A. (2011). Hypermedia Annotation Presentation: The Effect of Location and Type on the EFL Learners' Achievement in Reading Comprehension and Vocabulary Acquisition. *Computers & Education*, 57, pp. 1281-1291.
- Acha, J. (2009). The Effectiveness of Multimedia Programmes in Children's Vocabulary Learning. *British Journal of Educational Technology*, 40 (1), pp. 23-31.
- Ackermann Edith; David Gauntlett and Cecilia Weckstrom (2009). Defining Systematic Creativity, LEGO Learning Institute [Retrieved from <http://www.legofoundation.com/en-us/research-and-learning/foundation-research/defining-systematic-creativity>].
- Akbulut, Y. (2007). Effects of Multimedia Annotations on Incidental Vocabulary Learning and Reading Comprehension of Advanced Learners of English as A Foreign Language. *Instructional Science*, 35 (6), pp. 499-517.
- Alemi, M., Lavasani, M., and Daftarifard, P. (Constructivism, Connectionism, and Emergentism in SLA:Communalities and Differences [Retrieved from http://www.academia.edu/10811486/Constructivism_Connectionism_and_Emergentism_in_SLA_Communalities_and_Differences]
- Alesandrini, K. & Larson, L. (2002). Teachers bridge to constructivism. *The Clearing House*, 119-121.

- Alessi, S. and Dwyer, A. (2008). Vocabulary Assistance Before and During Reading. *Reading in a Foreign Language*, 20 (2), pp. 246–263
- Al-Seghayer, K. (2001). The Effects of Multimedia Annotation Modes on L2 Vocabulary Acquisition: a Comparative Study. *Language Learning and Technology*, 5 (1), pp. 202-232.
- Antonenko, P. and Niederhauser, D. (2010). The Influence of Leads of Cognitive Load and Learning in Hypertext Environment. *Computers in Human Behaviour*, 26, pp. 140-150.
- Anuradha, K.T. and Usha, H.S. (2006). Use of E-Books in An Academic and Research Environment. *Program*, 40 (1), p.48.
- Ariew, R. and Ercetin, G. (2004). Exploring the Potential of Hypermedia Annotations for Second Language Reading. *Computer Assisted Language Learning*, 17 (2), pp. 237-259.
- Authorware* (2005) (Version 7) [Computer program]. Adobe Systems Inc.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. (1992). Working Memory. *Science*, 255, pp. 556-559.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, 8, pp. 47–89. New York: Academic Press.
- Ballester, C., & Rodríguez, J. C. (2010). Developing Spanish Online Readings Using Design-Based Research. *CALICO Journal*, 27(3), pp. 540-553.
- Barbour, R., and Kitzinger, J. (1999). Introduction: The Challenge and Promise of Focus Groups, in Barbour, R., and Kitzinger, J. (eds) *Developing Focus Group Research: Politics, Theory and Practice*. 1-20. London: Sage Publications.
- Beck, I.L., & Mckeown, M.G. (2001). Text Talk: Capturing the Benefits of Read-Aloud Experiences for Young Children. *The Reading Teacher*, 55, pp. 10–20.

- Bensoussan, M. (1992). Learners' Spontaneous Translations in an L2 Reading Comprehension Task; Vocabulary knowledge and Use of schemata' in Arnaud, P. and Bejoint, H. (eds) *Vocabulary and Applied Linguistics*. London; Macmillan Academic and Professional Ltd. (pp. 102-112).
- Biemiller, A., and Boote, C. (2006). An Effective Method for Building Meaning Vocabulary in Primary Grades. *Journal of Educational Psychology*, 98(1), pp. 44–62.
- Blewitt, P., Rump, K., Shealy, S., and Cook, S. (2009, May 1). Shared Book Reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, 101(2), pp. 294–304.
- Brabham, E.G., and Lynch-Brown, C. (2002). Effects of Teachers' Reading-aloud Styles on Vocabulary Acquisition and Comprehension of Students in The Early Elementary Grades. *Journal of Educational Psychology*, 94(3), pp. 465-473.
- The British Council <http://www.britishcouncil.com.kw/en/programmes/education/kids-read> [Last accessed: 27 February 2015]
- Bryman, A. (2012). *Social Research Methods*. Oxford : Oxford University Press.
- Cameron, L. (2001). *Teaching languages to young learners*. Cambridge: Cambridge University Press.
- Candy, P. C. (1991). *Self-direction for lifelong learning* . San Francisco: Jossey-Bass.
- Chang, C. and Yang, F. (2011). Exploring The Cognitive Loads of High-School Students as They Learn Concepts in Web-Based Environments. *Computers & Education*, 55, pp .673-680.
- Chapelle, C. (2001). *Computer Applications in Second Language Acquisition*. Cambridge; Cambridge University Press.

- Chapelle, C. (2003). *English Language Learning and Technology: Lectures of Applied Linguistics in the Age of Information and Communication Technology*. Amsterdam: John Benjamins Pub.
- Chapelle, C. (2009). The Relationship between Second Language Acquisition Theory and Computer-Assisted Language Learning. *The Modern Language Journal*, 93, pp. 741-753.
- Chen, Z. (2006). *The Effects of Multimedia Annotations of L2 Vocabulary Immediate Recall and Reading Comprehension: A Comparative Study of Text-Picture and Audio-Picture Annotations under Incidental and Intentional Learning Conditions*. PhD. Thesis. University of South Florida.
- Chen, I. and Yen, J. (2013). Hypertext Annotation: Effects of Presentation Formats and Learner Proficiency of Reading Comprehension and Vocabulary Learning Foreign Languages. *Computers and Education*, 63 pp. 416-423.
- Cheng, Y. and Good, R. (2009). L1 Glosses: Effects of EFL Learners' Reading Comprehension and Vocabulary Retention. *Reading in a Foreign Language*, 21 (2), pp. 119-142.
- Chu, M-L. (1995). Reader Response to Interactive Computer Books: Examining literary responses in non-traditional reading setting. *Reading Research and Instruction*, 34, pp. 352-366.
- Chun, D. and Plass, J. (1996). Facilitating Reading Comprehension with Multimedia. *System*, 24 (4), pp. 503-519.
- Clahsen, H. & C. Felser (2006). Grammatical Processing in Language Learners. *Applied Psycholinguistics*, 27, pp. 3-42.
- Cohen, J (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.
- Cohen, J (1992). "A Power Primer". *Psychological Bulletin*, 112 (1), pp. 155-159.

- Cohen, L., Manion, L. and Morrison, K. (2011). *Research Methods in Education*. 7th ed. New York: Routledge.
- Coiro, J., & Dobler, E. (2007). Exploring The Online Reading Comprehension Strategies Used By Sixth-Grade Skilled Readers To Search for and Locate Information of The Internet. *Reading Research Quarterly*, 42 (2), pp. 214–257.
- Collins, M.F. (2010) ELL Preschoolers’ English Vocabulary Acquisition from Storybook Reading. *Early Childhood Research Quarterly*, 25, pp. 84-97.
- Cook, G. (1997). Key Concepts in ELT: Schemas. *ELT Journal*, 51(1), pp. 86.
- de Jong, M. T., and Bus, A. G. (2004). The Efficacy of Electronic Books in Fostering Kindergarten Children’s Emergent Story Understanding. *Reading Research Quarterly*, 39(4), pp. 378–393.
- Destefano, D. and LeFevre, J. (2007). Cognitive Load in Hypertext Reading: A Review. *Computers in Human Behaviour*, 23, pp. 1616-1641.
- Doughty, D. and Williams, J. (1998). *Focus of Form in Classroom Second Language Acquisition*. Cambridge; Cambridge University Press.
- Dziemianko, A. (2010). Paper or Electronic? The Role of Dictionary Form in Language Reception, Production, and the Retention of Meaning and Collocation. *International Journal of Lexicography*, 23 (3), pp. 257-273.
- Ercetin, G. (2003). Exploring ESL Learners’ Use of Hypermedia Reading Glosses. *CALICO Journal*, 20 (2), pp. 261–283.
- Ercetin, G. (2010). Effects of Topic Interest and Prior Knowledge of Text Recall and Annotation Use in Reading a Hypermedia Text in The L2. *ReCALL*, 22 (2), pp. 228-246.
- Ellis, N. (1995). ‘The Psychology of Foreign Language Vocabulary Acquisition: Implications for CALL’ *Computer Assisted Language learning*, 8, (2-3), pp. 103-128.

- Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
- Elman, J. L. (2001). Connectionism and language acquisition. In M. Tomasello & E. Bates (Eds.), *Language Development*. Oxford: Blackwell.
- Field, A. (2009). *Discovering Statistics Using SPSS*. London: SAGE Publications Ltd.
- Fischer, R. (2007). How do we Know What Students are Actually Doing? Monitoring Students' Behavior in CALL. *Computer Assisted Language Learning*, 20 (5), pp.209-442.
- Fry, A. and Hale, S. (1996). Processing Speed, Working Memory, and Fluid Intelligence: Evidence for A Developmental Cascade. *Psychological Science*, 7, pp.237-241.
- Gass, S. G., & Selinker, L. (2008). *Second language acquisition: An introductory course*. New Jersey: Rutledge.
- Gass, S. G., Torres, M. J. A. (2005). Attention when? An investigation of ordering effect of input and interaction. *SSLA*, 27, 1-31.
- Gasser, M. (1990) Connectionism and Universals of Second Language Acquisition. *SSLA*, 12, pp. 179-199.
- Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., and Eysink, T. (2009). Learning with Hypermedia: The Influence of Representational Formats and Different Levels of Learner Control of Performance and Learning Behavior. *Computers in Human Behavior*, 25, pp. 360-370.
- Gettys, S., Imhof, L., and Kautz, J. (2001). Computer-Assisted Reading: The Effect of Glossing –Format on Comprehension and Vocabulary Retention. *Foreign Language Annals*, 34, (2), pp. 91-99.

- Greenlee-Moore, M., and Smith, L. (1996). Interactive Computer Software: The effects on young children's reading achievement. *Reading Psychology: An International Quarterly*, 17(1), 43-64.
- Grimshaw, S., Dungworth, N., McKnight, C., and Morris, A. (2007). Electronic Books: Children's Reading and Comprehension. *British Journal of Educational Technology*, 38, pp.583-599.
- Gyselinck, V., and Tardieu, H. (1999). The Role of Illustrations in Text Comprehension: What, when, for whom, and why? In van Oostendorp, H., and Goldman, S. R. (eds.), *The Construction of Mental Operations During Reading*, Erlbaum, Mahwah, NJ.
- Henson, R. N. A., Burgess, N. and Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia*, 38, pp. 426-440
- Hollender, N., Hofmann, C., Deneke, M., and Schmitz, B. (2010). Review: Integrating Cognitive Load Theory and Concepts of Human-Computer Interaction. *Computers in Human Behavior*, 26, pp. 1278-1288.
- Huang, H. (2013). E-Reading and E-Discussion: EFL Learners' Perception of An E-Book Reading Program. *Computer Assisted Language Learning*, 26 (3), pp. 258-281.
- Jacobs, G., Dufon, P., and Hong, F. (1994). L1 and L2 Vocabulary Glosses in L2 Reading Passages: Their Effectiveness for Increasing Comprehension and Vocabulary Knowledge. *Journal of Research in Reading*, 17 (1), pp. 19-28.
- James, R. (1999). Navigating CD-ROMs: An Exploration of Children Reading Interactive Narratives. *Children's Literature in Education*, 30, pp. 47-63.
- Jordan, G. (Ed) (2004). Theory construction in second language acquisition. Amsterdam: John Benjamins Publishing Company.

- Juffs, A., and Harrington, M. (2011). Aspects of Working Memory in L2 Learning. *Language Teaching*, 44 (2), pp.137-166.
- Just, M. A., & Carpenter, P. A. (1992). A Capacity Theory of Comprehension: Individual differences in working memory. *Psychological Review*, 98, pp. 122–149.
- Kennedy, C., Kools, S., Krueger, R. (2001). Methodological Considerations in Children’s Focus Groups. *Nursing Research*, 50 (3), pp. 184-187.
- Kirschner, P. (2002). Cognitive Load Theory: Implications of Cognitive Load Theory of the Design of Learning. *Learning and Instruction*, 12, pp. 1-10.
- Knight, S. (1994). Dictionary Use While Reading: The Effects of Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities. *Modern Language Journal*, 78 (3), pp. 285–299.
- Krauss, S. E. (2005). Research Paradigms and Meaning Making: A Primer. *The Qualitative Report*, 10 (4), pp. 758-770
- Lam, P., Lam, S.L., Lam, J., and McNaught, C. (2009). Usability and Usefulness of eBooks on PPCs: How Students’ Opinions Vary Over Time. *Australasian Journal of Educational Technology*, 25 (1), pp.30-44.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say? *Organizational Research Methods*, 9(2), pp. 202-220.
- Laufer, B. And Hulstijn, J. H. (2001) Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning* , 51, pp. 539-558.
- Lee, M. J., & Tedder, M. C. (2003). The Effects of Three Different Computer Texts of Readers’ Recall: Based of Working Memory Capacity. *Computers in Human Behavior*, 19, pp. 767–783.

- Leeser, M. J. (2007). Learner-Based Factors in L2 Reading Comprehension and Processing Grammatical form: Topic Familiarity and Working Memory. *Language Learning*, 57, pp. 229–270.
- Lenders, O. (2008). Electronic Glossing- Is It Worth The Effort? *Computer Assisted Language Learning*, 25 (5), pp. 457-481.
- Loftus, G. R., and Masson, M. E. J. (1994). Using Confidence Intervals in Within-Subject Designs. *Psychonomic Bulletin and Review*, 1, pp.476-490.
- Long, M.H. (1996). The Role of the Linguistic Environment in Second Language Acquisition. In Ritchie, W.C., and Bahtia, T.K. (eds.), *Handbook of second language acquisition* (pp. 413-68). New York: Academic Press.
- Mackey, A. And Gass, S.M. (2005) *Second Language Research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.
- Madrid, R., Van Oostendorp, H., and Melguizo, M. (2009). The Effects of The Number of Links and Navigation Support of Cognitive Load and Learning with Hypertext: The Mediating Role of Reading Order. *Computers in Human Behavior*, 25, pp. 66-75.
- Marcus, G. (1998) Can Connectionism Save Constructivism? *Cognition*, 66, pp. 153-182.
- Marzban, A. (2011). Investigating the Role of Multimedia Annotations in EFL Reading Comprehension. *Procedia - Social and Behavioral Sciences*, 28, pp.72 – 77.
- Mayer, R.E. (1997). Multimedia Learning: Are We Asking The Right Questions? *Educational Psychologist*, 32, pp.10-19.
- Mayer, R.E., Dow, G., & Mayer, S. (2003). Multimedia Learning in An Interactive Self-Explaining Environment: What Works in The Design of Agent-based Microworlds? *Journal of Educational Psychology*, 95, pp. 806-813.

- Mayer, R. E., Heiser, J., and Lonn, S. (2001). Cognitive Constraints on Multimedia Learning: When Presenting More Material Results in Less Understanding. *Journal of Educational Psychology*, 93 (1), pp. 187-198.
- Maynard, S. and McKnight, C. (2001) Children's comprehension of electronic books: an empirical study", *The New Review of Children's Literature and Librarianship*, 7, pp. 29-53.
- McClelland, J. L., Rumelhart, D.E., and Hinton, G.E. (1986) The appeal of parallel distributed processing. In Rumelhart, D.E., and McClelland, J. L. (Ed), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, pp. 3-42.
- McKenzie, D. (2009). E-books and 21st century learning. *Multimedia & Internet @Schools*, 16 (1), pp.27.
- Mertens, D. M. (1998). *Research Methods in Education and Psychology: Integrating Diversity with Quantitative and Qualitative Approaches*. London: Sage Publications.
- Miller, G.A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits of our Capacity for Processing Information. *Psychological Research*, 63, pp.81-97.
- Moos, D. and Marroquin, E. (2010). Multimedia, Hypermedia, and Hypertext: Motivation Considered and Reconsidered. *Computers in Human Behaviour*, 26, pp. 265-276.
- Moreno, R., & Mayer, R. E. (1999). Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity. *Journal of Educational Psychology*, 91, pp. 358-368.
- Myong, H. (2005). Glosses, Comprehension, and Strategy Use. *Reading in a Foreign Language*, 17 (2), pp. 125-143.

- Nagy, W. And Scott, J. (2000). Vocabulary Processing. In Kamil, M., Mosenthal, P., Pearson, P.D., and Barr, R. (Eds.), *Handbook of Reading Research*, pp.269-284. Mahwah, NJ: Erlbaum.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Boston, Mass: Heinle & Heinle Publishers.
- Nation, I. S. P. (2001). *Learning Vocabulary in another Language*. Cambridge: Cambridge University Press.
- Nation, P. (2005). Teaching Vocabulary. *The Asian EFL Journal*, 7 (3), pp. 47-54.
- Naumann, J., Richter, T., Christmann, U. and Groeben, N. (2008). Working Memory Capacity and Reading Skill Moderate The Effectiveness of Strategy Training in Learning From Hypertext. *Learning and Individual Differences*, 18, pp. 197–213
- Nunan, D. (1999) *Second Language Teaching and Learning*. Boston, MA, Heinle & Heinle.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. Sydney: McGraw-Hill.
- Oakhill, J. (1993). Children’s Difficulties in Reading Comprehension. *Educational Psychology Review*, 5, pp.1-15.
- Paivio, A. (1986). *Mental Representations*. New York: Oxford University Press.
- Paivio (2007). *Mind and Its Evolution : A Dual Coding Theoretical Approach*. London: Lawrence Erlbaum Associates
- Parel,R. (2004). The Impact of Lexical Inferencing Strategies on Second Language Reading Proficiency. *Reading and Writing: An Interdisciplinary Journal*, pp. 847–873.
- Paolucci, R. (1998). The Effects of Cognitive Style and Knowledge Structure of Performance Using a Hypermedia Learning System. *Journal of Educational Multimedia and Hypermedia*, 7 (2/3), pp. 123-150.
- Paribakht, T. and Wesche, M. (1997). Vocabulary Enhancement Activities and Reading for Meaning in Second Language Vocabulary Acquisition. In Coady, J. and Huckin,

- T. (eds) *Second Language Vocabulary Acquisition*. Cambridge; Cambridge University Press. (174-200)
- Peters, E. Hulstijn, J. Sercu, L. and Lutjeharms, M. (2009). Learning L2 German Vocabulary Through Reading: The Effect of Three Enhancement Techniques Compared. *Language Learning*, 59 (1), pp. 113-151.
- Pike M, Barnes M, Barron RW. (2010) The Role of Illustrations in Children's Inferential Comprehension. *Journal of Experimental Child Psychology*, 105, pp. 243–255.
- Pollock, E., Chandler, P., and Sweller, J. (2002). Assimilating complex information. *Learning Instruction*, 12 (1), pp. 61-86.
- Renninger, K. A., Hidi, S., and Krapp, A. (eds.) (1992). *The Role of Interest in Learning and Development*. Hillsdale, NJ: Erlbaum, pp. 361-376.
- Rezaee, A. and Shoar, N. (2011). Investigating the Effect of Using Multiple Sensory Modes of Glossing Vocubular Items in a Reading Text with Multimedia Annotations. *English Language Teaching*, 4 (2), pp. 25-34.
- Ricci, C.M., and Beal, C.R. (2002). The Effect of Interactive Me- Dia of Children's Story Memory. *Journal of Educational Psychology*, 94, pp. 138–144.
- Rose, M., Rose, G. and Blodgett, J. (2009). The Effects of Interface Design and Age on Children's Information Processing of Web Sites. *Psychology & Marketing*, 26 (1), pp. 1–21.
- Sabelli, N. (2008). Constructionism: A New Opportunity for Elementary Science Education. DRL Division of Research on Learning in Formal and Informal Settings, pp.193-206. [Retrieved from <http://nsf.gov/awardsearch/showAward.do?AwardNumber=8751190>].
- Sadoski, M. (2005). A Dual Coding View of Vocabulary Learning. *Reading & Writing Quarterly*, 21(3), 221-238

- Sakar, and Ercetin, G. (2005). Effectiveness of hypermedia annotations for foreign language reading. *Journal of Computer Assisted Learning*, 21, pp. 28–38
- Salthouse, T. A. (1994). The Aging of Working Memory. *Neuropsychology*, 8 (4), pp.535-543.
- Samuels, S. J. (1970). Effects of Pictures on Learning to Read, Comprehension and Attitudes. *Review of Educational Research*, 40(3), pp. 397-407.
- Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11 (2), pp. 129-158.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Seidenberg and MacDonald, 1999 Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569-588.
- Shalmani, H. B., & Sabet, M. K. (2010). Pictorial, textual, and picto-textual glosses in E-reading: A comparative study. *English Language Teaching*, 3(4), pp.195.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford : Oxford University Press, 1998.
- Smeets, D., and Bus, A. (2012) Interactive Electronic Storybooks for Kindergartners to Promote Vocabulary Growth. *Journal of Experimental Child Psychology*, 112 (1), pp. 36-55.
- Smidt, E. and Hegelheimer, V. (2004). Effects of Online Academic Lectures on ESL Listening Comprehension, Incidental Vocabulary Acquisition, and Strategy Use. *Computer Assisted Language Learning*, 17(5), pp. 525–564.
- Swanson, H.L. and Berninger, V. (1995). The Role of Working Memory in Skilled and Less Skilled Readers' Comprehension. *Intelligence*, 21, pp. 83-108.

- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects of Learning. *Cognitive Science*, 12, pp. 257-285.
- Tabatabaei and Shams (2011). The Effect of Multimedia Glosses on Online Computerized L2 Text Comprehension and Vocabulary Learning of Iranian EFL Learners. *Journal of Language Teaching and Research*, Vol. 2, No. 3, pp. 714-725.
- Taylor, A. (2009). CALL-Based Versus Paper-Based Glosses: Is There a Difference in Reading Comprehension? *CALICO Journal*, 27(1), pp. 147-160.
- Thornbury, S. (2002) *How to Teach Vocabulary*. Harlow: Pearson.
- Trochim, W. M. (2006). *The Research Methods Knowledge Base*. Retrieved from <http://www.socialresearchmethods.net/kb/> [Last Accessed: 27 February 2015].
- Turk, E. and Ercetin, G. (2014). Effects of Interactive Versus Simultaneous Display of Multimedia Glosses of L2 Reading Comprehension and Incidental Vocabulary Learning. *Computer Assisted Language Learning*, 27 (1) pp.1-25.
- VanPatten B. (ed.) (2004). *Processing Instruction: Theory, Research, Commentary*. Mahwah, NJ: Lawrence Erlbaum.
- Walter, C. (2004). Transfer of Reading Comprehension Skills To L2 Is Linked To Mental Representations of Text and To L2 Working Memory. *Applied Linguistics*, 25 (3), pp. 315-339.
- Weir, C. (2005). *Language Testing and Validation*. New York: Palgrave MacMillan.
- Yanguas, I. (2009). Multimedia Glosses and Their Effect of L2 Text Comprehension and Vocabulary Learning . *Language Learning & Technology*, 13 (2), pp. 48-67.
- Yoshi, M. and Flaitz, J. (2002). Second Language Incidental Vocabulary Retention: The Effect of Text and Picture Annotation Types. *CALICO Journal*, 20 (1), pp. 33-58.
- Zhu, E. (1999). Hypermedia Interface Design: The Effects of Number of Links and Granularity Nodes. *Journal of Educational Multimedia and Hypermedia*, 8 (3), pp. 331-358.

