Effective Vocabulary Learning in Multimedia CALL Environments: Psychological Evidence

Saad Ahmed M Alzahrani

Doctor of Philosophy

University of York

Education

February 2018

Abstract

A wide range of technologies are now applied in the field of second language (L2) vocabulary acquisition. Nevertheless, intentional language-focused vocabulary CALL software has not been proven to effectively operationalise working memory. The research presented in this thesis contributes to the existing literature by identifying coding features from cutting-edge multimedia technologies that relate to L2 learning and memory research. The study participants were fifty undergraduate students from the University of York, UK. Their individual differences and memory abilities were assessed using the Automated Working Memory Assessment (AWMA). Initially, the participants were exposed to L2 novel words via the Computer-Assisted Vocabulary Acquisition software (CAVA) via three interactive interfaces: a verbal-based menu driven interface (L2-L1: MDI), a visual-based graphical user interface (L2-Picture: GUI) and a visuospatial-based zoomable user interface (L2-Context: ZUI), and immediate and delayed post-tests conducted. The first study results revealed that ZUI correlated significantly with AWMA, tending to be the most effective multimedia learning method in the immediate post-test, compared with GUI and MDI. However, in the delayed post-test, ZUI's effect experienced a dramatic decline, while GUI tended to be the most effective. In the second study, the participants were exposed to a second version of CAVA. Their accuracy and response times during the translation recognition task were measured and analysed, as were their pupillary responses. The findings revealed the participants were significantly more accurate and faster when judging the *No* translation pairs than the *Yes* ones. Of the multimedia representations, responses to MDI words were achieved significantly faster and more accurately than to GUI and ZUI words. Moreover, those participants with high verbal short-term memories were significantly faster and more accurate, experiencing a relatively reduced pupil size.

Contents

Acknowledgements

I owe a profound debt of gratitude to God; without Him, nothing would have been possible. I wish also to thank the many people who have contributed in so many ways to this research. I would like to extend my gratitude to my supervisor, Professor Leah Roberts, for her thought provoking guidance, which enabled me to deepen my research in a unique way, and for her continuing support, encouragement and assistance throughout recent years. I am also greatly indebted to Dr Zöe Handley for her rigorous thesis advisory meetings, her advice on the design of the CAVA software, and for providing me with the details of her own CALL work. I also wish to thank Dr Norbert Vanek, for his assistance and advice on the second study included in this thesis. I am very grateful to Dr Alice Cruickshank for her guidance and help with the pupillometry study, and for allowing me to use the eye tracking lab. I would of course like to thank all the participants for their cheerful cooperation and their patience. I am especially indebted to my parents, for their endless guidance throughout my life. Finally, I owe enormous thanks to my wife, for her continuous support for me.

Author’s Declaration

I hereby declare that, except where explicit reference is made, the work contained in this thesis is entirely my own, and it is the outcome of the studies conducted since the official start date of my PhD degree. This work has not, in whole or in part, previously been submitted for an award at York, or any other, University.

CHAPTER ONE: INTRODUCTION

1.1 Overview

This research falls within the domain of applied linguistics, an interdisciplinary field associated with language acquisition theory and language teaching practice. It aims to contribute to this discipline in the area of Second Language Acquisition (SLA) theory and research. The study concentrates specifically on the cognitive aspects of language learning theory, addressing second language vocabulary learning, specifically in relation to computer-assisted language learning (CALL) environments. The research aims to explore the differential influences of targeted features of instructional software designed by the researcher. To achieve this, it investigates multimedia coding elements, with potential for a differential impact on second language vocabulary learning and processing, to link the design of CALL software to earlier theoretically grounded SLA research. Moreover, the majority of SLA studies involving adults, with several exceptions (e.g. Klein & Dimroth, 2009; Gullberg, Roberts, Dimroth, Veroude, & Indefrey, 2010; Gullberg, Roberts, & Dimroth, 2012; Rast, 2010), have investigated second language learner's abilities in terms of processing or developing linguistic representations that are based on data drawn from intermediate stages. Unlike the majority of research, the current study controls for the influence of pre-existing linguistic knowledge by choosing participants for whom the target language is unknown. By doing so it intends to capture the abilities of learners when they first encounter the target language, when it is completely unknown and typologically distant. Overall, the principal aim of this thesis is to acknowledge and understand the effects of CALL research on vocabulary learning, as a means to more fully inform SLA theories.

The study explores the effects on the successful acquisition of Arabic vocabulary as a foreign language (AFL) via three different interactive interfaces in a multimedia CALL environment: a verbal-based menu driven interface (L2-L1: MDI), a visual-based graphical user interface (L2-Picture: GUI) and a visuospatial-based zoomable user interface (L2-Context: ZUI). As far as the researcher knows, ZUI is used here for the first time in the field of vocabulary CALL, and is provided in conjunction with other interfaces, using Computer-Assisted Vocabulary Acquisition software (CAVA) for intentional multimedia vocabulary learning that has been designed by the researcher.

The study also investigates learners' different cognitive capacities (as measured by working memory and short-term memory tasks). The present research examines how these three different interactive multimedia CALL interfaces might affect how learners encode, associate and recall novel words. Thus, the study does not concentrate only on achievement through different multimedia representations, but also sheds light on the underlying cognitive processes employed.

The research examines various aspects of vocabulary learning, including how learners, who have no previous knowledge of the target language, encode L2 words into their mental lexicon, and subsequently retrieve them. In addition, it explores what components of memory learners might employ to process new second language vocabulary items, and how differences in participants' verbal and visuospatial memory spans impact their performance. Cognitive abilities are measured using the Automated Working Memory Assessment (AWMA), to investigate their effect on learners' achievements and their progress using the three different multimedia CALL interfaces.

Furthermore, as the current research will look at the cognitive role during language processing, it is useful to consider that the cognitive load might arise from any of the independent study variables, teaching methods (MDI, GUI, and ZUI), or memory span (high, and low). To examine this, pupillometry will be applied in a follow up study within this thesis. This is significant as it has been shown that changes in pupil size relate to task demands during cognitive tasks (Kahneman, 1973), and pupillometry is therefore a sensitive indicator of cognitive effort (Beatty, 1982).

1.2 Motivation for the current study and its context

The present study was motivated by the observation that the majority of previous multimedia vocabulary CALL interventions failed to be efficiently applied to operationalize SLA theory and memory research, despite the potential to exploit cutting-edge technology. For instance, previous research has not taken into account the assumption of separation between visual and spatial abilities (Logie, 1995), as a means to take advantage of spatial memory to enrich the learning process. Taking such assumptions into consideration could result in more effective application of technology for language learning, and also improve the methods used to teach vocabulary.

A review of different CALL studies since the 1980s, conducted by Ma and Kelly (2006), reveals greater focus on vocabulary than grammar and other language skills, such as reading, writing, listening and speaking. However, communicating L2 words to present two aspects of word knowledge (word form + word meaning) and connect these are the most popular methods for introducing vocabulary in CALL environments. The methods of presenting L2 words to strengthen form-meaning links have an advantage in terms of vocabulary building, although the fact that "lexis is the core or heart of language" (Lewis, 1993, p. 89) promotes the increased effectiveness of other CALL methods to deliver improved vocabulary learning outcomes. In reflecting on word knowledge as being more than simply a mapping of form and meaning, Nation (2013) argued that:

> Words are not isolated units of the language, but fit into many interlocking systems and levels. Because of this, there are many things to know about any particular word and there are many degrees of knowing. (Nation, 2013, p. 44)

There is thus a need to address various aspects of word knowledge beyond the form and meaning links when designing vocabulary CALL programs. The current research employed technology to teach multiple significant levels of word knowledge (i.e. word visuospatial association), with the potential to create additional relationships between words. Knowing these relations is beneficial when understanding the meaning of the target words (Nation, 2013). This might lead to a more rapid and effective memory storage and retrieval of lexis, due to enhancing the encoding and retrieval of words and promoting the development of lexical networks. Furthermore, effective vocabulary acquisition is likely to be influenced by other variables. For instance, it is commonly assumed that individual differences in working memory play an essential role in vocabulary learning (e.g. Baddeley, Eldridge, & Lewis, 1981; Speciale, Ellis & Bywater, 2004; Service, 1992). Arguably, investigating these differences will provide information concerning the compatibility of different learners' cognitive ability with specific types of multimedia representation, by which they might acquire lexical items most efficiently. Therefore, in order to ascertain whether there is any relationship between the efficiency of each of the three multimedia representations (MDI, GUI, or ZUI) and learners' memory abilities (verbal, visuospatial), this study will examine the role of working memory and short-term memory. Further, pupillary dilation will be recorded as a measure of mental resource demand, during the study's experimental tasks.

1.3 Research questions

The present study aims to contribute to developing our understanding of the vocabulary acquisition process and, in turn, the field of SLA: specifically, how multimedia CALL environments enhance vocabulary learning for learners in the first phase of the learning process. To achieve this, two main studies will be conducted to answer the following principal research questions:

1- What is the effect of three different interactive interfaces types; namely verbal-based Menu Driven Interface (L2-L1: MDI), visual-based Graphical User Interface (L2-Picture: GUI), and spatial-based Zoomable User Interface (L2-Context: ZUI) on SLA?

2- What is the effect of individual differences in verbal and visuospatial working memory, and of short-term memory capacity on the initial uptake of a new vocabulary?

3- What is the impact of the three different interactive multimedia representation types; (MDI, GUI, and ZUI) on response accuracy, and response time, and pupil responses for L2 word retrieval?

4- What kind of performances (accuracy, response time, pupil size) can be observed across the three multimedia representations (MDI, GUI, ZUI), as indicated by the EyeLink Data Viewer? Is there a correlation between these performances and participants' individual differences in memory?

1.4 Research hypotheses

This research will test the following hypotheses:

1- L2 learners' long and short-term vocabulary acquisition varies significantly when using different multimedia representations (i.e. MDI, GUI, ZUI) to acquire L2 word meaning.

2- Learners with high proficiency in working memory and short-term memory will recall more L2 words than low-ability learners (i.e. verbal-MDI, visual-GUI and visuospatial-ZUI) in both immediate and delayed post-test scenarios.

3- Encoding of L2 lexis into the bilingual mental lexicon varies according to the teaching method, rather than the level of second language proficiency.

4- Learners' response accuracy and response time for L2 word retrieval will differ significantly relative to different multimedia representations (i.e. MDI, GUI, ZUI).

5- High-ability L2 learners with stronger short-term and working memory will recall L2 words faster and more accurately than low-ability learners.

6- Learners' mental effort (as revealed by pupillary dilation) differs significantly when retrieving L2 words, relative to different multimedia teaching representation types (i.e. MDI, GUI, ZUI).

7- Low-ability L2 learners with weaker short-term and working memory will show greater pupil dilation (as a physiological indicator of processing load) when recalling L2 words, than high-ability learners.

1.5 Conceptual framework and rationale

This research investigates and compares three interactive multimedia representation types (MDI, GUI and ZUI) to address vocabulary acquisition in CALL environments. This interdisciplinary thesis relates to several branches of knowledge, including educational technology, SLA, neuropsychology, cognitive development and the structure of the mental lexicon. Therefore, this thesis posits a theoretical framework at the junction of these fields of study. The following sections in this thesis comprise a review of relevant literature from related fields, to identify the features of effective vocabulary CALL software.

This research examines how novel L2 words are networked and stored in the mental lexicon, and which specific features, if any, of the three multimedia CALL interfaces facilitate acquisition. The study considers one area of linguistic knowledge, which is the acquisition of L2 lexis through three multimedia representations including different user interfaces. Therefore, the study focuses on the similarities and differences between the effects of the three representation types and how learners with different working memory abilities acquire novel L2 vocabulary items.

Cognitive Learning Theory is the primary theoretical foundation framing the current research. This theory can be used to evaluate the ability of vocabulary CALL interventions to operationalize SLA, and memory research principles of encoding and accessing lexis. The interventions designed in CAVA can be evaluated by focusing on three aspects of the vocabulary learning process: how the three CAVA representations network the words in a bilingual lexicon, the coding elements they employ, and their compatibility and interactivity with learners' working memory and short-term memory abilities. This interactivity will be more successful as this study will discuss vocabulary CALL design from a new angle. To date, no previous studies have applied pupillometry to the study of cognitive processes in multimedia CALL environments. Therefore, this thesis will add to the body of knowledge associated with vocabulary learning by exploring the effective features of the three multimedia representations under examination. As a result, this will provide insight to further improve vocabulary and CALL design.

1.6 Definitions of terms

Multimedia CALL

Multimedia CALL refers to language learning enhanced by multimedia technology. Multimedia for language learning includes a varied range of aurally and visually supported instructional materials. Multimedia CALL includes applications designed using multiple modalities, including text, images, drawings, graphics, animation, and sound in various interactive ways. These materials are used to create interactive educational software applications for language teaching.

Menu Driven Interface (MDI)

Menu Driven Interface (MDI) is a type of interface that allows the user to interact with a computer by working through a series of slides or menus. The computer screen presents the user with a menu form displaying choices that move them to the next menu, or to a slide that appears on the screen. This type of interface predominantly appears in written text, such as lists of choices of commands that might involve sounds.

Graphical user interface (GUI)

This is a program interface that takes advantage of graphics capabilities to facilitate interaction with the software. The use of visual design elements to interact with the computer is supported by GUI, rather text only.

Zoomable User Interface (ZUI)

A Zoomable User Interface (ZUI) is a type of graphical user interface (GUI) in which objects can be represented at different levels of scale, providing more detailed information. The user can change the scale of the area being viewed by zooming in or out to show details.

Computer Assisted Vocabulary Acquisition (CAVA) software

CAVA is software designed by the researcher of the current thesis for the purpose of intentional vocabulary learning. It employs three types of multimedia user interfaces (MDI, GUI, and ZUI).

Automated Working Memory Assessment (AWMA)

Designed by Alloway (2007), Automated Working Memory Assessment (AWMA) includes twelve subtests in its long form, and is used to measure individual differences in working memory and short-term memory capacity.

Pupillometry

Since the 1960s, the method of measuring the pupil diameter (in short - pupillometry) has played an important role within the field of psychology. Changes in pupil size have been demonstrated to successfully indicate intensity and cognitive demand in relation to mental activity.

CHAPTER TWO: FIRST STUDY

This chapter details the first experimental study conducted for the current thesis, and is organized as follows: First, the theoretical framework for the study is presented, focusing on reviewing literature pertaining to areas relevant to the first study. It begins by reviewing the role of working memory, and its potential to directly impact the lexical acquisition process. The focus then shifts to reviewing different types of CALL interventions for teaching vocabulary and their application in previous research, followed by a discussion of the potential added value of the ZUI. Second, the methodology section will be presented, beginning with in-depth discussions of AWMA and CAVA. Finally, the results from the first study will be provided, followed by a discussion of the findings.

2.1 Literature review

To increase our understanding of SLA, it is important to understand how language interacts with other cognitive processes, and the role of memory as a cognitive tool. There is little doubt that memory plays a crucial role in learning, and language learning is no exception. Pavičić Takač (2008) stated that one of the main concepts related to the cognitive theory of learning is that memory stores and retrieves linguistic structures and codes in the same way it does other information.

To achieve better learning outcomes, Jonassen (1992) suggested addressing and developing technological tools to display representations that facilitate cognitive processes. A learner's internal cognitive tools can be blended with instructional technology to construct knowledge (Jonassen, 1992). Therefore, it is essential to understand the elements involved in acquiring vocabulary knowledge, in particular how L2 words are stored and integrated cognitively, and the role of memory.

Ebbinghaus wrote the first classic book on the science of memory in 1885, two years after conducting a series of experiments on himself. His experimental findings resulted in the founding of the verbal learning approach (Ebbinghaus, 1913). This approach was later criticized by McGeoch and Irion (1952), due to its use of a limited range of methods; i.e. it involved memorizing lists of words only. A second major development in this field was Gestalt psychology, which was influenced by the study of perception as a mechanism to understand memory. Two of the main investigators were George Mandler and Endel Tulving, who developed an alternative approach to verbal learning, which became popular in both Europe and North America. The Gestalt approach placed greater emphasis on learner's activities when organizing materials. However, these materials were criticized as meaningless by Fredrick Bartlett (1932) in his book *Remembering*. As an alternative, Bartlett developed a third approach to memory in Britain, utilizing more complex materials, such as folk tales. He was particularly interested in the errors people make when remembering, claiming that these reflect their assumptions about the world. These initial approaches were subsequently influenced by the development of computers during the 1950s and 1960s, and forerunners of Cognitive Psychology. As the cognitive approach to psychology developed, the assumption that there is a single memory system was replaced by the notion of three memory systems. This led Atkinson and Shiffrin (1968) to propose the modal model, which hypothesizes that information is received by the sensory memory before being passed to the short-term memory, for later transference to the long-term memory.

2.1.1 The Multicomponent Model of Working Memory (M-WM)

In 1974, Baddeley and Hitch introduced the Multicomponent Working Memory model (M-WM). The revolution brought about by the M-WM model had been motivated by the requirement to resolve three major issues that had arisen with Atkinson and Shiffrin's (1968) earlier model (Baddeley, 1986). The first related to the assumption that holding materials in short-term memory for an extensive period of time is sufficient for learning. This notion was initially challenged by Craik and Lockhart (1972), who proposed the principle of levels of processing as an alternative way to describe how learning takes place when processing material, rather than relying on time. Second, the earlier model assumed short-term memory is essential to effectively access long-term memory. This was evidently not the case, as some patients had inefficient short-term storage capacity, but unimpaired long-term memory. Third, the model's assumptions about short-term memory playing an essential role in cognition were challenged by neuropsychological studies, which reported that patients with impaired short-term memory behave normally in other life situations (e.g. can successfully run their own businesses). Baddeley and Hitch (1974) renamed short-term memory *working memory*, to clarify that it serves as a system that not only stores information but also organizes, manipulates and transforms material for transference to the long-term memory (Baddeley, Eysenck, & Anderson, 2015). Many researchers have subsequently confirmed the applicability of this model's primary structure. The formulation of a multicomponent model arose from evidence accumulated from different sources, which revealed that different systems within the brain perform different cognitive functions in a highly independent way (Baddeley, 2012). The model is sufficiently developed to provide a framework within which to generate theoretically derived hypotheses that could be experimentally investigated, either with neuropsychological patients, or with normal individuals. A detailed review of these studies, found to be pertinent to the overall objectives of the present research, will be presented below.

Initial version of the M-WM model

Neuropsychological experiments conducted in the field of memory research during the 1960s and 1970s provided useful insights for Baddeley and Hitch, as they worked towards developing a theory of memory. However, because short-term memory deficit patients were difficult to access, Baddeley and Hitch used alternative methods to test this capacity. They utilized suppression tasks to disrupt individuals' short-term storage systems to varying degrees. For instance, they required participants to recall digits of varying length while completing another learning task concurrently. By posing different learning task challenges, such as visual reasoning and language comprehension, Baddeley and Hitch (1974) identified the system that underpins digit span. The findings revealed from their experiments, in addition to neuropsychological evidence from elsewhere, led to the development of the first version of the M-WM, which presents working memory as a three-component system, rather than a single unitary store (see figure 2.1) (Baddeley & Hitch, 1974). They referred to these three components as modules, whereby each module comprises processes and storage systems that are tightly interconnected within the module itself and also linked loosely across modules. Additionally, each of these components has more remote links to perceptual and long term memory processes. They all have a discrete function but work over a limited span, which constrains any capacity to process more than one task simultaneously. However, there is no interference with other subsystems when they work together, unlike when they work separately.

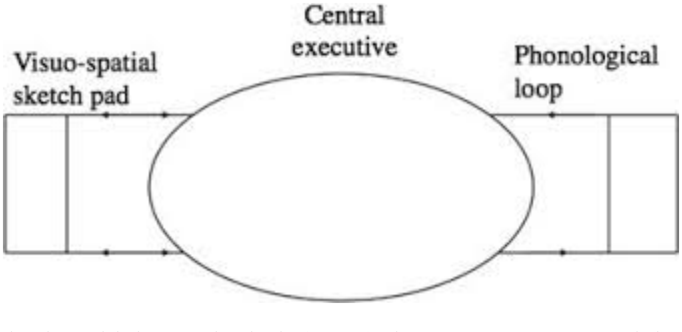


Figure 2.1. The original Baddeley and Hitch (1974) three component model of working memory. (Source: Baddeley, 2012, p.6)

The central executive

Regardless of its limited capacity, the central executive (CE) is thought to be the most important function responsible for dealing with cognitive tasks, with attention being on its main purpose. In order to explain the functions of the CE as a major component of working memory, Baddeley initially included the Norman and Shallice (1986) model of executive functioning as a mode of attentional control. Norman and Shallice argued that an action can be controlled in one of either two attentional ways: through a habit-based system (i.e. when driving a car), or the supervisory attentional system (SAS), for actions that cannot be controlled in a routine way (i.e. when responding to the closure of a road). In light of Norman and Shallice's (1986) research, and drawing on further theories of attention, Baddeley (1986) incorporated the CE system into the M-WM model, suggesting that the CE carried out three essential processes: focusing attention, dividing attention between two simultaneous tasks, and switching attention from one task to another. Directing and dividing attention between the tasks at hand is one function of the CE that enables the individual to carry out two tasks simultaneously (i.e. driving a car and talking to a passenger). Eysenck and Keane (2010) concluded that people utilize the CE heavily for every cognitive activity. They observed that the CE supports tasks such as reading and problem solving. In another context, speaking for instance, Stuss and Alexander (2007) mentioned that there are three functions of the CE involved in speaking: the speaker plans what he is going to say (task setting), concentrates on delivery (energization), and then checks that what he is actually saying is as planned (monitoring). It is widely concluded that the prefrontal cortex is the most active part of the brain when using the CE (Eysenck & Keane, 2010).

Baddeley (1986) views the CE as an attentional controlling system, linked with the phonological loop and visuospatial sketchpad; the two CE slave systems that provide short-term storage as well as manipulation of information. While the phonological loop temporarily stores and manipulates verbal information, visual and spatial information is preserved in the visuospatial sketchpad (Baddeley et al., 2015). The assumption that verbal and visual short-term memory systems are separate corresponds to evidence from neuropsychological studies. Some patients have impaired visual and preserved verbal short-term memory (Della Sala & Logie, 2002), whereas others show intense deficits in verbal short-term memory, yet their visual short-term memory is intact (Shallice & Warrington, 1970). The assumptions here are also supported by behavioural data collected from unimpaired participants (Meiser & Klauer, 1999; Della Sala et al., 1999).

To evaluate the significance of the separation of working memory components, Farmer, Berman and Fletcher (1986) performed tasks that relied on specialized resources, without a requirement for other working memory resources. Each task designed was expected to interfere with the storage and processing of one working memory component but not with the other. For example, it was assumed that tapping metal plates would load the visuospatial sketchpad. However, according to Farmer et al. (1986) this visuospatial suppression task did not interfere with the concurrent reasoning task, which is thought to rely on the CE. However, the phonological loop was inhibited by an articulatory suppression task. These findings suggest that Baddeley's M-WM model offers a better explanation for performance on short-term memory tasks than other theories, which assume a unitary-store model of memory (e.g. Nairne, 1990; Cowan, 1999).

The phonological loop

The phonological loop is another important component of working memory. It is thought to be a slave system to the executive centre with two subcomponents. The first subcomponent of the phonological loop is a passive store of speech-based items, and the second is an articulatory rehearsal process (Baddeley et al., 2015). The phonological loop is modular in nature, with only a limited capacity to temporarily hold and manipulate verbal information. This limited short-term storage is thought to be adversely affected by the displacement of information or overwriting. The phonological loop model was proposed by Baddeley and Hitch (1974), after implementing a method of examining that involved evaluating evidence associated with different phenomena, such as the phonological similarity effect, the word length effect, and irrelevant sound effects. The availability and applicability of simple common tools to test the phonological loop, such as suppression effects, phonological similarity, and word length, perhaps explain why it is the most explored aspect of working memory.

One of the crucial characteristics of the phonological loop, which relates to the current thesis in terms of the menu driven user interface (MDI) condition, is that verbal information is intended to be fed into the phonological store in two ways. Spoken material can be encoded directly into the phonological store. Whereas, written material, such as digits or letters need to be subvocalized, whereby the reader verbally rehearses the items, for encoding and processing in the phonological store.

The majority of research suggests that phonological working memory, as measured by different tasks, such as digit span and word recall, is involved in the execution of text comprehension (e.g. Baddeley, Eldridge, & Lewis, 1981; Just & Carpenter, 1992), first language vocabulary development (e.g. Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992), SLA in formal settings (e.g. Dufva & Voeten, 1999; Speciale, Ellis & Bywater, 2004; Palladino & Cornoldi, 2004; Service, 1992), and second language vocabulary acquisition (e.g. Atkins & Baddeley, 1998; Baddeley, Papagno, & Vallar, 1988; Gathercole & Baddeley, 1990; Papagno & Vallar, 1992; Martin & Ellis, 2012). To exemplify this; Baddeley, Papagno and Vallar (1988) studied the case of patient PV, who was known to have very poor phonological loop abilities. She had difficulties understanding long sentences, since she would forget the first word in a sentence. The researchers hypothesized that this resulted from her phonological loop deficit, which meant that she could not recall the first few words of the sentence. Another hypothesis, of relevance to the current thesis, is that the phonological loop is involved in SLA. To test this, the researchers taught patient PV, whose first language was Italian, eight Russian words along with other unimpaired participants. After ten trials, the control participants learned all eight words, whereas PV learned none. Therefore, the researchers concluded that the phonological loop serves as a language acquisition device (Baddeley, Papagno, & Vallar, 1988).

However, to avoid reaching a misleading conclusion from an individual case, and, in view of the lack of such patients, Papagno, Valentine and Baddeley (1991) conducted another study, disrupting the phonological loop of normal participants when learning foreign vocabulary. They used articulatory suppression (asking the participant to repeat irrelevant sounds when learning). This disruption of the phonological loop created more challenges when learning second language vocabulary, than it did when associating pairs of words in a native language (Papagno, Valentine, & Baddeley, 1991). These results support the conclusions drawn from patient PV, that the phonological loop is important to second language learning.

Further evidence regarding this was derived from a study in the classroom setting conducted by Service (1992), which suggested there is a relationship between phonological representations in working memory and foreign-language learning. In a large educational experiment, the verbal working memory of 44 Finnish primary-school children (aged 9-10 years) was measured using two tasks related to working memory codes; the oral repetition of spoken pseudowords, and the delayed copying of visually presented pseudowords. Divided into ten lists, fifty pseudowords were constructed obeying English phonotactics and phonology, and another fifty Finish-sounding pseudowords. *Rendence* and *disajoinance* are examples of English-sounding pseudowords and *laira* and *hainuksia* were Finnish-sounding pseudowords. The subjects received two lists of pseudowords in each of four separate test sessions. Their responses were recorded individually in a room at their school. Correlations between the learners' accuracy, which reflect their abilities to create phonological representations in working memory, were compared with test scores when learning English 2.5 years following the first assessment. Service (1992) concluded that the ability to represent unfamiliar phonological material in working memory probably facilitates the acquisition of new foreign language vocabulary.

Although only one limited component of working memory, the significance of the phonological loop in providing a short-term store and verbal information processing centre might also enhance learning words via MDI representation. It is a system that is highly beneficial and worthy of consideration when designing CALL interventions.

The visuospatial sketchpad

The second slave of the CE is the visuospatial sketchpad. Baddeley's conceptualization of the visuospatial sketchpad was influenced by advances in visuospatial memory research during the 1970s. This suggested visual short-term memory has a limited capacity, serving as a system for temporarily storing and manipulating visual and spatial stimuli. As a memory system, there needs to be an opportunity to feed the sketchpad from exterior and interior sources. There also should be explanations for forgetting, such as memory decay or interference from new information. In addition, the sketchpad should rehearse ways to extend storage time. Moreover, as a buffer, there must be relationships between storage codes and the features of visuospatial materials. These memory characteristics are common among other memory systems, such as the phonological loop, and both memory systems complement each other. The characteristics of the system have been explored by designing and conducting experiments, taking into consideration basic memory functions thought to temporarily store, process, and retrieve visual and spatial information.

Thus far in the visuospatial working memory literature, the investigation of sketchpad features focused on three areas: visual similarity, visual and spatial distinction, and visual imagery. With regard to visual similarities, it emerged that confusion can arise, as errors effect the retrieval of comparable visual materials (e.g. Hitch, Halliday, Schaafstal, & Schraagen, 1988). For instance, Hitch et al. (1988) reported that the retention and drawing of a series of distinct pictures, such as a pen, a ball, and a pig, is easier than when the images are a pen, a brush and a rake. This indicates that the short-term visual store is limited in capacity, relying on visual codes that are interfered with and affected by the presence of similar visuals input.

The visual and spatial distinction approach to investigating visuospatial working memory has provided a tractable way of studying the nature of this element of memory. Logie (1995) suggested the partitioning of visuospatial working memory to create an *inner scribe*, comprises a perceptual system and a *visual cache* for temporary storage. Although they work together in practice, the visual cache temporarily only stores visual form and the colour of a stimulus, whereas the inner scribe processes spatial and movement information. For example, if you were in your office and the light suddenly turned off, your ability to remember the location of the door would be a spatial memory function, relating to the answer to the question (where?). Remembering (what?) objects were on your desk is a visual memory function, pertaining to the physical description of an object, such as its shape or colour (Baddeley et al., 2015).

Cognitive tasks have been developed previously to differentiate spatial memory (remembering where) from visual memory (remembering what). Corsi span is an example of a spatial task requiring the participant to imitate a sequence of taps on a number of blocks made by an experimenter, who then increases the length of the sequence to the point at which performance breaks down. The participant in this task must remember *where* the location of each of the taps was and also recall the correct order. In contrast, there are multiple task options for assessing visual span. One of them is to use matrix patterns, in which the participant sees a pattern and, when it disappears, they have to reproduce it by pointing to the cells previously filled in the empty matrix. The number of cells in the matrix is then increased until performance breaks down. The matrix begins with four cells and reaches 16 (Phillips, 1974). When completing this task, the participant must remember *what* the pattern looks like. However, it is still a crucial issue, which has been widely debated, whether the working memory system combines visual and spatial processing or of they are separate systems. Other studies have reported double dissociations between visual and spatial tasks, which support the division of the visuospatial sketchpad into separable visual and spatial subcomponents (for a review of this, see Klauer & Zhao, 2004).

After presenting their review, Klauer and Zhao (2004) applied two working memory tasks: a visual task (memory for Chinese characters) and a spatial task (memory for dot locations) with secondary discrimination tasks, both with and without visual and spatial interference. They found the spatial interference task proved a greater disturbance in the main spatial task than the main visual task, whereas the visual interference task disturbed the main visual task more than the main spatial task.

The notion of separate visual and spatial systems has been supported by additional neurological evidence. According to Eysenck and Keane (2010), studies upholding the notion of separation have shown that different brain regions are involved in visual and spatial tasks. During visual tasks, the ventral prefrontal cortex is activated, whereas the dorsal prefrontal cortex is generally activated during spatial tasks. These findings are consistent with other re-cognition processing studies (e.g. Diana, Yonelinas, & Ranganath, 2007) that have suggested there are two different pathways for visual and spatial stimuli. The binding-of-item-and-context model (Diana et al., 2007) detailed evidence specifying the brain areas involved in different areas of memory:

- Information about particular items (what) is received by the perirhinal cortex.
- Information about context (where) is received by the parahippocampal cortex.
- (What) and (where) information is received and bound in the hippocampus, to create item-context associations, which permit recollection.

The binding-of-item-and-context model is supported by functional neuroimaging studies. Research by Diana et al. (2007) reviewed several studies of brain activation. As anticipated, the perirhinal cortex was activated by familiarity. In contrast, the parahippocampal and hippocampus were activated by recollection (Diana et al., 2007).

Further support for the assumption that there is a distinction between spatial and visual memory spans arises from neuropsychological studies performed with numerous patients. For example, after suffering a head injury in a car accident, Patient LH experienced impaired visual memory. The impairment affected colours and shapes, but he had normal verbal and excellent spatial memory (Farah, Hammond, Levine, & Calvanio, 1988). Another patient, MV, suffered brain damage; following this, his visual memory was preserved but his spatial abilities were impaired (Carlesimo, Perri, Turriziani, Tomaiuolo, & Caltagirone, 2001).

Investigating the retention of visual and spatial information, Logie and Marchetti (1991) conducted a study exploring the role of the visuospatial sketchpad. Their subjects were shown two types of visual representation. The first were visual in nature, as the participants were presented with coloured hues. The second task was intended to be spatial, and the participants were required to retain a sequence of squares presented at different locations on a screen. This task employed spatial memory, as the participants rehearsed the sequence of locations in which the squares had appeared. After the presentation, the participants were exposed to either irrelevant pictures, or a movement task, before completing a subsequent recognition test. As predicted, the results revealed that the participants' ability to remember the shade of the colours was disrupted by the display of irrelevant pictures, whereas the movement task disrupted the retention of spatial locations. This reaffirmed the hypothesis that there are separate memory systems involved in retrieving visual and spatial materials. This finding supported Logie's (1995) view that there are two discrete subcomponents of the visuospatial working memory, which are a passive visual store (visual cache) for processing visual information, and an active rehearsal process (inner scribe) for rehearsing and retrieving spatial information.

Overall, the research seems to suggest the visuospatial sketchpad is involved in short-term storage and in the processing of visual and spatial materials. Although the visuospatial sketchpad is a slave system in the working memory, which is assumed to be responsible for the encoding and manipulation of temporary visuospatial representations, it is sufficiently developed to be able to play an essential role in learning processes that involve pictorial and visuospatial representations, as arises under GUI and ZUI conditions. Moreover, although connected within the sketchpad, the spatial and visual aspects of working memory seem to be detachable, as was taken into account when designing ZUI representations.

The episodic buffer

A fourth component of the working memory model was proposed by Baddeley in 2000. This is the *episodic buffer*, which integrates information from the phonological loop, the visuospatial sketchpad and long-term memory, briefly storing it (see figure 2.2.). According to Repovš and Baddeley (2006, p. 15), the episodic buffer

> …is episodic by virtue of holding information that is integrated from a range of systems including other working memory components and long-term memory into coherent complex structures: scenes or episodes. It is a buffer in that it serves as an intermediary between subsystems with different codes, which it combines into multi-dimensional representations.

Baddeley (2000) stated that the CE and episodic buffer work together to integrate visuospatial and verbal information from the phonological loop and the visuospatial sketchpad, and then assimilate it according to prior knowledge stored in the long-term memory. However, the significance of the functions of the episodic buffer have not been integrated into multimedia learning theories.

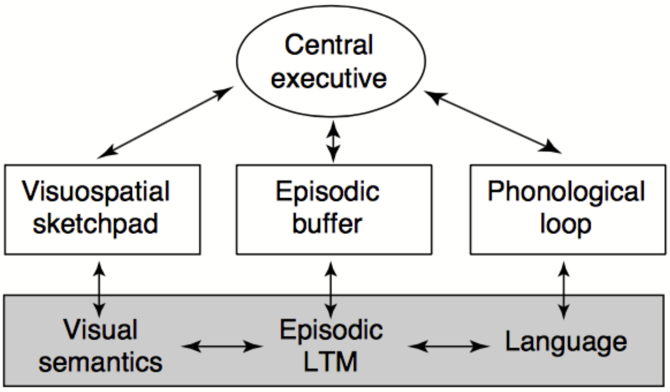


Figure 2.2. A later development of the multicomponent working memory model. (Source: Baddeley, 2012, p.16)

As touched on above, diverse reviews have confirmed the claim that working memory is highly associated with second language processing. In a recent meta-analysis review, Linck, Osthus, Koeth and Bunting (2014) concluded that working memory is instrumental in second language learning, as it reports differential contributions from executive and phonological subsystems. The other direct correlation between working memory and second language learning and processing was confirmed in a more recent meta-analytical review based on empirical studies, conducted by Grundy and Timmer (2017). Their results, collated from 27 independent studies involving 2,901 participants, showed that bilinguals outperformed monolinguals in terms of their working memories. In another review of experimental, developmental and neuropsychological studies, Gathercole and Baddeley (2014) mentioned the close link between learners' phonological working memory abilities and language acquisition, for both native and foreign language learners. They emphasized that the distinctive features of newly learned words facilitate their retention in the phonological working memory, allowing them to be manipulated before transference into the long-term lexical memory. Factors found to influence the phonological loop, such as articulatory suppression, similarities, and word length, impact the long term learning of new foreign words. Gathercole and Baddeley (2014) considered the evidence in detail, suggesting that the visual forms of the new words are translated into sequences of subvocal articulations for storage in the phonological loop. This suggests that the modularity domain of the phonological loop is primarily language, including sign language and lip reading. Moreover, Williamson, Baddeley and Hitch (2010) also assume the phonological loop processes musical sounds.

The M-WM model has been applied in several multimedia learning experiments when teaching subjects other than language. Gyselinck, Ehrlich, Cornoldi, Beni and Dubois (2000) asked participants to study physics (e.g. static electricity), by showing them written texts under one condition and written texts with pictures in another. Learners with greater spatial abilities outperformed low spatial ability learners when information was presented pictorially. The researchers used the Corsi block task as a visuospatial measurement, and checked that all the participants were comparable in their verbal abilities when using the digit span task. In addition to using the Corsi block task as a capacity approach, the researchers applied a dual task approach in a follow up study (2002), involving the inclusion of a spatial tapping task during learning. Again, the high spatial ability participants were found to rely more on their visuospatial sketchpads when studying pictures. It was concluded that they were exceedingly disturbed by the spatial tapping task, which had no effect on the low spatial ability participants. In the second experiment, Gyselinck, Cornoldi, Dubois, De Beni and Ehrlich (2002) found that the spatial tapping task did not interfere with text processing, indicating the visuospatial sketchpad is not involved in text manipulation. Thus, the results revealed the involvement of the phonological loop in text processing, but not in picture processing.

Another study (Pazzaglia, Toso and Cacciamani, 2008) supported the suggestion that working memory components contribute to multimedia learning. Pazzaglia et al. (2008) asked Italian students to study Germany's geography via a hypermedia representation, consisting of written text, spoken text, and pictures. The researchers measured the participants' visuospatial abilities using the capacity approach, in both its simple and complex span assessments, by applying a simple span task, the Corsi block task, as well as a complex span task, the dot matrix task, which was assumed to assess CE in relation to Visuo-Spatial Sketch-Pad capacity (CEVSSP). In addition, the researchers tested the participants' verbal abilities. In a similar way to that described above, they used a simple span task, the digit span, and a complex span task, the listening span. The latter was intended to assess CE engagement with the Phonological Loop (CEPL). Their results revealed that only complex capacity tasks, CEVSSP and CEPL, were significantly correlated with semantic knowledge acquisition and map recognition, respectively. The simple span was not found to relate to learning outcomes.

The CEPL was found to be involved in multimedia learning from different perspectives (Austin, 2009; Doolittle, Terry, & Mariano, 2009; Doolittle, & Mariano, 2008). Using a complex span task, the operational span Task, Doolittle et al. (2009) found learners with a high CEPL capacity outperformed low-capacity learners when working from a pump in its three conditions; animation and written text, animation and spoken text, or animation, spoken, and written text. Similar results were reported in Austin's (2009) study covering learning about lightning under the same multimedia conditions. The contribution of the CEPL span to multimedia learning was indicated by Sanchez and Wiley (2006), who tested their participants using the operational span and reading span tasks. Both are complex tasks for measuring the phonological loop, as well as the CE. The participants were assigned to text-only, text and relevant picture, or text and irrelevant picture. The high CEPL learners suffered less interference from irrelevant pictures than the low CEPL learners. These results were confirmed with an eye tracking study, which showed low CEPL learners choose to spend more time looking at seductive pictures. Dutke and Rinck (2006) tested the CEPL and CEVSSP of their participants using reading span and spatial span tasks. They distributed them into four groups: low CEPL, low CEVSSP, high CEPL, and high CEVSSP. The results indicated that low CEPL learners encountered the greatest difficulties integrating words and icons, but not when the presentation included only words. Whereas the low CEVSSP found both words and icons challenging, probably due to the spatial arrangements associated with the presentation.

To sum up, a range of experimental and neuropsychological data provides support for Baddeley and Hitch's working memory model, as a system underpinning our capacity to perform complex cognitive activities. It can be concluded from the studies surveyed so far that the working memory plays an essential role in SLA and the multiple working memory model helped to form our understanding of second language processes. According to Baddeley's most recent model (2012), shown in figure 2.3., separate buffers describe the different forms of information. These buffers, in turn, are separate from LTM. Therefore, this model is used in the present thesis, as grounded theory underpins the three conditions and frames the design of the CAVA interfaces.

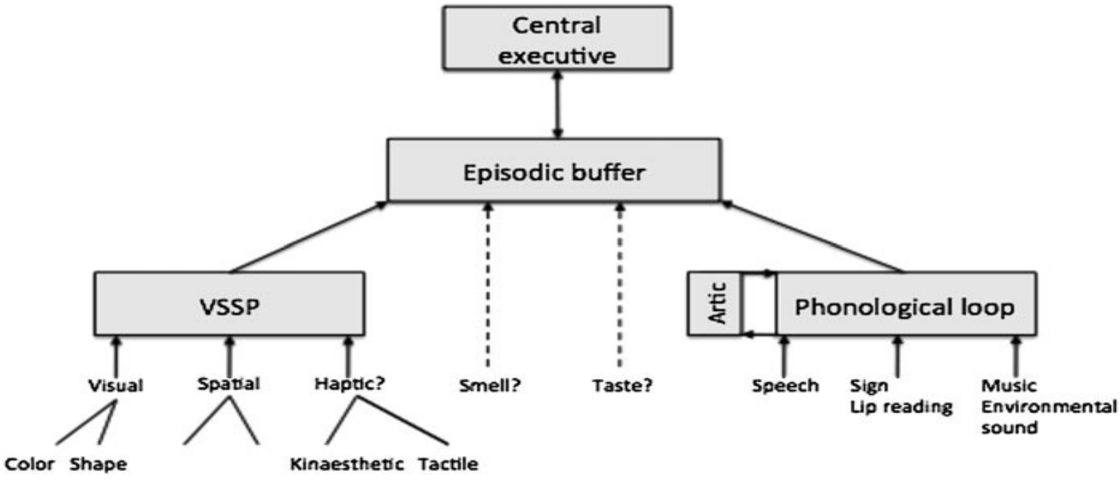


Figure 2.3. Current elaboration on the original Baddeley and Hitch model, showing the flow of information from perception to working memory. (Source: Baddeley, 2012, p.23)
N.B. VSSP = Visuospatial Sketchpad.

In the first condition, the MDI will employ verbal based design elements (written texts) as the major medium for communicating the meaning of L2 words in CAVA. This involves showing the written form of translation equivalents. This mode of presenting written L2-L1 lists is assumed to rely on the phonological loop. The second condition of CAVA software is designed to be a GUI, which comprises pictures and L2 words. According to this type of representation, the meaning of target L2 words will be communicated visually through an illustrated picture. The information in this interface is principally manipulated by the visual cache of the visuospatial working memory. The third condition refers to the ZUI. This combination of multiple visuospatial media is thought to employ not only the visual cache, but also the inner script of the visuospatial sketchpad, since it shows the objects in relation to their locations in a complete virtual context. It can be assumed that participants scan this visuospatial context in order to study target words; thus, the recall process may then also be spatial. Thus, spatial processing may be involved when processing L2 word meanings that involve visuospatial word associations.

All the studies reviewed to date, however, suffer from the fact that the researchers relied too heavily on the role of the phonological loop in language learning. These studies did not fully investigate the impact of additional components on working memory; such as the effect of the visuospatial sketchpad, on language acquisition. In fact, no attempt was made to explore the association between spatial memory and vocabulary learning. By focusing on the formal system of language, researchers overlooked the fact that one property of the language is that it is "arbitrarily symbolic: Language creates an arbitrary relationship between a symbol and what it represents: an idea, a thing, a process, a relationship, or a description" (Sternberg, 2012, p. 362). Words have different levels of information associated with them, and presenting these aspects of word knowledge through visual and spatial representations could stimulate visuospatial working memory components, thereby playing a role in vocabulary learning. It is therefore possible to develop vocabulary CALL interventions to increase the quantity of information learned, by reducing dependence on the phonological loop. Involving visuospatial annotations might deepen the mental processing that occurs when learning vocabulary, which might then lead to more effective vocabulary acquisition.

2.1.2 Multimedia learning of second language vocabulary

This section reviews experimental studies about multimedia learning in the context of second language vocabulary in CALL environments. It explores the most striking cognitive theories associated with SLA; in particular, considering what the designers of multimedia vocabulary CALL based their interventions on. Specifically, it investigates whether the design of digital resources used in previous multimedia vocabulary CALL research was theoretically grounded, and if it focused on multimedia technology being employed to strengthen the link between the word form and its meaning, or to deliver other aspects of word knowledge beyond form-meaning connections. Moreover, how intentional language-focused and incidental meaning-focused input software has been used for multimedia vocabulary CALL will be reviewed in light of the cognitive theories associated with SLA.

Researchers in the field of CALL have typically focused on vocabulary rather than other language areas and skills (Ma & Kelly, 2006). Vocabulary learning has attracted instructional designers leading to the introduction of a wide range of technologies, including courseware, dictionaries, online activities, corpora and concordances, as well as computer-mediated communication technologies (Stockwell, 2007). Computer-based lexical databases, for instance, are being developed, and are derived from conceptual-semantic and lexical relations. WordNet, for example, is an online thesaurus that groups words together based on their meanings. The first version of WordNet was released in 1991, and at the end of 2007 a third update was made available. Words are mainly interlinked by their semantic relations (such as IS-A-KIND-OF, IS-A-PART-OF, IS-AN-ANTONYM-OF and ENTAILS), meaning that WordNet functions as a huge network of linguistic nodes (Miller & Fellbaum, 2007).

A further example of a form of technology used to communicate vocabulary is concordance. A concordance comprises a list of contexts that reveal how the word is used, and can provide the learner with extensive knowledge about the word. Tom Cobb's (1997) website (Lextutor) contains an online concordance that provides information about a word, such as its collocates, grammatical patterns, and related meanings and homonyms. In an experimental study, Cobb (1997) found that learners who used the concordance scored higher on vocabulary quizzes than those who did not. In a recent longitudinal study, Karras (2016) found English as a foreign language (EFL) learners who used a data-driven learning (DDL) approach (i.e. concordances) outperformed those who were not exposed to DDL. Additional experimental studies have confirmed the value of using concordances in L2 vocabulary learning. For example, using LexisBOARD as vocabulary CALL software, Mirzaei, Domakani and Rahimi (2016) taught L2 lexical items via concordances, polywords, or formulaic sequences to a group of Iranian EFL learners who then outperformed a control group, who had studied mainstream EFL textbooks in the vocabulary post-test. More recently, Lee, Warschauer and Lee (2017) observed that EFL learners given the definitions of target L2 words after having been exposed to concordance lines during online reading tasks, achieved higher vocabulary gains than those provided only with concordance lines. A data-driven approach was also found to effectively enhance second language vocabulary acquisition, both inductively and deductively (Lee & Lin, 2019; Tsai, 2019).

A further early striking example of vocabulary instructional technology is the French Learner's Workbench, a hypermedia French language learning system (Hayet, 1994). This system integrated a collection of modules and packages. Among these packages is Dicologique, a dictionary presenting a wealth of information intended for the study of French. It consists of 105,000 words and phrases and 20,000 notions, which can enter into semantic relations. The learner can view the entry item according to its position in the semantic field, and infer its meaning (Hayet, 1994). The use of a web-based workbench was found to be effective in other experimental studies. For example, Lim (2014) reported that participants were more effective at handling an interpreting task when building their English L2 vocabulary by looking for the required vocabulary via a web-based workbench. In addition, the use of the workbench was also found to be efficient for those participants building vocabulary portfolios that were larger in size and richer in variety.

Other types of technology were designed to employ *multimedia* to enhance second language vocabulary learning, and different multimedia annotations were developed over the course of recent decades. In fact, the term *multimedia* has very different meanings according to the context in which it is used. In a phone company, for instance, the consumer entertainment vendor might consider multimedia as a cable-TV-like package provided over a high-speed internet connection, or an interactive TV running thousands of digital channels. In a computer company, on the other hand, a laptop with advanced sound capability and super multimedia-enabled microprocessors might come into the mind of a hardware vendor when they think of multimedia. In educational environments, instructional designers have a more application-oriented view of multimedia, which involves applications designed using multiple modalities consisting of text, images, drawings, graphics, animation, video, and sound in various interactive ways. This is in contrast to media that uses traditional printed materials and computer displays with simple text-only contents. However, multimedia is not simply about mixing those materials together; rather, it places great emphasis on integrating them with rich interactions amongst the materials, as well as such interactions between the materials and human beings (Li, Drew, & Liu, 2014).

A widely employed feature of common instructional technology, aims to link the learner in vocabulary CALL environments with dictionaries, using simple hyperlinks (e.g. Dilenschneider, 2018; Dziemianko, 2019). This function has also been programmed in flashcard software, such as SuperMemo 2008, vTrain 5.2, and Memory Lifter 2.3, to present a standard definition, a pronunciation of the word, its meaning and its translation. Advanced features of more recent multimedia technologies made it possible for designers to expand the textual/audio content of the flashcards by offering additional pictorial annotations, such as pictures and videos to illustrate the meaning of words. WordChamp, P-Study System Ver.8.3, Quizlet, iKnow!, Word Engine and Learn ThatWord are other examples of online flashcard software that go beyond a simple presentation of L2 word definition and employ multimedia capabilities, such as audio recordings, contexts, and images on flashcards. Other flashcard CALL programs (e.g. iKnow! & WordChamp) provided quizzes to increase retrieval efforts and enhance vocabulary through retrieval practices (for a detailed review on flashcard software, see Nakata, 2011). Practice exercises have also been provided in other types of vocabulary CALL technology, such as discrete-point activities. A common example of online discrete-point activities is Hot Potatoes, which provides learners with six vocabulary and grammar tutorial activities (Levy, 1997).

Multimedia CALL designers have been influenced by the assumptions made by many SLA researchers who argued that intentional vocabulary learning, along with the incidental acquisition of vocabulary, can play an important role in second language vocabulary learning (e.g. Cobb, 2007; Godwin-Jones, 2010; Laufer, 2003; Min, 2008; Nation, 2013; Read, 2000, 2004; Schmitt, 2010; Tozcu & Coady, 2004). Incidental vocabulary learning refers to the non-deliberate learning of vocabulary that arises when learners perform other tasks, such as communication, reading or listening. On the other hand, intentional vocabulary learning refers to the deliberate acquisition of vocabulary when directing explicit attention towards a lexical item for the purpose of learning it. Laufer (2006) argued that if a teacher could not create a second language learning environment with similar input conditions to those associate with first-language acquisition, then intentional vocabulary learning becomes particularly important. Read (2000) observes:

The trend in the 1990s is for many language teachers to discourage learners from memorizing lists of isolated words, on the basis that vocabulary should always be learned in context. Nevertheless, research shows that systematic learning of individual words can provide a good foundation for vocabulary development, especially in foreign-language environments where learners have limited exposure to the language outside of the classroom.

In addition, Nation (2001) argues that 25% of any balanced well-designed foreign language course should be language-focused to support intentional language learning. Other strands include, learning from comprehensible meaning-focused input, meaning-focused output, and fluency development. Consequently, according to Nation (2001), direct vocabulary learning should equate to roughly a quarter of the time made available for vocabulary learning. Vocabulary learning in this context then complements instruction and learning in context, rather than representing a conflict between two distinct modes of learning, which then reinforces and enhances learning as it arises (Schmitt, 2010).

The use of multimedia CALL has been found to be more effective as regards incidental L2 vocabulary learning (e.g. Dilenschneider, 2018; Sato, Matsunuma, & Suzuki, 2013) and intentional learning (e.g. Esit, 2011; Hirschel & Fritz, 2013; Khezrlou, Ellis, Sadeghi, 2017; Rusanganwa, 2015), when compared with the traditional pen and paper learning methods. For example, Sato, Matsunuma and Suzuki (2013) could emphasize automatized L2 word recall with their subjects by exposing them to incidental vocabulary learning and multimedia software with a time-control function, which would then be compared to other subjects using printed text. They concluded that this CALL application facilitated the decoding of word forms, which then led to freeing up working memory to support better comprehension of a reading text. The advantage of CALL over non-CALL vocabulary learning was also found in L2 intentional learning. For instance, Esit (2011) reported a positive effect from intentional vocabulary CALL software, which overlapped traditional methods for teaching English as a foreign language to Turkish learners. In a more recent study, Hirschel and Fritz (2013) compared the use of notebook and CALL approaches to learning L2 vocabulary. In their experiment, 140 Japanese students participated and took pre-, post-, and delayed post-tests. Although there were no significant differences between the two groups on the immediate post-test, the multimedia CALL group performed slightly better in the delayed post-test. In Rusanganwa's (2015) study, multimedia presentation including visual stimuli in combination with aural presentation, enabled EFL students to learn L2 vocabulary more effectively through multimedia and thus outperform their peers learning via a traditional method through chalk and talk. In a recent study, however, Khezrlou, Ellis and Sadeghi (2017) found that if learners explicitly learned a non-CALL L2-L1 list of the target words before reading computer-based texts including multimedia glosses, they made greater gains in terms of L2 vocabulary in word recognition test, as well as maintaining a long-term knowledge of the target words, than participants who did not.

Other studies moved away from comparing vocabulary CALL with non-CALL interventions. Researchers investigated the impact of multimedia annotations on different gloss types for incidental vocabulary learning (e.g. Acha, 2009; Akbulut, 2007; Al-Seghayer, 2001; Boers, Warren, Grimshaw, & Siyanova-Chanturia, 2017; Çakmak & Erçetin, 2018; Chen, 2002; Chen, 2016; Chun & Plass, 1996; Jones, 2009; Jones & Plass, 2002; Kost, Foss, & Lenzini, 1999; Lin & Tseng, 2012; Mohsen, 2016; Plass, Chun, Mayer, & Leutner, 1998; Ramezanali & Faez, 2019; Rouhi & Mohebbi, 2013; Sun & Dong, 2004; Türk & Erçetin, 2014; Varol & Erçetin, 2019; Yanguas, 2009; Yeh & Wang, 2003; Yoshii, 2006), and investigated the effects of using different multimedia during intentional communication of the meaning of lexical items (e.g. Duquette, Renié, & Laurier, 1998; Dubois & Vial, 2000; Fagehi, 2013; Kaplan-Rakowski & Loranc-Paszylk, 2019; Kim, 2019; Kim & Gilman, 2008; Rimrott, 2010).

The majority of previous studies on multimedia vocabulary CALL programs have investigated incidental vocabulary acquisition. A comparatively smaller body of research on multimedia vocabulary CALL has been conducted to support intentional learning. For example, Duquette, Renié and Laurier (1998) taught English students 30 French words as a second language under three conditions. The first group was exposed to a multimedia interactive package called *Vi-Conte*, which provides images, sounds, and written text. The second group learned through a video, and a third control group learned in a traditional manner. The results revealed that the multiple appearance of L2 words in different multimedia presented animated and still images supported by texts stimulating lexical learning. Dubois and Vial (2000) taught Russian vocabulary to 60 French learners in their first to sixth year at the University of Grenoble, and discovered that co-referencing of the different annotations of a new word (i.e. the sound, image and written form) enhanced memorization considerably. For the study, learners who had zero knowledge of Russian were divided into four experimental groups (see figure 2.4.). Students in the first group (P1) were presented with (L2-L1) words. An illustration was added in the case of P2, while a pronounced example sentence and a written sentence were added for P3 and P4 respectively. All the learners could hear the word in each of the experimental groups. The results showed that groups P3 and P4 were significantly better at word memorization than P1 and P2. The researchers explained these results by stating that the inclusion of extra information about a word (for P3 and P4) reduced the cognitive load. For P2, the presentation of an image possibly divided the learner's attention, overloading their memory by requiring extra semantic processing.

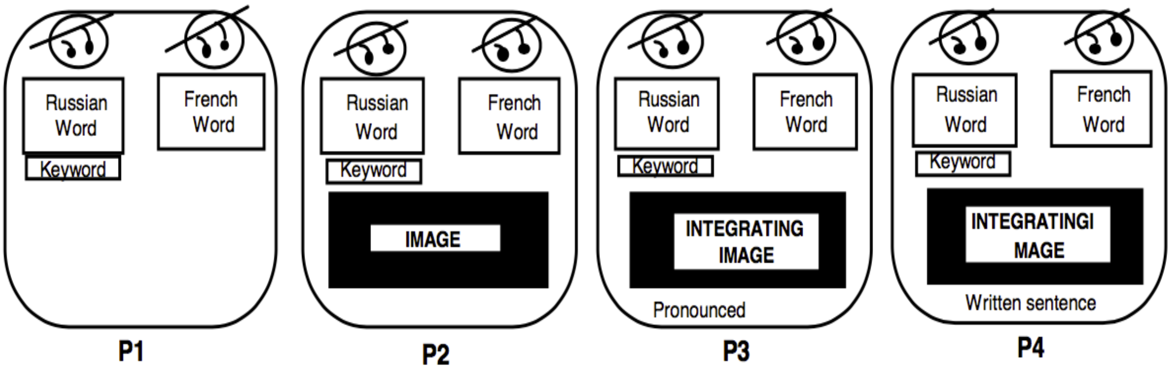
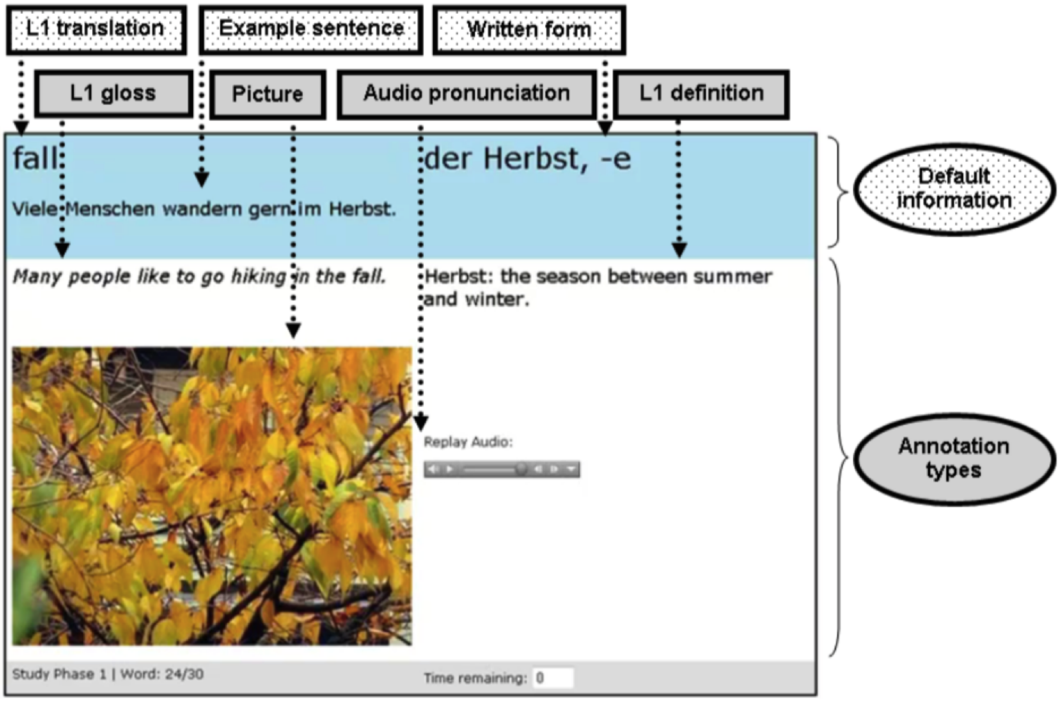


Figure 2.4. Material used in Dubois and Vial's experiment. (Source: Dubois & Vial, 2000, p.161)

Kim and Gilman (2008) investigated the use of different multimedia components when teaching English vocabulary to Korean students. In a web-based program, the researchers compared visual texts, spoken texts, and graphics when communicating the meaning of target words. The study was designed to explore the effects of six methods of multimedia instruction on a Web-based self-instructional software: visual text (Group A), visual text and added spoken text (Group B), visual text, and added graphics (Group C), visual text, added graphics, and added spoken text (Group D), reduced visual text, and added spoken text (Group E), and reduced visual text, added graphics, and added spoken text (Group F). 172 14-year old students participated in the study, completing a pre-test, a post-test, a retention test, and an attitude inventory. The results of the study revealed findings that were consistent with the multimedia principle of the cognitive theory of multimedia learning. That is, adding a picture to a textual item leads to more effective acquisition of a word. However, one distinct criticism of this study was that the participants were not separated equally into the six groups (e.g. 43 students in group A and 22 students in group B) because of lack of audio facilities for use with the computers in the classroom. Other limitations affected the design of the software for this study, sharing features in common with the following study, as will be discussed below.

A further example of studies that employ multimedia to represent the meaning of vocabulary are intentional language-focused learning programs described by Rimrott (2010). In her *Voka* (a flash card multimedia software), the researcher used different multimedia combinations to teach German vocabulary to L2 learners. 72 participants were separated into five groups and either received a picture and a gloss of example sentences (PG), a definition and a gloss (DG), a picture and audio pronunciation (PA), a definition and audio (DA), or a picture, audio, gloss and definition (PAGD), as shown in Figure 2.5. In addition to the specific multimedia combinations for each group, all the participants received a translation and example sentence including the word. Results of the immediate post-test showed that learners in groups provided with pictorial annotations significantly outperformed the learners in the other groups. However, these differences were no longer apparent on the delayed post-test, which revealed an equal effect from all types of multimedia combinations. Although Rimrott's study drew on cognitive theories of multimedia learning, as mentioned above, it failed to explain the reduced effect of the pictorial representations used in Voka over time.



fall

der Herbst, -e

Viele Menschen wandern gern im Herbst.

Many people like to go hiking in the fall.

Herbst: the season between summer and winter.

Replay Audio:

Study Phase 1 | Word: 24/30

Time remaining: 0

Figure 2.5. Labelled Voka flashcard for the target word Herbst (fall). (Source: Rimrott, 2010, p. 54)

Figure 2.5. presents a screenshot of a flashcard in Voka, depicting all the annotations associated with the German word 'Herbst', which means 'fall'. One striking weakness of the software is that it fails to take multimedia design principles into account. This is evident in the disrupted nature of the elements on the screen, as well as the random variations in the size and type of the font. Moreover, when so much information is presented randomly in this way, there is a higher likelihood that it could result in the user experiencing cognitive overload. Consequently, the learner could avoid reading much of the information presented in small-size texts by focusing their attention on the large photo, the item most likely to attract their attention. This could explain the temporary effect of the pictorial annotations found in this study.

The representation of all the information in the display, as the researcher explained, was to consider the interaction of form, meaning, and the aspects of word knowledge used. Although this is a recommended procedure when teaching vocabulary, as suggested by linguists, such as Nation (2001), it might have been improved by splitting the representation into different slides, rather than challenging the learner's working memory by gathering all the information together on one slide. In a more recent study, Fagehi (2013) found that conveying the meaning of technical terms when using text and video proved significantly more effective for second language learners than using texts only.

In incidental vocabulary learning environments, a number of researchers examined vocabulary annotations when reading second language texts (Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Plass et al., 1998, 2003; Rouhi & Mohebbi, 2013; Türk & Erçetin, 2014; Yanguas, 2009; Yeh and wang, 2003; Yoshii, 2006). Other researchers (e.g. Çakmak & Erçetin, 2018; Jones & Plass, 2002; Jones, 2009; Mohsen, 2016) investigated the effect of different vocabulary annotations on vocabulary acquisition when listening to passages in the target language. One of the striking findings of these studies is that communicating the meaning of a word through verbal and pictorial annotations leads to more effective learning than presenting either one of the two annotations; i.e. findings that have often been related to Paivio's (1991) dual-coding theory. For instance, Chun and Plass (1996) conducted a study using *CyberBuch*, a hypermedia application for reading German texts, and compared the effects of different annotations for acquiring words during reading. 160 university student participants were divided into three experimental groups. Learners in the first group received a definition of the word. In the second group, the learners received a definition and image of the word. In the last group, the learners received a definition and a short video. The results revealed that learners who were given a definition with a picture significantly outperformed the learners in the groups exposed to video and definition or definition only.

In their study conducted in a computer reading environment, Yoshii and Flaitz (2002) confirmed the results of a non-CALL study conducted by Kost, Foss, and Lenzini (1999), which showed significant discrepancies between picture and text annotations. This combination was better than the picture only or text only glosses for supporting L2 incidental vocabulary retention. 151 ESL students read a text of 392 words and received 20 word glosses of text only, picture only, or text and picture annotations. The results of the immediate and delayed vocabulary tests, which consisted of both production and recognition tasks, revealed that performance was best when learners received a combination of picture and text glosses.

Plass, Chun, Mayer, and Leutner (1998) aimed to answer the question "For whom is multimedia instruction effective?" by looking at individual differences between being a verbalizer or visualizer, based on the participant's choices during their interactions with multimedia representations. They presented learners with an English language story to teach them L2 words through translations, spoken words, pictures, and video clips. In their study of 103 participants, reading comprehension of passages was found to improve if new vocabulary was annotated with both verbal input and images, rather than with only one of these. Plass et al. (1998) obtained results supporting multiple annotated representations of learning efficacy, as they found significant effects on vocabulary acquisition when presenting multiple annotations. Similar results were found by Plass, Chun, Mayer, & Leutner (2003). English-speaking college students (152 participants) enrolled in a German reading course. Participants received either no annotation, written annotations, pictorial annotations, or both when reading a story presented via a multimedia program. Learners in the multiple annotation group (written and pictorial) achieved higher scores than those from any of the other groups. Yeh and Wang (2003) also explored the preference for these type of vocabulary annotations within the context of vocabulary learning when reading an English passage. 83 EFL college students in Taiwan were presented with a passage about Thanksgiving in courseware with colour-coded novel words. When the learner clicks on a word, a pop-up window shows either text only, text and a picture, or text and a picture with audio annotation. The results of the study revealed that the presence of the text along with a picture was the most effective.

In a within-subject design, Al-Saghayer (2001) investigated the effects of different vocabulary annotations presented in a written passage within an interactive multimedia program. ESL college students at the University of Pittsburgh (30 participants) read an English narrative text with annotations for target words developed according to different modes: text, graphics, video and sound. In an immediate post-test, he found that the use of video clips proved more effective for vocabulary learning than a still picture.

Yoshii (2006) conducted a study in which 195 students from two universities in Japan participated when learning English vocabulary. The learners were presented with glosses of unfamiliar words when reading a passage within a multimedia learning environment. Different annotations were presented within each condition: a translation, a definition, a picture and a translation, or a picture and a definition. In the productive recall test, learners from the groups involving multiple annotations outperformed those in the single annotation groups. However, there were no significant differences in the productive post-test between the learners' scores and the other groups.

Jones's (2009) study built on research that looked into the effects of learners' cognitive abilities on L2 reading and listening abilities, including vocabulary acquisition in multimedia environments (Chun & Plass, 1996; Plass et al., 1998, 2003). In particular, Jones investigated the effects of spatial and verbal abilities in the context of vocabulary learning within a listening task. 171 French English-speaking students participated in the study, and were categorized into either high spatial ability (87 students) or low spatial ability (84), and either high verbal ability (92), or low verbal ability (79), based on the score from the Kit of Factor- Referenced Cognitive Tests (Harman, Ekstrom, & French, 1976). Software developed using Adobe Premiere 4.2 (Adobe, 1994) and Authorware 4.0 (Macromedia, 1997) was used to present 27 key-words and different parts of speech during a listening task. The learners were then separated into four groups: In group 1, the students could see the L2 key word and could only hear it if they clicked on the L2 key word (no annotations). Similarly, in group 2, the students could see the L1 translation, in addition to viewing the L2 written and spoken form (written annotations). In group 3, the students could view a picture annotation (pictorial annotations). In group 4, the students could see a target word, in addition to the ability to view any of all the previous annotations (pictorial and written annotations) (see figure 2.6.).

Figure 2.6 Screenshot of French listening comprehension materials: pictorial and written annotations treatment. (Source: Jones, 2009, p.273)

A pre-test, an immediate post-test, and a delayed post-test in vocabulary recognition and recall protocol comprehension were administered to students. The vocabulary recognition immediate post-test results revealed significant effects on vocabulary learning from pictorial annotations, with no significant effects in terms of spatial abilities. With regard to verbal abilities, high verbal ability learners within the second and third treatments outperformed low verbal ability learners within the immediate post-test, when only pictorial or written annotations alone were presented. However, low verbal learners within the fourth treatment, in which they could view all the annotations, outperformed high verbal ability learners. Among the high verbal ability learners, group 1 learners with no annotations performed significantly less well than learners from the other groups. Among the low verbal ability learners, group 4 learners who were able to view all the annotations outperformed those learners within the other treatments. A delayed vocabulary recognition test revealed equivalent results. The researcher concluded the study by relating the results to the cognitive theory of multimedia learning; however, no clear or full explanation of why spatial abilities did not interact with multimedia annotations appeared in the vocabulary recognition tests.

More recently, Çakmak and Erçetin (2018) reported no effect from different types of multimedia glosses on facilitating recognition and producing vocabulary while listening to a story using mobile phones. Mohsen (2016), in an incidental multimedia listening learning environment, found that learners performed slightly better in the delayed vocabulary recall test when provided with captions more than transcripts during the learning phase. This included the presentation of animations and annotations. Türk and Erçetin (2014) fond that presenting both pictorial and verbal annotations in the glosses led to better L2 incidental vocabulary learning than when learners were given the option to access other types of glosses (verbal or pictorial). Yanguas' (2009) study results revealed that recognition of target L2 words was significantly better when using multimedia glosses incidentally to communicate the meaning of words visually and verbally, as compared with presenting only one medium. This led to more effective reading comprehension. Rouhi and Mohebbi (2013) also found a positive effect from multimedia glosses on L2 vocabulary learning, with both high and low spatial ability learners. Sun and Dong (2004) argued that L2 vocabulary multimedia tools should be accompanied by additional learning support, such as sentence-level translation or target warming-up, to provide greater efficiency to young learners.

Although a variety of multimedia technologies have been produced to date, previous research in the field of L2 multimedia vocabulary CALL has been largely restricted to limited comparisons of the effects of different annotations, such as texts, audio, pictures and videos. In addition, the majority of these vocabulary CALL studies have failed to effectively draw on SLA theory, and only a limited amount of research used multimedia CALL environments to operationalize and test SLA theory. The inability to link multimedia vocabulary CALL with SLA and memory theory was also observed in a more specific research area. Handley (2014) conducted a comprehensive examination and updated systematic reviews of research on vocabulary CALL for primary and secondary ESL and EFL. She found that of all the research published between 2004 and 2013, only six studies drew on SLA theory and current research at that time (Handley, 2014). With regard to studies that focused on introducing vocabulary through different modalities as intentional language-focused learning, Handley found only one study (i.e. Kim and Gilman (2008)) related to SLA theory, and Mayer's (2005) cognitive theory of multimedia learning.

The cognitive theory of multimedia learning is the theory that unites cognitive load theory (Sweller, Van Merrienboer, & Paas, 1998) with dual coding theory (Paivio, 1991). These theories are the most striking SLA theories that designers of vocabulary CALL draw upon when designing instructional software. They are used in their studies when explaining and discussing their results. Therefore, it is important to briefly revise these three theories, adapting Baddeley's (1986) M-WM model to provide a theoretical foundation.

M-WM model's assumption about the separation between verbal memory, which is manipulated in the phonological loop, and visuospatial memory, processed in the visuospatial sketchpad, has been incorporated into both Cognitive Load Theory and a Cognitive Theory of Multimedia Learning. The latter relied more heavily on reflection than on the idea of separation. Mayer (2005), in his Cognitive Theory of Multimedia Learning, assumes that humans have two channels (verbal and pictorial) to process information. It is vital to note that Mayer considers written text as a visual representation processed along a pictorial channel, whereas only auditory texts are manipulated in the verbal channel. This is not in line with Baddeley (1986), who states that both auditory and written texts are processed in the phonological loop. The second assumption of the M-WM model concerns the limited span of the working memory subcomponents. This assumption is also apparent in the Cognitive Theory of Multimedia Learning, which assumes that a limited amount of information can be handled in each channel. While the two assumptions were highlighted by the Cognitive Theory of Multimedia Learning, the Cognitive Load Theory placed greater weight on the limited capacity of Baddeley's model than on the dual coding assumption. Instead of considering the distinction between verbal and visuospatial channels, Cognitive Load Theory emphasized the cognitive load that might overcome the unitary construct resources that inform intrinsic load; as a result of the learner's prior knowledge of extraneous load as it emerges from one's surroundings (Sweller et al., 1998). Both of these intrinsic and extraneous types of load can generate a cumulative load, which might surpass the span of working memory, and thus hinder the learning process.

Cognitive load theory

Based on the notion of the limited capacity of both visual and auditory channels, Sweller (2002) proposed a cognitive load theory, suggesting that information presented via instructional design must be guided in a manner that does not impose a heavy load on working memory. Following the notion that working memory load should be decreased to encourage better schema construction, cognitive load theory aimed to offer guidelines that facilitate information processing (Sweller, Van Merrienboer, & Pass, 1998). According to this theory, cognitive load could be intrinsic (by the intrinsic nature of the material) or extraneous (by the manner of presenting the material), and is determined by the instructional design (Sweller et. al, 1998). Sweller argued that since many of the instructional procedures and designs used were not designed with respect to working memory limitations, they are insufficient (Sweller et al,, 1998).

Cognitive theory of Multimedia learning

The cognitive theory of multimedia learning combines cognitive load theory and dual coding theory, maintaining that three cognitive science principles form the basis of the theory: dual-channel assumption (pictorial and verbal processing of information); limited capacity assumption (both channels have limited span); and active processing assumption (coordinating sets of processes) (Mayer, 2005). Thus, when a multimedia lesson is presented to a learner (figure 2.7.), they receive pictorial and verbal information in their cognitive system, through visual and the auditory modalities.

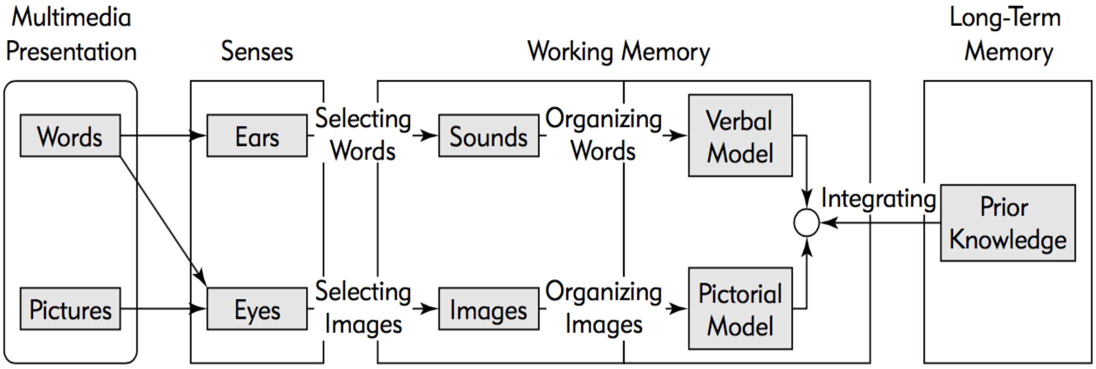


Figure 2.7. Cognitive theory of multimedia learning. (Source: Mayer, 2005, p.37)

When a learner pays attention to the materials shown, they enter their working memory to access additional processing. There, the learner mentally organizes the materials into either pictorial models or verbal models. Finally, the materials will be integrated with existing knowledge, gleaned from the long-term memory. Engaging in such cognitive processes, multimedia environments can result in effective learning (Clark & Mayer, 2011).

Dual-coding theory

Paivio (1991) theorized that there is a separation between visual and audio interpretation in our minds. The first of these two cognitive representations is termed imagens, and these are visuals that are close to their physical representations in reality. Thus, when we see a cat in real life or in a picture, our minds see a four-legged animal. The second, on the other hand, is Logogens, which are more abstract. These are the words and symbols we use daily. Thus, when we see the word cat, we interpret this to mean a four-legged animal, even though the word itself is simply a series of symbols. According to Paivio (1991), information as it is represented in the mind is organized by verbal and pictorial mental codes. He added that mental images are analogous to the objects they are representing, whereas words are represented in symbolic codes and these can be any arbitrarily written forms denoting things other than themselves. For example, the number 5 represents the concept *fiveness*, but has no physical qualities that suggest this meaning. The assumption that there are different channels for processing verbal and visual information was studied by Paivio (1991), who found participants processed verbal and pictorial information differently, recollecting sequences of words more readily than a sequence of pictures. However, the participants recalled more pictures when allowed to do so.

The researchers found evidence, in line with the dual-coding theory, suggesting the possibility that there are two different systems for processing verbal and pictorial information. For example, Brooks (1968) designed two main tasks for his participants; one verbal and one pictorial. To interact with these tasks, Brooks asked participants to respond verbally by saying yes or no, physically by waving their hands, or mentally by simply tapping with the right hand or left hand for correct and incorrect responses, respectively. The researcher then measured the interference caused by any of the three response types with the two main tasks, by looking at response speed. He found the verbal task interfered with the verbal response (saying yes or no), but not with other response mediums (i.e. either visual or manual). Moreover, the pictorial task created more interference for the participants when they needed to respond physically. The findings suggest two distinct channels for mental representations of knowledge using analogical codes of pictorial objects and symbolic codes for verbal stimuli. The dual coding effect on memory has been identified in other experiments (e.g. Paivio, 1975; Paivio & Lambert, 1981). The learning process proposed in this theory is consistent with the historical trend towards knowledge construction through pictures and imagery (Paivio, 2006). In congruence with what Paivio posited in his theory, Nation (2010) asserted that connecting the meaning of lexical items to real objects directly, such as visual aids, is a strategy used widely by beginners. He also viewed it as a procedure helping learners to understand words, and at the same time work as a cue for remembering them. Such activities had a greater effect if complemented by, for instance, a verbal definition. Such a situation would not only decrease the possibility of wrong guessing, but also result in ‘dual encoding’, i.e. visual and linguistic storing of information (Nation, 2010).

In sum, cognitive theories of learning see second language acquisition as a cognitive skill that is acquired by the engagement of different cognitive systems like perception, memory and information processing. Just as when acquiring other skills, language learning process have to be designed to overcome the mental capacity limits that might hinder performance (Ellis, 2000).

Overall, from the aforementioned discussion in the literature, and drawing on findings from the multimedia vocabulary CALL experimental studies mentioned, it can be concluded that the majority of previous studies focused on assessing the impact of technology on strengthening the connection between the form of a word and its meaning (mainly comparing between different multimedia annotations). Additionally, the majority of intentional language-focused vocabulary learning interventions had failed to effectively employ multimedia to present and practice other aspects of word knowledge beyond form-meaning connections. Moreover, it can be concluded from this review that multimedia vocabulary CALL software had not been efficiently employed when operationalizing SLA cognitive theories. The majority of previous studies investigating the use of multimedia to communicate L2 word meaning via different modes, whether intentionally or incidentally, did not relate their CALL design persuasively to memory theories. For instance, assumptions about a visual-spatial distinction were not considered in previous vocabulary CALL studies. Thus, the CAVA software used in the current study will draw on Baddeley’s most recent model (2012) of working memory, to better acknowledge and understand the effects of CALL research on vocabulary learning. In order to investigate the role of the visuospatial sketchpad, the zoomable interface will be designed and used for the first time in the field of vocabulary CALL. The zoomable interface will provide the learner with an ability to zoom in and out over a slide, showing a thematic set of words associated visuospatially creating an overview of the environment (context) in which they are normally used.

2.1.3 The potential added value of the Zoomable User Interface (ZUI)

The design elements of ZUIs are commonly applied in computers (e.g. Apple's IOS, google maps, prezi, and emaze), since the first zoomable sketchpad program was developed by Ivan Sutherland in 1962 (Sutherland, 1963). In the 1970s, the Architecture Machine Group at Massachusetts Institute of Technology (MIT) developed a more general zoomable interface in a media room, which later influenced the Apple Mac (Bolt, 1979) (see figure 2.8.).

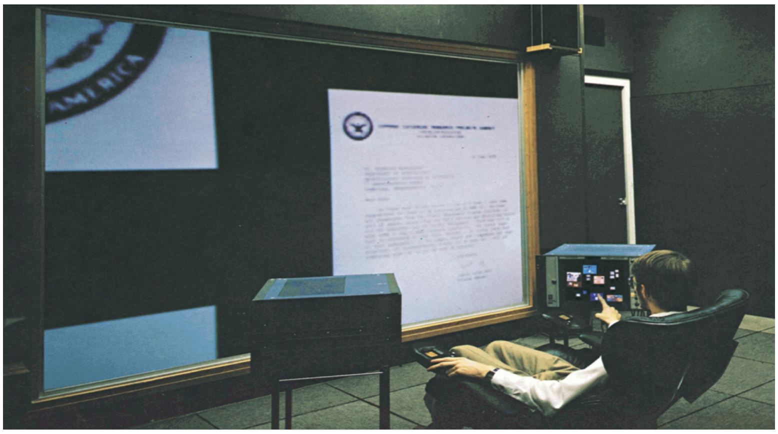


Figure 2.8. Dataland and a spatial data-management system. (Source: Bolt, 1979, p.11)

Surprisingly, this piece of extraordinarily popular technology has yet to be employed or examined in vocabulary CALL environments yet. The potential added value for ZUI, in comparison with GUI and MDI, in current intentional multimedia L2 vocabulary learning software can be linked to the number of theoretically grounded features that rationalize its design.

To attain a more precise comparison of CAVA's three multimedia representations, it is essential to address the difference between words and pictures when used to communicate knowledge from the outside world, within educational software, as is the case in the current study. In addition, it is necessary to ask if we store this knowledge in the form of images, or both words and images, or some other format. The CAVA conditions were designed in ways that might influence how L2 knowledge is stored and represented in learners' minds. Before looking more closely at the coding elements of the three CAVA interfaces as external representations communicating L2 word meaning, we need to briefly consider different representations of mental knowledge, and the form they are thought to be encoded in cognitively, as well as how they are retrieved.

If someone wants to describe a photo of a person or an object, they first need to imagine how that person or thing looks, and perhaps also sounds. Obviously, the person or object described does not actually physically exist in the mind. In order to successfully describe them, one has to use a type of previously encoded mental representation, conveying some information about them. These mental *knowledge representations*, as cognitive psychologists call them, are what we recall when imagining and describing the things that surround us. Imagery, which is the mental representation of an object or a thing, is no longer sensed by sense organs (Moulton & Kosslyn, 2009). Generally speaking, mental imagery not only involves visual cues, but also mental representations of the other encoded senses, such as auditory information, touch, smell, or taste. For example, it is possible to imagine the taste of fried eggs or the sound of your doorbell. However, cognitive psychologists have long focused their research on visual imagery, as the majority of people use visual images more frequently than other forms (Kosslyn, Seger, Pani, & Hillger, 1990). Visual mental imagery can be used to answer questions and resolve problems by visualizing a setting, and then creating and manipulating a mental image of that setting (Kosslyn & Rabin, 1999; Kosslyn, Thompson, & Ganis, 2006). Arguably, the mental image is represented in the mind delivering a similar copy of an external physical image (Kosslyn, 2006).

The functional equivalence hypothesis supports the mental imagery theory for representing knowledge of concrete objects. It proposes that mental images of concrete objects, although formed inside the brain, are functionally similar to what we perceive through our sense organs. Thus, mental imagery and visual perceptions are strongly analogous, although not identical (Farah, 1988; Finke, 1989; Jolicoeur & Kosslyn, 1985; Rumelhart & Norman, 1988). One of the important principles of visual imagery was suggested by Finke (1989), and taken into account when designing ZUI. This concerns representations of spatial information, such as location, and how they relate to mental visual imagery. The principle states that, “the spatial arrangement of the elements of a mental image corresponds to the way objects or their parts are arranged on actual physical surfaces or in an actual physical space” (1989, p.61). Therefore, mental imagery is not only effective as a means of retrieving memories about the physical characteristics of objects that we perceive, but it also recreates spatial relationships among objects, even where implicitly encoded (Galotti, 2017).

Another important hypothesis related to the current research is the functional equivalence hypothesis, which underlies the idea of zooming in on our mental images as with our representations of precepts. For instance, if one examines a large bookcase with many books of different sizes, it is necessary to zoom in and out to be able to read the titles of the books if they have different font sizes. The image scaling of ZUI could work as a supportive factor to facilitate the encoding and manipulating of mental imagery relating to the overview picture presented to the learner.

The ability to mentally zoom in and out of a mental image is supported in previous research. As we respond rapidly to questions about larger perceived objects rather than small ones, keeping in mind the functional equivalent between perception and mental imagery, researchers were able to ask participants similar questions about the attributes of mentally imagined objects. They found they do answer more quickly if the mental image is large. Moreover, if a mental image included small and large objects (a rabbit paired with an elephant), it takes the learner longer to answer questions about the rabbit than the elephant (Kosslyn, 1975). In another study, Kosslyn (1976) asked participants to answer questions about the physical attributes of an imaginary cat. He found that they answered more quickly when the question was about a large part of the body (Does the cat have a head?) than when the question was about small parts (Does the cat have claws?). Different results emerged when the participants were asked to use their verbal-propositional knowledge about the cat to answer the same questions. No differences were associated with the large and small parts of the cat; however, differences in response time emerged based on the distinctiveness of the features of the cat as an animal. The participants responded faster when answering questions about claws (which distinguished the cat from other animals), but were slower to answer questions about undistinctive characteristics such as the head. However, representations of knowledge cannot be measured or observed applying direct methods. Alternative cognitive psychologists ask people what they can see, which is an approach that is considered highly unreliable (Pinker, 1985). They might also ask participants to deduce logically how knowledge is represented in the mind using a rationalist approach based on empirical data taken from one of two main sources; i.e. neuropsychological studies and standard laboratory experiments.

After briefly considering internal mental representations, the next section will look at CAVA conditions as external representations, and their potential role in creating the internal representations of the L2 word knowledge. The potential added value for ZUI over GUI and MDI can be predicted, due to its richer and more advanced design. To clearly measure this effect, and to minimize extraneous differences between the three conditions of CAVA (MDI, GUI, and ZUI), all the target L2 words given were concrete nouns. This avoided the potential confounders from abstract words, as concrete nouns are more easily represented pictorially. For example, to answer the question: What is the shape of the house?, it might be easier to draw it than to describe it in words. However, using words to communicate an idea might convey the distinctive characteristics of the represented object. Neither a picture of a mouse, for instance, nor the word *mouse* communicates all mouse characteristics. Thus, three CAVA conditions were designed to deliver different distinctive characteristics associated with the target L2 word, with the expectation of a greater effect from the new multimedia design ZUI over MDI and GUI.

MDI presents the word *mouse* as a translation equivalent, to communicate the meaning of the L2 word meaning. This representation is merely arbitrary, meaning the word *mouse* offers a symbolic representation with no other information to relate it to the real-world object it is representing. In addition, this symbolic representation differs across languages. In contrast with the picture, any missed part of the word renders it impossible to grasp the concept MOUSE from the constituent parts. Also, rules must be followed when reading the word, from left to right; for example, following a fixed order of letters. The MDI could transmit additional abstract information about target concepts; i.e. information, such as class membership (mice are mammals), might be captured by showing the word *mouse*. However, learning a language involving one rule for reading from a translation equivalent and a different rule for reading (RL from a LR language) might well prove more challenging.

GUI, by contrast, is designed to present the concrete attributes of the target concept, such as shapes and colours, which are an analogous representation of real-world objects. The features and characteristics of the mouse can be observed by looking at a picture of a mouse with parts covered, the parts of the picture that remain visible may be sufficient to identify it as a picture of a mouse. The GUI representation is distinctive in that a picture of the target word is presented in isolation, apart from other related objects that might cause interference. GUI presents a high definition picture with easily distinguishable parts. There are no restrictions on how the object is viewed; from the top to bottom, right to left, or however it pleases the learner. These traits could prolong the GUI effect, and the encoding of a clear visual stimulus along with the Arabic written form as a verbal stimulus might create an efficient dual coding learning process.

Finally, the ZUI representation is designed to communicate additional aspects of target L2 word knowledge. Showing the whole view of an office, for instance, will represent the spatial relationship between the chair and the desk. This spatial (positional) relationship (e.g. above, behind, below, beside) is concretely represented in ZUI. These kinds of attributes, as a distinctive characteristic of ZUI, are similar to what the learner would observe in real-life situations. Most of the aspects of the view presented can be repeatedly grasped individually for each object by zooming in for a closer look, or as a holistic overview when zooming out to see the big picture. ZUI can serve as a platform to view layouts that can stimulate the human spatial memory. In addition, the learner will be able to view words together in a single layout in which the relationship between the words can be seen and associations easily made between items. This could facilitate building a spatial model to describe the information, as an internal mental representation.

By writing words in close proximity to their related objects, ZUI in this study can be exploited to employ the spatial contiguity principle, as proposed by Moreno and Mayer (1999), who found that participants learned better when texts and their related visuals were physically close on the computer screen. Additionally, through constant interaction, and zooming in and out of the labelled items and reading and listening to them, a *spatial memory* of where things are arises. It can be hypothesized that through this interaction with ZUI, a mental image of the layout view will appear in the mind. A script for word retrieval becomes encoded in the space where the item is located. This linking of the word with its location could serve as an extra memory cue. This goal is more likely to be achieved with multimedia lessons containing both words, corresponding pictures, and locations that work together to explain the to-be-learned content.

ZUIs are engaging (Ware, 2004), and all learning requires behavioural and psychological engagement. *Psychological engagement* describes the cognitive processing of content in ways that can lead to the acquisition of new knowledge and skills. Some cognitive processes that lead to learning include paying attention to relevant material, and mentally organizing it so that it is coherently represented (Clark & Mayer, 2011). Clark and Mayer (2011) found that adding relevant graphics to words can be a powerful way to help learners engage in active learning. Moreover, a well-known psychological principle is that organized information is easier to learn than unorganized information (Schmitt, 2007). This finding would suggest the benefits of using ZUIs to present an overview that groups target words together visuospatially during learning. Bederson (2011) stated that users of ZUIs may be more engaged, and are more likely to remember the spatial structure of content better. In addition, there is the hope that animated transitions are understood preconsciously by the human visual system, and thus may not in reality place a significant load on STM (Bederson, 2011). However, relatively few studies have looked at the value of animated zooming more generally. Indeed, there is the potential that ZUIs tax human short-term memory (STM), because users must integrate the spatial layout of information in their heads. Retrieving knowledge representations might involve watching a videotape of continuing representations of knowledge in the mind. It is probable that ZUI, with its ongoing display of spatially related items, might create a similar kind of *videotape*.

The essential function that a ZUI aims to offer in the present study is a fundamental one; i.e. that there is more information about word knowledge than fits on the screens of the other two interfaces (MDI and GUI) used in CAVA. In contrast to the majority of previous studies, which limit their contributions in terms of providing insight into our understanding of SLA, the present study will explore the features of the three user interfaces that make them effective, by researching features with a differential effect on vocabulary learning. Providing and testing CALL environments contributes to relating CALL research to SLA theory and research that can develop our understanding of SLA.

There is thus a need to address various aspects of word knowledge beyond form and meaning links when designing vocabulary CALL programs. The current research employed technology to teach multiple significant levels of word knowledge (i.e. word association), with the potential to create additional relationships between words. This leads to more rapid and effective storage and retrieval of lexis, by enhancing the encoding and retrieval of words and promoting the development of lexical networks. This depth of processing can then result in faster and more effective storage and retrieval of lexis.

Depth of processing

The depth of processing hypothesis, as proposed by Craik and Lockhart (1972), refers to the degree of effort placed on analysis of a stimulus, stating that its retention as a memory trace depends on the degree to which its processing has been enriched during encoding; i.e. deeper semantic analysis yields richer and more elaborate encoding of a memory trace. This extensive elaboration supports improved retention as follows. Firstly, the elaborated trace is differentiated from previous memory traces, ensuring it is distinctive and easier to retrieve. Secondly, deeper processing leads to the organization of an elaborated trace, followed by its integration into the previous knowledge in an arranged structure that facilitates the retrieval process (Moscovitch & Craik, 1976). It should be noted that, in order to ensure improved retention, deep processing must be implemented on the stimulus, not only at shallow levels (i.e. analysis of colour, surface form, and vividness), but also to enrich it meaningfully in a broader sense (i.e. analysis of inference, meaning, implication, and interpretation) (Craik & Watkins, 1973; Craik, 2002; Lockhart & Craik, 1990). Goldstein and Chance (1971) found that, in order to differentiate snowflake patterns in a meaningful way, learners required semantic knowledge of such patterns. Thus, despite each pattern being unique, learners remembered meaningfully distinctive patterns more accurately. In addition, the degree of attention paid to a stimulus adds a further factor when elaborating on depth (Craik & Lockhart, 1972).

The depth of processing hypothesis is relevant to the current study, as it is applied by CAVA software in its third condition (i.e. ZUI). This ensures that learners focus, and undertake deeper processing, in order to analyse additional aspects of a word as displayed on the layout, i.e. in comparison to the L2-picture representation (GUI). In addition, it is expected that the smallest degree of attention and processing is likely to be devoted to L2-L1 in the MDI condition.

Word knowledge

The three CAVA conditions explicitly teach different aspects of word knowledge. Prior to discussing these in more detail, it is first vital to examine issues surrounding the knowledge of a word. According to Nation (2001), the depth of word processing (i.e. images of the word, elaboration, or deliberately making inferences) creates a strong explicit awareness of the meaning of a word. Nation (2001) further noted that word meaning is one of nine aspects involved in the knowledge of a word (Nation, 2001). Thus, the learner knows the spoken form of the word when they are able to firstly, recognize it when spoken and secondly, vocalize it to produce a meaning. Learners also need to be able to correctly pronounce each sound within the word, placing clear degrees of stress on the appropriate syllables. These sounds are combined according to underlying rules, known as phonotactic grammaticality. Nation (2001) also considered spelling a key feature used to gain familiarity with a word's written form; noting that this can be improved by early training and playing with rhymes. In addition, learning can be easier if a word is comprised of known parts, due to stems and affixes being potentially familiar to a learner from his/her first language or other second language vocabulary. Nation (2001) further noted that general word knowledge includes understanding that certain aspects can be employed in other words, as well as that familiarity with members of a word family might serve to suggest knowledge of a word and be a sign of proficiency. In addition to acquiring a word's meaning and appearance (i.e. its written form), as well as its sound (i.e. its spoken form), the learner needs to be able to combine these pieces of information (Nation, 2001).

A dictionary can list different meanings, particularly for high frequency words. Thus, knowing the dictionary entries of a single word forms a notable aspect of word knowledge, as well as recognizing it as part of speech, in order to use it in accordance with appropriate grammatical patterns. A further aspect of knowing a word is familiarity with the vocabulary alongside which it generally appears; i.e. *I ate some fast food*, rather than *speedy food* or *quick food*. Pawley and Syder (1983) argued that a native speaker becomes fluent by storing a large number of memorized sentences in the brain, because producing memorized sequences is easier than engaging in spontaneous construction during speech. An additional aspect of the knowledge of a word relates to constraints concerning its use under specific circumstances (Nation, 2013).

The teaching of different aspects of word knowledge across the three multimedia CALL interventions in this study occurred during the earliest stages of learning, i.e. when learners had not yet connected the meaning of the word with its form, due to a lack of familiarity with the target language. This ensured that learners' lack of experience with Arabic assisted in their exploration of how different multimedia representations might result in differing levels of L2 processing.

2.1.4 Summary

Since learner's internal cognitive tools collaborate with instructional technology when constructing knowledge (Jonassen, 1992), we should concern ourselves with developing technological tools that display activities that facilitate cognitive processes, to achieve better language learning outcomes. In accordance with Clark and Mayer (2011), I believe that new technologies need to be compatible with our cognitive learning processes to support learning rather than hinder it. Instructional designers should not become so concerned about advanced technology features that they ignore our mental abilities. Clark and Mayer (2011) suggested that the most crucial factor when designing an effective e-learning program is to support human cognitive learning processes.

Baddeley and Hitch's (1974) multiple working memory model provided a theoretical foundation to underpin multimedia learning design. The model was sufficiently well developed to provide a framework for the current study, enabling the researcher to generate theoretically derived hypotheses to be investigated experimentally. M-WM assumptions were supported by a range of neuropsychological and experimental findings that assisted in the formation of knowledge regarding second language acquisition. However, not all M-WM components were investigated in language learning, and previous research relied more commonly on the role of phonological loops. Most previous vocabulary CALL designs to have investigated interactions between cognitive abilities and multimedia representations failed to consider assumptions about visual-spatial distinction. The software used in previous studies was mostly designed to compare pictorial annotations, such as still pictures or videos and verbal annotations on the one hand and verbal and visual abilities on the other hand. Thus, they limit their usefulness in terms of furthering our understanding of the role of other cognitive abilities, such as spatial memory.

This is demonstrated in discussions of examples associated with multimedia studies of language-focused and meaning-focused vocabulary CALL, as detailed above.

In contrast with previous vocabulary CALL studies, the present study aims explicitly not only to employ human visual memory, but also to take advantage of spatial perception and memory. It assesses the effect of teaching the associated dimensions of a word's meaning. It aims to determine the veracity of the hypothesis that L2 learners' long and short-term vocabulary acquisition varies significantly when employing different multimedia representations (i.e. MDI, GUI, ZUI) to acquire L2 word meaning, because of the differential impact of each of the three teaching methods. To test this, there will be three exposure groups, to investigate the efficacy of three multimedia representations (i.e. verbal-based (L2-L1: MDI), visual-based (L2-Picture: GUI), and a spatial-based (L2-Context: ZUI), designed for vocabulary learning. Each of these three methods will present the participants with 24 concrete words, divided across six lessons (four words in each lesson).

These three different user-interfaces will be drawn on to present words in CAVA software, as designed by the researcher. One of the interfaces used is menu-driven (which is a widely accepted way of presenting new vocabulary items); i.e. a verbal menu showing a list of L2-L1 words. The second is a GUI, presenting an L2-picture list. The third interface is used for the first time when teaching vocabulary; it is a zoomable interface describing a thematic set of words that afford an overview of the environment (context) in which they are normally used. The research aims to identify the role of each user interface in the storage, as well as the retrieval of L2 lexis.

In addition, the research data will involve a working memory test to allow a comprehensive analysis of how novel words are learned naturally. Unlike previous vocabulary CALL studies, which investigated the effects of verbal and visual working memory spans only when perceiving different types of stimuli, this study aims to examine working memory capacity in addition to verbal and visual elements, applying the assumption that there is a separation between visual and spatial abilities. To evaluate this, the research tests whether learners with a high proficiency in working memory and short-term memory will recall more L2 words than low-ability learners (i.e. verbal-MDI, visual-GUI and visuospatial-ZUI), in both immediate and delayed post-test scenarios.

Quantitative data associated with verbal, visual and spatial cognitive abilities will be collected using an Automated Working Memory Assessment (AWMA) to achieve a clearer understanding of the role of each of the separate working memory components, and to test this for correlations with learners' vocabulary acquisition, as measured by CAVA. Corresponding to the three subsystems of the M-WM model, it can be hypothesized that high verbal learners will recall more MDI words, and high visual learners will recall more GUI words, and high visuospatial learners will recall more ZUI words. To investigate this, AWMA has been selected, as it comes with several unique subtests thought to be appropriate to assess the effect of individual differences in memory on L2 language learning. AWMA was designed by Alloway (2007), who took Baddeley's theory a step further, abandoning the CE label. Drawing on the M-WM model, Alloway divided up the working memory test to cover other components, instead of measuring verbal processes only. The result was that the latest version of AWMA included twelve subtests in its long form. Rather than using short forms of AWMA, which have fewer subtests, the long version was used here to attain more reliable results as a way to assess the participants' individual differences in working memory capacity, which would then be expected to play an essential role in language learning.

2.2 Methodology

This section will present the methodology set out in the first experimental study conducted for the current thesis. This section will be organized in the following way: First, the apparatus employed in the study (the automated working memory assessment (AWMA) and the CAVA software) will be introduced. Then, a description of the method used in the initial study will be presented. It describes the participants, the design of the experiment, the interventions undertaken in the experiments, and finally, the process of collecting, handling, and analysing the data.

2.2.1 The Automated Working Memory Assessment (AWMA)

Previous studies have generally implemented one of two approaches to measure the contribution of working memory components to learning; the capacity approach or the dual-task approach. The latter is used when the learner is required to complete more than one task while learning. Both tasks are then processed as one component (e.g. phonological loop). If interference occurs between the dual task and the learning task, it can indicate the contribution of that component to the learning process. The capacity approach, on the other hand, involves assessing the component span of different participants wishing to participate on the same learning task, comparing their learning outcomes with their individual differences and the capacity of components (Andrade, 2001). With regard to the capacity approach, some researchers (e.g. Carretti, Borella, Cornoldi, & De Beni, 2009) have distinguished between complex and simple capacity tasks. Simple tasks only measure the maximum amount of information that can be stored permanently in a single slave system; i.e. the phonological loop and the visuospatial sketchpad. The complex task not only measures the span of storage, but at the same time assesses how much information can be processed by each of the two subcomponents in addition to the central executive, which is then assumed to engage in the information manipulation process (Daneman & Carpenter, 1980). The current study followed a capacity approach, using the Automated Working Memory Assessment (AWMA) to assess participant's cognitive abilities, and then compared them with their L2 vocabulary learning outcomes.

The Automated Working Memory Assessment (AWMA), a computer-based assessment of working memory skills (Alloway, 2007; Alloway, Gathercole, & Pickering, 2006; Alloway, Gathercole, Willis, & Adams, 2004), was chosen to assess the participants' cognitive abilities in the current research. AWMA was developed in the United Kingdom to screen for significant working memory problems in individuals from childhood onwards, through to early adulthood. Minimal training is essential, as its administrative, scoring, and interpretive functions are automated. There are three forms of AWMA (i.e. screener, short form, and long form) according to the number of the subtests. The long form comprises twelve subtests, and was the version used in the current study (see table 2.1). The presentation of stimuli in all of the AWMA subtests is consistent across the participants as it is fully automated in its presentation and scoring. The AWMA was applied in the current study as it approximates to the Multicomponent Working Memory Model proposed by Baddeley and Hitch, and used in the current thesis as a theoretical framework of CAVA design. The following subsections describe the content of the subtests implemented for the AWMA, used in this study to measure participants' capacity to temporarily access and manipulate information.

Alloway (2007) required both verbal and visuospatial working memory to be measured by the AWMA separately from verbal and visuospatial short-term memory abilities. Distinguishing working memory subtests from the short-term memory counterparts was a striking feature of the AWMA, and was achieved by including additional processing tasks. These new processing tasks were designed not only to test verbal abilities, but also to assess visuospatial factors. When explaining this, Alloway et al. (2006) revised empirical studies that supported their decision to add visuospatial memory tests.

Table 2.1. Memory subtests of AWMA's long form version.

	Subtest	Memory component
1	Digit Recall	Phonological STM
2	Word Recall	Phonological STM
3	Non-word Recall	Phonological STM
4	Listening Recall	Verbal WM, Executive WM
5	Counting Recall	Verbal WM, Executive WM
6	Backwards Digit	Verbal WM, Executive WM
7	Dot Matrix	Visuospatial STM
8	Mazes Memory	Visuospatial STM
9	Block Recall	Visuospatial STM
10	Odd-One-Out	Visuospatial WM
11	Mr. X	Visuospatial WM
12	Spatial Span	Visuospatial WM

2.2.1.1 Validity Evidence Regarding Memory Components

Alloway et al. (2006) used the AWMA to measure four different components of memory, as a means to investigate the cognitive mechanisms that underpin it. Confirmatory factor analyses of the results supported Baddeley and Hitch's (1974) multiple component model of working memory. The results revealed the processing factor of working memory tasks, and were supported by the executive centre as a domain-general process. On the other hand, storage features rely on domain-specific verbal and visuospatial components. Considering these findings, the AWMA structure contains tasks intended to measure verbal and visuospatial factors, testing the systems' ability to both store and process codes.

2.2.1.2 AWMA Subtests and Tasks

In all of the tests, the instructions were presented as sound files, while the computer screen remained blank. After delivering the instructions, the participant completes practice trials to master the task. The test trials followed the practice phase. The researcher recorded the participants' responses to the stimuli by clicking the right or the left arrow keys on the keyboard for true and the false responses, respectively. A correct answer was credited by the computer with a score of one. Each test contained a block of trials. The participant was allowed to move from one block to the next one up if they responded correctly to four trials within a block of trial. However, if they could not achieve four trials in the block, the test stopped automatically, and the computer advanced to the next test. The number of correct responses for each test was reflected in the score the participants got by the end of the test.

2.2.1.3 AWMA four major components

Verbal short-term memory

There were three subtests conducted to measure verbal short-term memory. The first being the digit recall task. In this task, the participant hears a sequence of digits, and is asked to recall each sequence in order. The second subtest is the word recall task. In this task, the participant hears a sequence of words, and is then asked to recall each sequence in the correct order. The third and final subtest for verbal short-term memory is the non-word recall task. In this task the participant hears a sequence of non-words, and is asked to recall each sequence in the appropriate order.

Verbal working memory

Three different subtests were used to measure verbal working memory. The first task was listening recall. In this case, the participant hears a series of spoken sentences. T, and then has to judge if the sentence heard is ''true'' or ''false''. Lastly, they have to recall the final words of all the sentences in the sequence. The trials for the listening recall task start with one sentence, and then the number of sentences is increased until the participant becomes incapable of recalling three correct trials in a block. The second subtest measuring verbal working memory is the counting recall task. In this task the participant is presented with a visual array of blue triangles and red circles. The participant needs to give two responses after viewing the array presented. The first aim is to count the circles in the array. Secondly, to recall the tallies of the circles. This test trial beings with a single visual array, and continues by adding more visual arrays in each of the following blocks, until the participant becomes unable to remember the four trials correctly. The third subtest is a backwards digit recall task. In this task, the participant hears a sequence of spoken digits, and is then required to recall them in reverse order. The trials for this test start with two numbers. Then the numbers are increased by one number in each block. When the participant becomes incapable of recalling four trials correctly, the test ends.

Visuospatial short-term memory

The AWMA measures the visuospatial short-term memory via three subtests: all were administered in this study. The first is the dot matrix task, in which the participant is presented with the position of a red dot in a series of 4 * 4 matrices for 2 seconds. They are then asked to remember the position, and show the dot's place by pointing to the appropriate squares on a computer screen for the examiner. The number of the red dots is then increased until performance breaks down. The second subtest is a maze memory task. In this task, the participant is presented with a maze. For three seconds, a red line showing the correct path is displayed to the participant. After these three seconds, the line disappears and the participant has to trace the same path on the blank maze. The third subtest for measuring visuospatial short-term memory is the block recall task. In this task, the participant views a video of a series of blocks being tapped. He or she then has to recall the sequence in the correct order.

Visuospatial working memory

To assess visuospatial working memory span, three measures were conducted using the AWMA. Just as the verbal working memory tasks was administered, the processing and storage factors of the tasks were conducted separately and later combined. The first subtest was the odd-one-out task. In this task, the participant is shown three shapes consecutively. They are then asked to identify the odd-one-out. The shapes then disappear and the participant is asked to recall the position of the odd-one-out. The number of rows is later increased until the participant is unable to recall the location of the odd-one-out shapes in the correct order. The second task is the Mr. X task. The participant views a picture of two Mr. X figures and is then asked to judge whether blue Mr. X is holding the ball in the same hand as yellow Mr. X. Later blue Mr. X will be rotated, and the participant needs to remember the location of the ball in Mr. X's hand in sequence. The third subtest to measure visuospatial working memory is spatial span. The participant is presented with a picture of two arbitrary shapes, with a red dot on the shape on the right. The participant then has to identify whether the shape on the left is the same as or opposite to the shape on the right. As in the Mister X task, the red-dotted shape is also rotated. After each trial, the participant must recall the position of the red dots in the correct order.

By comparing the AWMA with other working memory assessments tools, Dehn (2011) found it to be the best application in terms of distinguishing between short-term memory and working memory components. Dehn added that the AWMA supports the notion that functions of working memory must be separated, rather than joined under an executive factor. Also, the AWMA is an efficient instrument for screening the working memory's ability, as it effects a large number of participants easily and sufficiently. Figure 2.9 shows the relationship between the AWMA measures and the Multicomponent model of memory divisions.

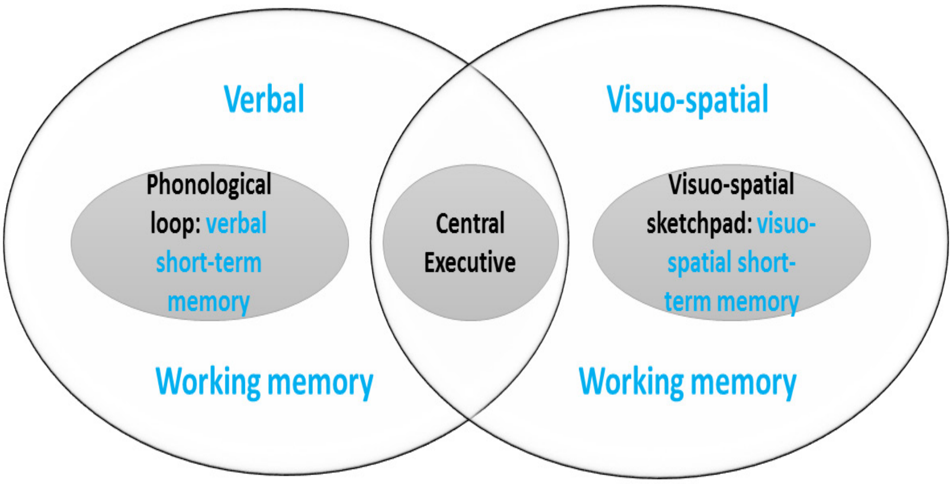


Figure 2.9 Baddeley's working memory model and its relationship to AWMA measures (in blue). (Source: Absatova, 2015, p.22)

The AWMA test is used to generate data by measuring the memory abilities of all fifty study participants. This study, as mentioned above, uses the measure to explore similarities and differences in memory scores between participants. It also seeks to ascertain whether differences in ability to acquire a second language through different modes of teaching could be explained by individual variations in memory abilities.

As mentioned in the literature review, age has been correlated with working memory abilities (Gathercole, Pickering, Knight, & Stegmann, 2004). Consequently, to ensure whether any differences in the memory abilities of the participants in this study correlated with factors other than their age, all the participants were drawn from a similar age group.

2.2.2 Computer-Assisted Vocabulary Acquisition (CAVA) software

In order to develop the CAVA software in a consistent and reliable way, it was designed according to a systematic instructional design procedure. There are many versions of instructional design models, which have been used to develop a variety of training and educational programs. Although these models have produced diverse versions of instructional design processes, they often share a similar set of phases (Gustafson & Branch, 2002). To ensure objectives are achieved, the CAVA design process was created as an efficient and effective software tool, following a widely accepted instructional design model for producing interactive video/multimedia (Bergman & Moore, 1990) (see figure 2.10).

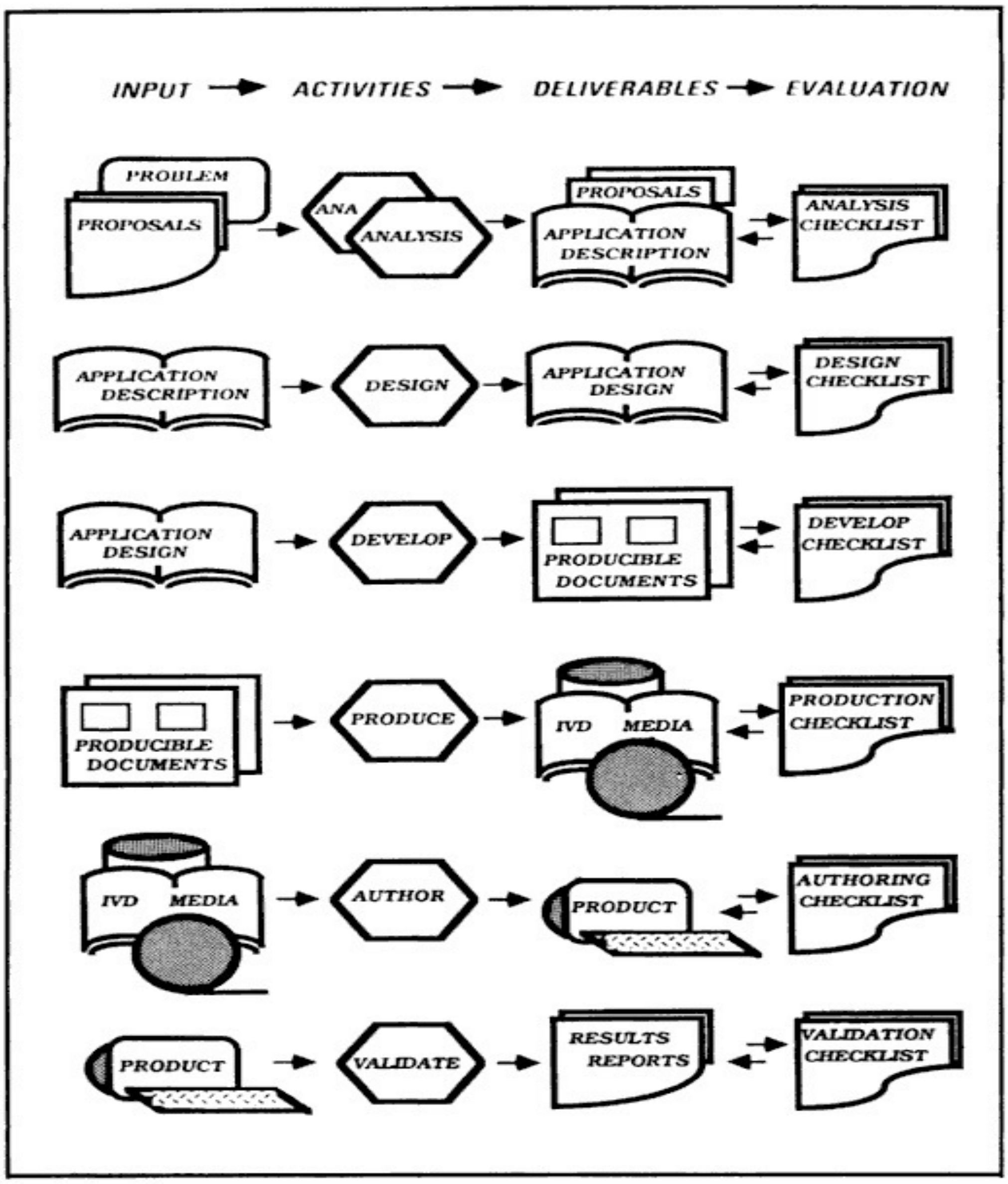


Figure 2.10 The Development Model. Adapted from 'Managing interactive video/multimedia projects'. (Source: Bergman and Moore, 1990*,* p.19)

Bergman and Moore's Model is a product orientation model that provides excellent and detailed checklists to guide the process of high multimedia technology production. Bergman and Moore's Model was selected due its main focus on the instructional production of new technology rather than editing exciting materials (Gustafson & Branch, 1997). Moreover, unlike other system models, the primary aim of Bergman and Moore's Model is to develop instructional video/multimedia productions, which can be exposed across a shorter amount of time, perhaps several hours or days (Gustafson & Branch, 1997). More recently, Branch (2009) incorporated Bergman and Moore's Model into a multifunctional process, involving the common fundamental elements: Analyse, Design, Development, Implementation, and Evaluation (ADDIE), as shown in figure 2.11. These components are interactive and require consistent evaluation and movement back and forth in a non-linear way, in a self-correcting manner.

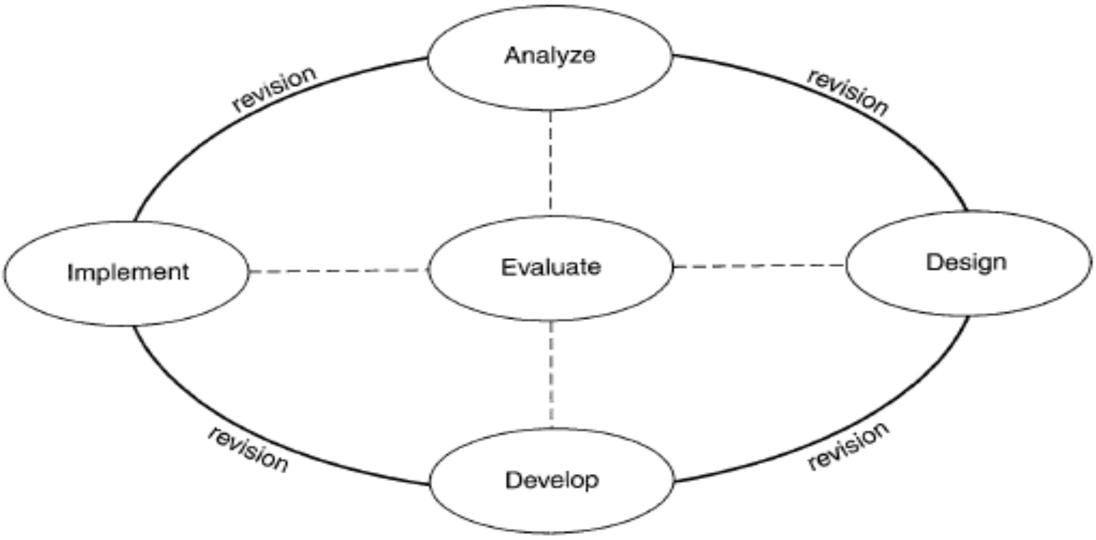


Figure 2.11. The five fundamental elements of ADDIE. (Source: Branch, 2009*,* p.2)

Bergman and Moore's six stages (analyse, design, develop, produce, author, validate), as shown in figure 2.10 above, were incorporated into Branch's first three ADDIE phases. The purpose of using both models when developing CAVA is to create an effective multimedia learning resource to be efficiently used to achieve set goals. The following sections will describe in detail each of the five ADDIE phases, and the working activities conducted in each, which are designed to manage CAVA instructional design process.

2.2.2.1 *Analysis*

This phase involves defining the scope of the work by conducting a needs analysis to identify performance problems in the educational environment, and then establish the main instructional goals of the software to respond to competency gaps. The preliminary analysis in this stage involved conducting a literature review describing how to use multimedia in L2 vocabulary learning. The instructional goals were intended to be answered in reference to research questions that respond to the knowledge gap. Another comprehensive analysis was conducted to document the characteristics of the software's target audience. CAVA is an instructional software intended to be delivered for beginner learners of Arabic as a foreign language. The target audience were intended to be total beginners, to ensure they have no baseline knowledge of target L2 Arabic items, and no other knowledge of Arabic as a second language. The course in CAVA provides intentional vocabulary learning, the effectiveness of which is assessed in the experiments conducted for this research. A third level of the analysis stage was conducted, requiring resources such as content, technology, human, and potential delivery methods to be analysed and determined.

In terms of the instructional approach, CAVA will employ one component of the four-strand balanced foreign language course (Nation, 2013), namely the strand that emphasizes language-focused learning in multimedia CALL environments. Language-focused learning is also called form-focused instruction, and is one of Nation's four equally essential strands. It is most applicable to high-frequency word teaching. High-frequency words are those words that account for a large proportion of the running words in spoken and written texts and language uses (Nation, 2013). 16 Arabic high-frequency words, four words divided thematically into four lessons, were initially chosen for delivery through CAVA. The number of words was increased to 24 over six lessons after the first pilot study. After applying all the required changes and corrections, the first version of CAVA (CAVA-1) consisted of six lessons with four high-frequency concrete L2 Arabic words grouped thematically in each lesson. This format proved adequate for addressing the research questions set out in the first experiment. For the second experiment, a second version of the software (CAVA-2) was developed with 108 words divided into twelve lessons.

Moreover, to eliminate the effect of the grammatical classification of the word, which is assumed to have an impact on the ease the learning process (James, 1996), the lexical items in CAVA were restricted to nouns. The word frequency, and the language features of the lexical items taught in CAVA were carefully selected to control the learning burden. In addition, the planned treatment and assessment of CAVA's vocabulary learning were based on views concerning the receptive and / or productive coverage of the nine aspects of word knowledge (see table 2.2). For instance, CAVA's treatment covers productive knowledge associated with the written target form of the Arabic word, and involves asking students to practice typing the word on the second slide in the first study phase. The aim was to test items productively during CAVA assessment (i.e. post-tests), by requiring the learner to recall and write down the learned word. More information about the phases of the software will be discussed later in the Design and Development phases of the current ADDIE process.

Table 2.2. Aspects of word knowledge contained in CAVA's treatment and assessment.
(*N.B.* R = receptive knowledge, P = productive knowledge)

Aspects of word knowledge			CAVA treatment	CAVA assessment
Form	Spoken	R	✓	✓
		P	No	No
	Written	R	✓	✓
		P	✓	✓
	Word parts	R	✓	✓
		P	✓	✓
Meaning	Form + meaning	R	✓	✓
		P	✓	✓
	Concept + reference	R	✓	✓
		P	✓	✓
	Associations	R	✓	✓
		P	✓	✓
Use	Grammatical functions	R	✓	No
		P	No	No
	Collocations	R	✓	No
		P	No	No
	Constraints in use	R	No	No
		P	No	No

In addition, CAVA does provide explicit coverage of associations as an independent variable in the ZUI condition. However, this can serve as a test of Nation's (2001) argument, which states that introducing new words with their related words usually does not result in effective learning but can cause interference and confusion. To examine this claim, word associations were included in the ZUI method of CAVA to communicate meaning and investigate their impact on vocabulary learning in the early stages.

As discussed in the literature review above, the direct vocabulary instruction, which is provided in CAVA, proved an effective way of teaching; one applied across a range of educational settings over the past decades. However, in the context of second language vocabulary research, it has been argued that one of the most important initial processes when learning new vocabulary is to establish a memory link between the word's form and its meaning (Laufer & Goldstein, 2004; Nation, 2001; Schmitt, 2010). Accordingly, the majority of the previous studies assessed the creation of this form-meaning link by communicating the meaning of the L2 word by presenting its translation equivalents, or an illustrated picture. The CAVA design aimed to go beyond teaching and assessing word associations, to evaluate another level of the word's meaning. The impact of this level of knowledge will be examined if it helps to strengthen form-meaning links.

With regard to the vocabulary learning strategies involved in CAVA, the computer's capabilities will be used to incorporate the best features of intentional vocabulary learning strategies, and also to avoid some associated pitfalls. CAVA will be designed to combine several complementary consolidation strategies. The following table (2.3) shows the forms of vocabulary learning strategies involved, according to Schmitt's (1997) taxonomy, as listed next to each feature of CAVA.

Table 2.3. Intentional vocabulary learning strategies involved in CAVA.

Learning strategy	CAVA feature	Interface
Social Strategies (SOC)		
Ask for an L1 translation	The program is doing the role of the teacher in providing L1 translation.	MDI
Ask for a sentence including the new word	The learner can read and listen to short sentences in which the word is used. The translation of the conversation can be seen, too.	All
Memory Strategies (MEM)		
Study word with a pictorial representation of its meaning	The user can view lists of words with pictorial representation of their meanings.	GUI, ZUI
Group words together to study them	The home page of each lesson groups the words together.	All
Group words together spatially on a page	The user can see the words organized spatially on the display.	ZUI
Use new word in sentences	All the target words are used in simple sentences.	All
Study the spelling of a word	The letters of each word are color-coded and the learner can hear the sound of each letter in addition to seeing its typed form.	All
Study the sound of a word	The learner can click a button to hear any word in the software pronounced by a native speaker.	All
Say new word aloud when studying	The learner can practice saying the word after listening to it.	All
Cognitive Strategies (COG)		
Verbal repetition	Since the software is self-paced, the learner has enough time to repeat the word aloud.	All
Written Repetition	The software is programmed to let the learner practice writing the target word as often as they want.	All
Word Lists	The learner is provided with lists of words displayed in the home page.	MDI
Put target language labels on physical objects	The words are written on the objects.	ZUI

Conducting the above analysis of the target audience, content, and delivery approaches and methods, helped generate an overall instructional plan that addresses the issues of who, what, when, where, why, and how? The analysis phase provided the answers to these questions and built the foundations for the remainder of the ADDIE designing process. This instructional software is going to be developed for undergraduate students of Arabic as a foreign language at the University of York, assisting them to intentionally learn high frequency concrete words, while also investigating effective vocabulary learning methods in multimedia environments using three types of user interface.

2.2.2.2 *Design*

In the design phase, the data for analysis is organized into a master plan. The main software segments will be sequenced, and their treatments defined. To achieve this, the objectives were written in measurable terms, and learning was categorized according to type, and the motivation, learning activities, media, interaction, and testing strategies then specified. These design specifications are documented in greater detail for production in multimedia materials (audio, text, graphics, or video). Each individual event associated with the software, such as menu choice, touch areas, and messages will be described to create reproducible documentation. This procedure will involve writing a storyboard of the software.

To complete measurable learning objectives for CAVA, the researcher considered Bloom's Taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956), which consists of six levels. Specifying learning objectives according to Bloom's taxonomy will have a direct impact on CAVA's learning phases, activities, assessments, and content. The taxonomy outlines a hierarchy of cognitive-learning levels, ranging from low to high levels associated with cognitive learning processes. The six major categories in the cognitive domain were *Knowledge, Comprehension, Application, Analysis, Synthesis*, and *Evaluation*. In a review of the taxonomy, Krathwohl (2002) renamed the first category *knowledge* to *remember* (see table 2.4). It includes, but is not limited to, four main action verbs for retrieving relevant knowledge from long-term memory, reflecting four levels of knowledge: list, recognize, recall, and identify. This first level of the taxonomy is deemed appropriate to the target audience, as they are beginner level learners of Arabic vocabulary. The choice of the first category, *remember*, allowed the developer of CAVA to customize the statements associated with its learning objectives according to anticipated learning outcomes, and helped to determine ideal testing strategies.

Table 2.4. Structure of the cognitive process dimension of the revised taxonomy. (Source: Krathwohl, 2002, p. 215)

1.0 Remember	Retrieving relevant knowledge from long-term memory. *1.1 Recognizing, 1.2 Recalling*
2.0 Understand	Determining the meaning of instructional messages, including oral, written, and graphic communication. *2.1 Interpreting, 2.2 Exemplifying, 2.3 Classifying, 2.4 Summarizing, 2.5 Inferring, 2.6 Comparing, 2.7 Explaining*
3.0 Apply	Carrying out or using a procedure in a given situation. *3.1 Executing, 3.2 Implementing*
4.0 Analyze	Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose. *4.1 Differentiating, 4.2 Organizing, 4.3 Attributing*
5.0 Evaluate	Making judgments based on criteria and standards. *5.1 Checking, 5.2 Critiquing*
6.0 Create	Putting elements together to form a novel, coherent whole or make an original product. *6.1 Generating, 6.2 Planning, 6.3 Producing*

Some of the other terms found to be suitable for use to describe the action verbs for *Remembering* are define, describe, identify, label, list, match, name, outline, reproduce, select, and state.

The following specific learning objectives were set for each of the CAVA learning conditions (MDI, GUI, and ZUI) and written in a SMART way (Doran, 1981), to apply a more comprehensive definition of the CAVA instructional goals in Specific, Measurable, Assignable, Realistic, and Time-related way:

- By the end of the program, the learners will be able to recognize the written form of the target L2 words when answering multiple choice questions in the post-test.
- By the end of the program, learners will be able to recognize the spoken forms of the target L2 words when answering multiple choice questions in the post-test.
- By the end of the program, the learners will be able to recognize the written form of the associated lexical set of target L2 words when answering word association questions in the post-test.
- By the end of the program, learners will be able to recall the written form of the target L2 words when answering the writing questions in the post-test.

These objectives helped to generate CAVA content and learning activities designed to focus on key areas of word knowledge, which were chosen in the analysis phase of the current ADDIE process, in an intentional multimedia vocabulary learning course:

- Teaching the written form of the target L2 word.
- Teaching the spoken form of the target L2 word.
- Teaching word associations with their lexical set in visuospatial representations.
- Teaching sounds and letters associated with L2 words (to increase phonological awareness).
- Practicing writing the target word using the keyboard.
- Practicing reading the word and finding its matching meaning representation (translation equivalent, pictorial representation, or visuospatial representation) form in a multiple choice context.
- Practicing listening (listen to a word and then find its matching word form in a multiple choice).
- Practicing associating the target L2 word with its lexical set (find the hidden word among a group of distractors to complete a list).
- Teaching the use of the target word in short sentences.

Testing word form-meaning links delivered in three ways by CAVA (translation equivalent- MDI, illustrated picture- GUI, or visuospatial associations- ZUI) will be done using more than one assessment type in the CAVA post-test, including drawing on both receptive and productive knowledge. To better achieve reliable results, receptive and productive learning will be encouraged during the study phases, and also through practice exercises, conducted in a consistent way using multimedia representation (MDI, GUI, or ZUI) as preparation for the post-test. Thus, the post-test in CAVA will aim to match the content and test behavioural actions in the *Remember* category of Bloom's revised taxonomy. Therefore, the multiple choice questions and assessment statements that are frequently used for this category of learning objectives, were developed to assess the following:

- Recognizing the written form of a target L2 Arabic word.
- Recognizing the spoken form of a target L2 Arabic word.
- Recognizing a word that is not from the same lexical set to exclude it.
- Recognizing a word that completes the lexical set by choosing an appropriate word from multiple options.
- Recalling the written form of a target word to write it in order to complete a lexical set.

Selected words

The number and type of the target Arabic words has changed as a result of the interactive nature of ADDIE, which includes consistent evaluation and editing. Following the selection criteria mentioned in the analysis phase, CAVA started with sixteen words split thematically into four lessons in a Latin square design of four words per lesson (see table 2.5).

Table 2.5. The first four four-word lessons in CAVA.

Lessons	Words
مكتب (Office)	طاولة (Desk) , كرسي (Chair) , حاسوب (Laptop) , دولاب (Bookcase)
طعام (Food)	بيض (Eggs) , حليب (Milk) , خبز (Bread) , ماء (Water)
العائلة (Family)	والدي (Father) , والدتي (Mother) , أخي (Brother) , أختي (Sister)
ملابس (Clothes)	بنطال (Trousers) , قبعة (Hat) , قميص (Shirt) , حذاء (Shoes)

After the first pilot study (see the evaluation phase for the current ADDIE process), one of the lessons has been replaced, and the number of words increased to 24 words over six lessons, as shown in Table 2.6.

Table 2.6. The six four-word lessons in CAVA.

Lessons	Words
مكتب (Office)	طاولة (Desk) , كرسي (Chair) , حاسوب (Laptop) , دولاب (Bookcase)
طعام (Food)	بيض (Eggs) , حليب (Milk) , خبز (Bread) , ماء (Water)
وظائف (Occupations)	إطفائي (fireman) , شرطي (Policeman) , ساعي (postman) , دهان (Painter)
ملابس (Clothes)	بنطال (Trousers) , قبعة (Hat) , قميص (Shirt) , حذاء (Shoes)
مطبخ (Kitchen)	قدر (Pot) , سخان (Kettle) , فرن (Oven) , مغسلة (Sink)
فصل (Class)	لوح (Board) , كتاب (Book) , حقيبة (Bag) , قلم (Pen)

The overall learning burden of all the lessons is approximately similar, since the linguistic features of the chosen words are matched.

Delivery methods of word knowledge in CAVA

With regard to the delivery methods for the different aspects of L2 word knowledge targeted in CAVA, the developer considered Nation's (2001) list of a range of aspects involved in the receptive and productive knowledge of a word, as discussed in the literature review. In CAVA, information is provided pertaining to all three aspects of word knowledge, including form, meaning, and the use of the target words, as shown in Table 2.7. In addition, CAVA makes use of three user interfaces (MDI, GUI, and ZUI) to teach different levels of the *meaning* aspect of the target words, whereas the *form* and *use* aspects were consistent across all three conditions.

Table 2.7. Aspects of word knowledge covered by the information in CAVA.

Aspect of word knowledge	Information	Delivery method in CAVA
Form	Written form Spoken form	Default information (all methods) Default information (all methods)
Meaning	L1 translation Picture Word association	Representation type (MDI) Representation type (GUI) Representation type (ZUI)
Use	L2 example sentence	Default information (all methods)

Three different sources of information of the word's meaning

In addition to the word's written and spoken forms, CAVA provides three different sources of information about the word's meaning: L1 translation, pictures, and visuospatial word associations. The target word كرسي *(chair)* for example has to be covered in CAVA with different aspects of the word knowledge involved in each representation. Default information in CAVA (i.e. L2 written and spoken form, L2 example sentence) is presented to all the students for every word. On the other hand, representational information (i.e. L1 translation, picture, word associations) is only provided to learners when they are studying the word with a user interface, to include this type of representation.

These three sources of information of the word meaning are chosen as *representational types*, and each is presented as a different condition of the experiment, in order to compare their effects. In the first condition (*MDI representation)*, the learner will be provided with the L1 translation. Under the second condition (*GUI representation)*, learners will see an illustrated picture. In the third and final condition (*ZUI representation)*, the learner will be exposed to visuospatial associations with the target words. One of the aims when including the latter source of meaning is to address the lack of research on the role of spatial memory in vocabulary CALL. The three sources are designed as alternatives, rather than as complementary ways of expressing the meaning of target lexical items, and each source is responsible for delivering a different type of information about the word's meaning. To achieve this, CAVA will be designed to compare these three ways of communicating word meaning and building L2 word overall knowledge. These approaches were considered by Nation (2001) as 'indirect' when he stated that “all ways of communicating meaning involve the changing of an idea into some observable form, are indirect, are likely to be misinterpreted, and may not convey the exact underlying concept of the word.” (p. 85). However, it has been argued that providing the second language learner with a word based translation is particularly effective for beginners with no previous knowledge of the target language (Chun & Plass, 1996; Plass & Jones, 2005).

Program Flow

In the design phase of the ADDIE process, the CAVA program flow was created to establish clear links between learning objectives and the course content. As presented in figure 2.12., the program flow was identical in each lesson for all three experimental representations. The program did not contain a pre-test, since all the learners were beginners, knowing nothing of the target language. The program consisted of five phases (see Table 2.8): two studying phases, a practice phase, an immediate post-test (*follow-up test 1*), and a delayed post-test (*follow- up test 2*). As shown in Table 2.8, each phase has either a recognition or a recall format. Phases 1 to 4 will take place together, whereas the intention is to plan a follow-up test 2 (phase 5) conducted one month later.

Table 2.8. The five phases in CAVA.

Phase	Corresponding name in CAVA	Format	Function
1- Studying phase 1	Studying phase 1	Recognition	Treatment
2- Practice phase	Practice phase	Recognition	Treatment
3- Studying phase 2	Studying phase 2	Recognition	Treatment
4- Immediate post-test	Follow-up test 1	Recognition, Recall	Assessment
5- 1-month delayed post-test	Follow-up test 2	Recognition, Recall	Assessment

Spaced repetition of the word’s presentation over the period of the program was also taken into account. Each target word was presented in CAVA on each slide of the lesson, across all the treatments. Spaced repetition was shown to lead to more efficient L2 vocabulary acquisition than massed repetition (Nation, 2013). The target item appears to the learner in the study phases, as well as in the practice phase.

The exposure time expected to complete six lessons of CAVA is about one and a half hours, including approximately fifteen minutes in each lesson. The post-test will take around thirty minutes.

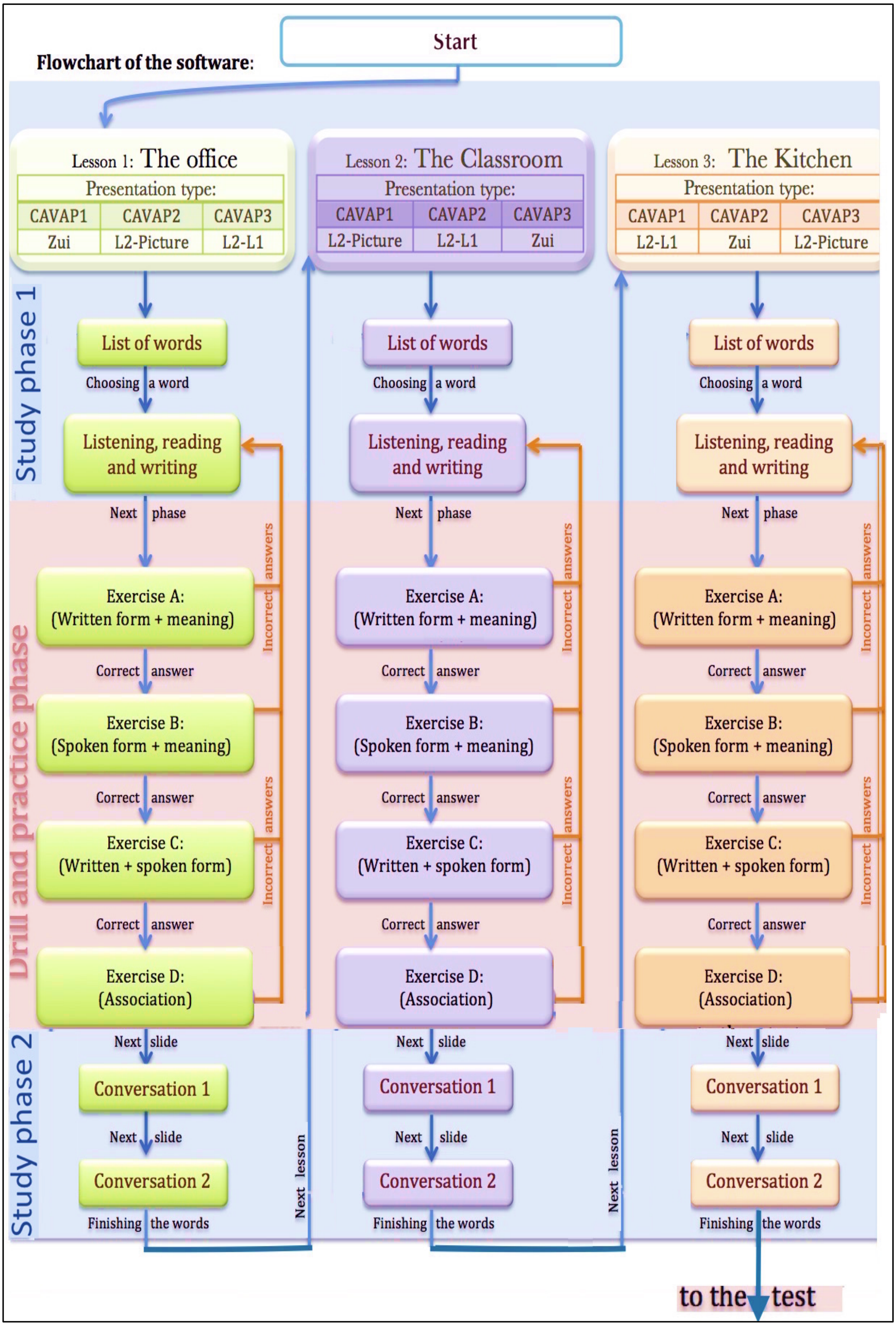


Figure 2.12. CAVA program flow.

Storyboard

After determining the main approach, the learning objectives, the content of the software, assessment strategies, delivery methods, program plot and duration, and the designer storyboarded the content of the software by creating a prototype. This was done by drawing the presentation and description of interactive events, and the required multimedia design elements to build each of the individual user interfaces. Storyboarding was used at the design stage of the ADDIE approach when identifying the CAVA specifications for a particular set of user interfaces. During the storyboarding, the screens were drawn on papers and the main elements of each frame also illustrated and timed. The control type and navigation systems were also written and revised. Storyboarding helped support editing and changes, making it easier to reach the final prototype of the content and learning activities that should support the learners' construction of knowledge. This prototype was written in a clear logical sequence, to guide development activities.

2.2.2.3 Develop

In the development stage of the ADDIE process, the documents produced and created in the design phase, will be transformed into their matching digital mediums, such as audios, texts, or graphics. To do this, sounds, pictures, colours and fonts were selected and/or created. Guided by the storyboard, production activities are performed at this stage to convey content in the desirable setting for which they were designed. Once appropriate media was produced, the authoring stage starts by integrating each of the individual media elements into an interactive presentation, to generate instructional software. To do this, the authoring activities of the software will involve programming user interfaces, control types, transitions, navigation styles, and the timing of presentation slides. At the end of the development stage, debugging and testing has to be performed on the software. According to any feedback given, the software is then reviewed and revised to prepare for the next ADDIE phase; implementation.

The audio (spoken form) for all the target words was recorded in a high-quality stereo recording using the Voice Memos app, a built-in app in MacOS Mojave. Since Voice Memos files are saved using the MPEG-4 format, with an extension .m4a, all the files were transformed to the extension .wav to be imported and played by PowerPoint as the main authoring program. All the sounds used in CAVA were recorded by the researcher and evaluated by an instructor of Arabic as a foreign language (both native Arabic speakers). After recording all the audios required for all the Arabic target words, the sounds for each letter were also recorded and saved in the same way. As planned in the storyboard, the audio was provided to the learners along with instructions about how to use it. With regard to the written form of the target Arabic word, the font (Al Nile, size 32) was selected and used for all the words, due to its high readability. To increase phonological awareness, the Arabic word is written in the studying CAVA phases in color-coded letters. Consequently, the learner can see and distinguish each letter of the word clearly. Moreover, the audio files for each letter sound, in addition to the whole word's pronunciation were imported to the authoring program, and linked to a speaker icon, which appeared next to the target word. This, as will be presented in detail below, provides the learner with the ability to hear the audio by clicking on the icon to listen to an Arabic native speaker voice.

The title of the lesson was written in Century Gothic font, size 36, in white over a pink background shape, positioned in the top right corner of the slide. The instructions were provided in English on the left side of the slide, following the same design as the title but in a smaller sized font, 24. These design features were default, and consisted of aligning all the phases of the software across the three multimedia representations methods.

The illustrations used to represent the meaning of the target words were selected and downloaded from google images. Some of the pictures were found to suffer a loss of quality when zoomed in on, and thus, higher resolution images were purchased from the dreamstime.com website. After downloading all the pictures from the Internet, some of the pictures were cropped and / or enhanced using Preview (Version 8.0 '859.21' 2014); an image editor software produced by Apple. During the first pilot study (see the evaluation phase of ADDIE), four L1 English native speakers (three male, one female), aged 18 – 22 years old, used CAVA and were asked to indicate whether they thought that the photographs were adequate pictorial representations of the target words. The interview results indicated that the four raters believed all the pictures visually represented the meaning of the target words effectively.

After careful production of selected media, the authoring stage started. These design elements were integrated to develop an interactive presentation, based on instructional phases (software plot) written and storyboarded during the design stage of the ADDIE process. The following subsections will demonstrate how the selected individual media was put together to develop three user interfaces (MDI, GUI and ZUI) using PowerPoint versions 2013 and 2016; a presentation software by Microsoft.

Developing the Menu-driven Interface (MDI)

The first type of representation in CAVA is a MDI, which presents a list of target second language words (L2), with their equivalent translations (see figure 2.13.). In this representation, L1 translations are delivered as the meaning information pertaining to the target Arabic words.

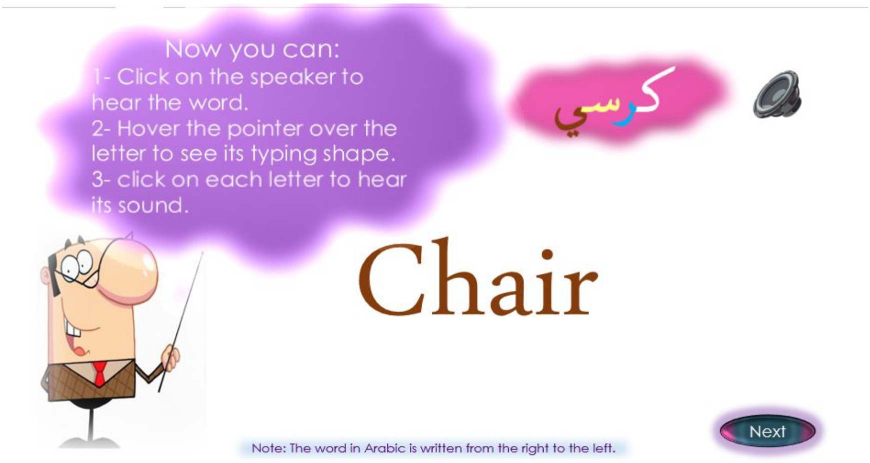
3- My office

You can find these items in my office. Please click the first word in the list:

Desk	مكتب
Chair	كرسي
Laptop	حاسوب
Bookcase	خزانة

Figure 2.13. The home page of the lesson (My office) in MDI.

Following the instructions given, the learner can click on the directed word, which is hyperlinked to another slide (see in Figure 2.14.). This new slide then appears in a *fade* transition effect, presenting the selected target word, which is translated by the single word that expresses its most common meaning, and both are written in a larger font size (e.g. *chair* as the translation of كرسي). The translation equivalent is positioned at the centre of the slide, in 166 font size, and the Arabic target word is located on the top right corner in 88 font size.



Now you can:
1- Click on the speaker to hear the word.
2- Hover the pointer over the letter to see its typing shape.
3- click on each letter to hear its sound.

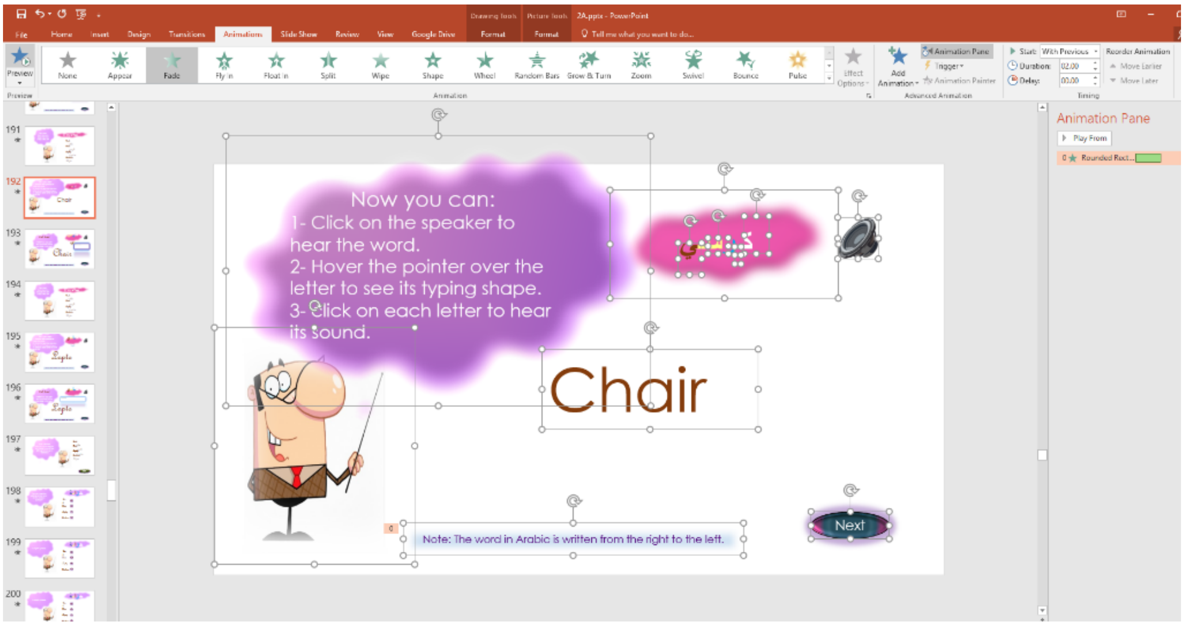
كرسي

Chair

Note: The word in Arabic is written from the right to the left.

Figure 2.14. CAVA screenshot of the target word كرسي *(chair)* in MDI representation.

The navigation hyperlinks to related audio, transition effects, or other slides behind the scene, can be shown in figure 2.15. These hyperlinks, which were programmed on this slide are the underpinning structure that provided the user with multimedia, as planned in the ADDIE design stage for this kind of user interface.



Now you can:
1- Click on the speaker to hear the word.
2- Hover the pointer over the letter to see its typing shape.
3- Click on each letter to hear its sound.

كرسي

Chair

Note: The word in Arabic is written from the right to the left.

Next

Figure 2.15. Navigation hyperlinks underpinning MDI interface.

Developing the Graphical User Interface (GUI)

Pictures are used to communicate the meaning information related to the target words in the second type of multimedia representations in CAVA. This representation employed a GUI to display the list of target second language words (L2) along with the illustrative pictures. In this representation, a set of coloured photographs is presented to illustrate the meaning of target nouns (see Figure 2.16). To eliminate any possible impact from the cultural dimension on the meaning and use of the vocabulary (Nation, 2011), all the pictures were taken from western culture. All the pictures in this user interface have an approximately similar size of 95 x 85 pixels.

2- My office

These tools are used in my office. Please click the first word in the list:

حاسوب

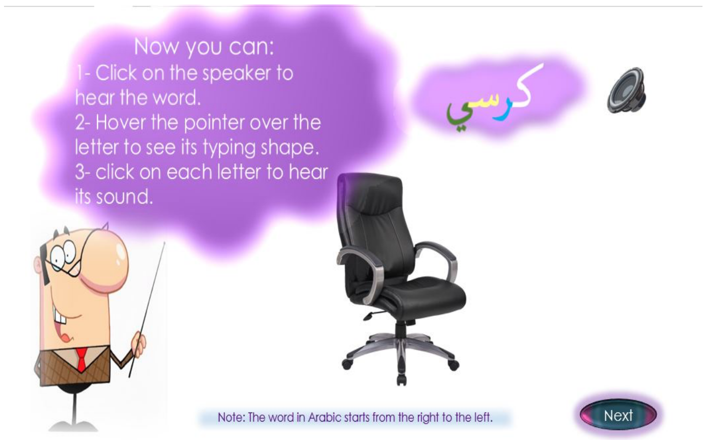
مكتب

خزانة

كرسي

Figure 2.16. The home page for the lesson (My office) in GUI.

Each pair of words and its picture were joined in an unseen background box, hyperlinked to the related studying slide (see Figure 2.16.). Once the learner follows the instructions and clicks on the word, he will be navigated to the study slide, where he can study the target word at his own pace. On the study slide, as shown in figure 2.17, the target Arabic word is positioned near to the top write corner in 88 font size, whereas the relevant illustration is positioned in the centre of the slide, in a larger size area of around 384 x 288 pixels.



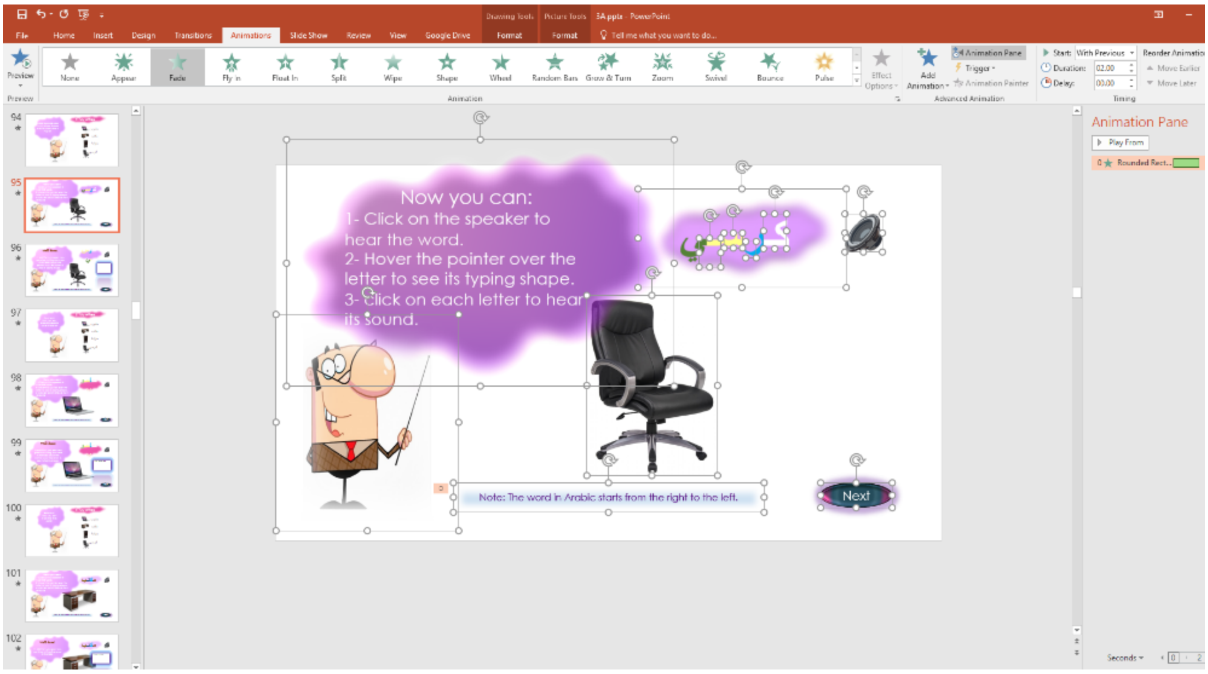
Now you can:
1- Click on the speaker to hear the word.
2- Hover the pointer over the letter to see its typing shape.
3- click on each letter to hear its sound.

كرسي

Note: The word in Arabic starts from the right to the left.

Figure 2.17. CAVA screenshot of the target word كرسي *(chair)* in GUI representation.

The hidden navigation hyperlinks connect the designed objects to their related audio, transition effects, or elements that were linked to subsequent slides, as shown in figure 2.18.



Now you can:
1- Click on the speaker to hear the word.
2- Hover the pointer over the letter to see its typing shape.
3- Click on each letter to hear its sound.

Note: The word in Arabic starts from the right to the left.

Figure 2.18: Navigation hyperlinks underpinning GUI interface.

In the MDI user interface, the slide transition effect used is *fade*, and the hyperlinks were programmed to provide the learner with multimedia, as planned for this kind of user interface in the design stage of ADDIE.

Developing the Zoomable User Interface (ZUI)

The third type of multimedia representation in CAVA is the ZUI, which presents a visuospatial layout of target second language words (L2), as an overview of the environment in which the words are actually being used. In this interface, word associations are presented to the learner as meaning information pertaining to the word, in addition to including the word picture as a part of a larger visual representation involving pictures of other related words. This show of word associations is encouraged in second language vocabulary learning as “understanding these relations is useful for explaining the meanings of the words and for creating activities to enrich learner’s understanding of the words” (Nation, 2013, p. 79).


This is my office. Please click on the underlined word on the photo:

1- My office

Figure 2.19. An overview of an environment (the office) where related words are used.

Again, all the pictures in this representation are taken from western culture. They were downloaded from the Internet and some were cropped and / or enhanced using the same image editor software; that is Preview (Version 8.0 ‘859.21’ 2014). All the pictures in this representation were approximately 825 x 588 pixels in size, with a standard 8:6 width-height-ratio. Since the ADDIE process is interactive, formative evaluation took place alongside each stage. As a part of this type of evaluation process, the participants in the pilot studies believed all the pictures served as adequate pictorial representations of the target words. Each picture in this representation presents four target Arabic words, which are thematically related to each other, and usually exist in a common environment. For example, the office picture shows a desk, chair, laptop and a bookcase (see figure 2.19.). The names of these parts are written in Arabic, and were placed near the corresponding pictures of the item to apply the spatial contiguity principle, which it is strongly recommended to apply to low-knowledge learners (Mayer, Steinhoff, Bower, & Mars, 1995) to help minimize extraneous cognitive load.

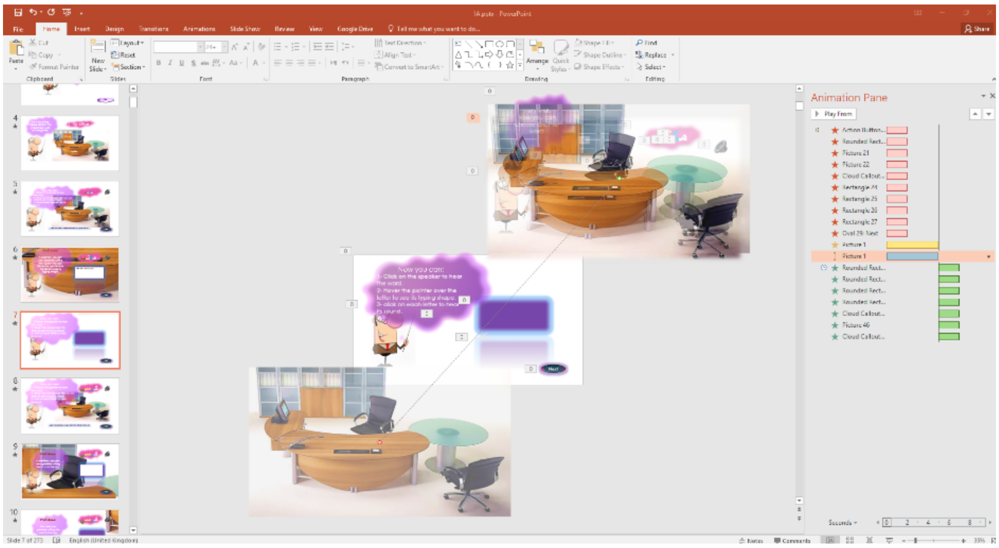
When the learner clicks on the each of the target words, the program zooms out to enlarge the size of the selected object to fill a new screen, on which the learner can then study the word (see figure 2.20.).


Now you can:
1- Click on the speaker to hear the word.
2- Hover the pointer over the letter to see its typing shape.
3- click on each letter to hear its sound.

Note: The sentence in Arabic starts from the right to the left.

Figure 2.20. CAVA screenshot of the target word كرسي *(chair)* in ZUI representation.

The programming behind this zooming transition can be seen in figure 2.21. It requires the addition of a special animation type (grow in 400%) to zoom in, accompanied by a directed path. This transition takes place over two seconds, in which a group of elements related to the main overview (shown in red in the animation pan) play out, and another group of elements played in (green) relate to the new slide, which will present the selected word to be studied.



Now you can:
1- Click on the speaker to hear the word.
2- Hover the pointer over the letter to see its typing shape.
3- click on each letter to hear its sound.

Figure 2.21. Navigation hyperlinks underpinning the ZUI interface.

As shown above, the three multimedia representations were developed to deliver different aspects of the word’s meaning; however, the representation theme (MDI, GUI, or ZUI) is consistent in all the phases of CAVA, as a consistent background to each condition, even in the slides that present default information. Developing the slides that display both the default and type of information used in CAVA are described in detail in the following subsections, throughout the software phases.

Integrating the software phases

Guided by the storyboard discussed above, the phases of instructional software will be integrated and developed to teach lessons that introduce four thematically grouped words. Although there are three multimedia representation types, that fulfil the three main conditions of the experiment, the program flow will be developed to present the same learning process into five phases.

The first studying phase (phase one)

The first studying phase introduces the learners to the words. This phase starts with the home page, on which the learner can see all the lessons. In this study phase the learner chooses one word from the list. As they click on the word, the program will navigate to another slide where the learner studies the chosen word. They will be able to hear the word as often as they want in addition to listening to the sound of each letter of the word. The letters of the word are color-coded so that the learner can recognize them clearly in order to raise their phonological awareness. In addition to providing the learner with these forms of word based information, the meaning of the word is introduced through its translation, an illustrated picture, or a pictorial-based word associations. In the second slide, the learner is given the chance to practice writing the word as often as needed. The learner is always given instructions in a clear and polite way (see the implemented phase of the current ADDIE process). The learner will be guided to write the word correctly, as he can see the shape of the letter typed, in a pop-up window, when hovering over it using the pointer. The typed shape is identical to the shape shown on the keyboard. That is because the shape of the letter in Arabic looks slightly different from its corresponding shape in isolation (see figure 2.22.). Nation (2013) said that spelling is one of the features used to gain familiarity with the word's written form, and can be improved by early training. Thus, if the word is comprised of known parts, the learning burden will be relatively light. Similarly, the spoken form is often provided at these two levels; that is the letter and the word. When clicking on a letter, the learner hears a sound. In addition to listening to letters, another icon is designed and placed next to the word to be clicked on in order to hear the pronunciation of the full word (see figure 2.22.). When the slide is being displayed the participant has the option to replay the audio pronunciation as often as needed. Moreover, the audio recording is played automatically once the learner answers correctly in the practice exercise phase of CAVA.

Well done!

In addition, you can now practice writing the word in the box:

Note: The word in Arabic starts from the right to the left.

Figure 2.22. A pop up window to show the typed shape of the letter to facilitate writing the word in the box.

When the learner finishes writing, he can click the *Next* button, which is programmed to navigate to the homepage, where the learner has to then choose another word from the list to study. After studying all the words, the learner then moves to the next phase, the practice phase.

The practice phase (Phase two)

Following the first study phase, the participant moves on to the practice phase. Practice exercises were accompanied by feedback. Three productive recognition tasks were developed in a multiple choice format for the practice phase. The first task is an exercise that links the word's meaning with its written form. The meaning (L1 translation, picture, or visuospatial) of the Arabic target noun is presented along with three other meanings as distractors. The learner has to then identify the meaning of the written form of the Arabic target word, as displayed on the screen (see figure 2.23.).

Figure 2.23. The first exercise in the practice phase (meaning-written form).

The same process is repeated in the second task of the practice phase, but as an exercise in linking meaning with the spoken form of the Arabic target word, rather than its written form (see figure 2.24.).

Figure 2.24. The first exercise in the practice phase (meaning-spoken form).

The third and final task is an association practice exercise, in which the learner practices finding target Arabic words among other distractors, to complete a list of semantically related lexis presented with their translation equivalents for MDI learners (see figure 2.25.), illustrated pictures for GUI learners, or a visuospatial overview for ZUI learners.

Desk	مكتب
Chair	كرسي
Laptop	حاسوب
Bookcase	....

Figure 2.25. The third exercise in the practice phase (word associations) in the MDI interface.

In all three tasks of the practice phase, the program was developed to provide the learner with the required instructions, as well as written audio feedback about their responses (see figure 2.26.). In addition, the feedback message displayed the correct answer in a new slide, and linked the word with its correct meaning in a specific representation type, which matched the learner’s study phase theme (MDI, GUI, or ZUI).

Brilliant!

That's the correct answer. Thank you.

كرسي

Figure 2.26. Feedback for the correct answer in the GUI interface.

The second studying phase (phase three)

After developing the practice phase, which aimed at consolidating the vocabulary knowledge required for the L2 target word, the second study phase was developed to produce information concerning word use. The target word appears in two audio/written conversations, each on a separate slide. These slides were developed to provide the learner with the capability to hear each word in short sentences and also to see the English translation of Arabic words, except for the target ones. This was done by following the same mode of programming when linking media events (audio and visual) with each other in the first study phase, as discussed above. The default was to use information consisting of Arabic example sentences relating to the target word in two short audio/written conversations. For example, for the word كرسي *(chair)*, the example sentence used was هذا كرسي (This is a chair.) (see Figure 2.27.). This sentence answers the question (what is this?). Then there is another question about the characteristics of the item (what does it look like?), followed by another short answer (It is black.) This was the first conversation written to teach the learner about how the word is used in w/h questions.

Conversation 1

(Wh questions)

1- Hover your mouse over the word to see its translation.

2- Click on the word to hear it.

ـ ما هذا ؟

ـ هذا كرسي.

ـ كيف يبدو ؟

ـ إنه أسود.

Note: The sentence in Arabic starts from the right to the left.

Figure 2.27. CAVA screenshot of the second study phase of the target word كرسي *(chair)* in a GUI representation.

The second audio/written conversation consists of yes/no short sentences (Is this a chair? Yes, it is) and (Is it brown? No, it is black) (see figure 2.28.). The learner can see an English translation of each Arabic word in the conversation (except for the target word) in a pop up window, by hovering the pointer over the word. The target word did not have a pop-up translation, since its meaning was presented representationally.


Conversation 2
(Yes/no questions)

1- Hover your mouse over the word to see its translation.
2- Click on the word to hear it.

ـ أهذا كرسي؟
ـ نعم.
ـ أهو أحمر؟
ـ لا إنه أسود.

Is it red?

Note: The sentence in Arabic starts from the right to the left.

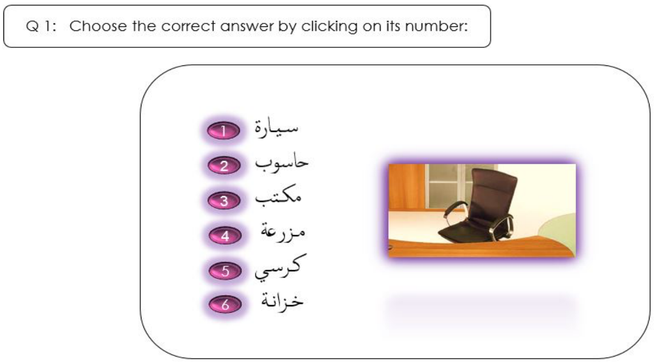
Figure 2.28. The translation of the sentences appears when hovering.

In addition, the slide was developed to enable the learner to hear any word in the conversation by clicking on it. All the target Arabic words are nouns and presented in the example sentences in singular form. Again, the sentences in the two conversations were written on the storyboard by the researcher during the design phase of the ADDIE process, and then revised by an instructor of Arabic as a foreign language.

To sum up, the first three phases of the software were integrated to present default information in CAVA; these focused on the form (written and spoken L2 word) and use (L2 example sentences in two written and audio conversations) of the target word. This information was developed to be identical for all the participants. At the same time, the three phases of the software were developed to present the participants with one of three representation types that differ in the way they provide the meaning (L1 translation, pictorial, or visuospatial representation) for each word. The software is designed to be used to investigate the effect of these three ways of communicating the meaning of the word, based on the premise that meaning is a more important aspect of word knowledge than either form or use.

The two post-tests (phases four and five)

As planned in the storyboard, the post-tests in CAVA were developed in a receptive and productive recall format. The same design elements used when developing the study phases were used to create the test slides. The tests consisted of five questions. The first question tested the participant’s ability to link the target meaning of the Arabic word (L1 translation, picture, or visuospatial representation) with its written form. Five other distractors are then presented along with the correct answer. The learner has to then identify the meaning of the written form of the Arabic target word, as displayed on screen (see figure 2.29.).

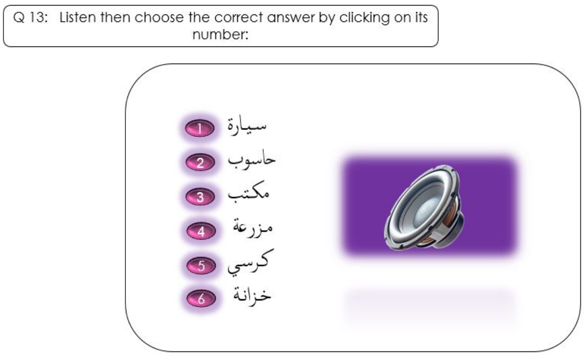


Q 1: Choose the correct answer by clicking on its number:

1 سيارة
2 حاسوب
3 مكتب
4 مزرعة
5 كرسي
6 خزانة

Figure 2.29. The first type of questions in the post-test (meaning-written form) in the ZUI interface.

Once the learner responds to the question, he will then be automatically moved to the next question. In the second type of question format on the post-test, the same process is repeated, but to test its ability to link meaning with the spoken form of the Arabic target word, rather than its written form (see figure 2.30.).

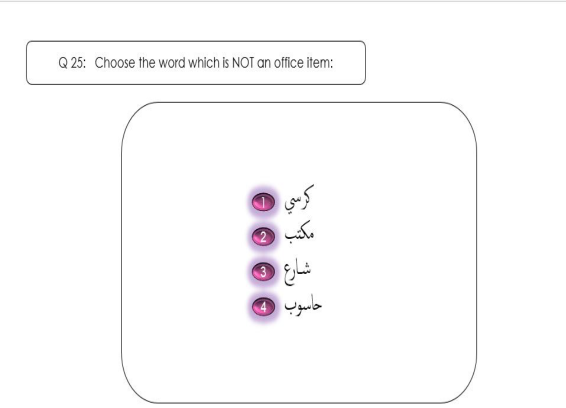


Q 13: Listen then choose the correct answer by clicking on its number:

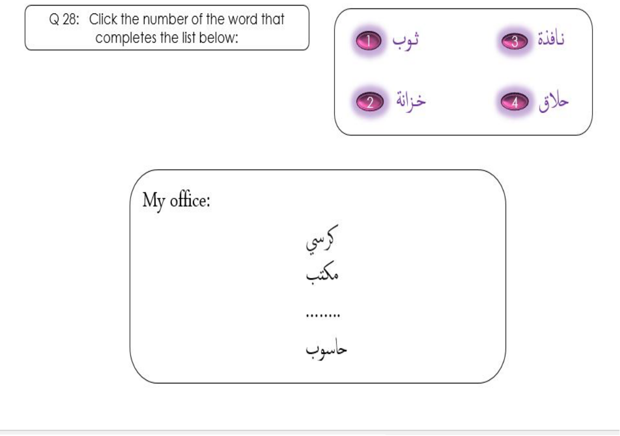
1 سيارة
2 حاسوب
3 مكتب
4 مزرعة
5 كرسي
6 خزانة

Figure 2.30. The first type of questions in the post-test (meaning-spoken form).

The third, fourth and fifth types of questions on the post-test were association based. The third type were developed to test the learner’s ability to select the odd word out from a list of related words. The fourth assesses the learner’s ability to find the hidden target Arabic word among other distractors, as a way to complete a list of semantically related Arabic words. Lastly, for the fifth type of question, the learner is required to write the Arabic word that completes the list of four semantically related words (see figures 2.31., 2.32., and 2.33.).

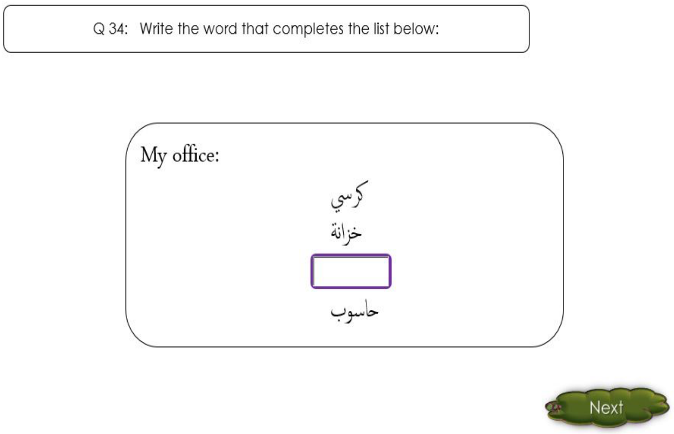

Q 25: Choose the word which is NOT an office item:

1 كرسي
2 مكتب
3 شارع
4 حاسوب


Q 28: Click the number of the word that completes the list below:

1 ثوب
2 خزانة
3 نافذة
4 حلاق

My office:
كرسي
مكتب
....
حاسوب


Q 34: Write the word that completes the list below:

My office:
كرسي
خزانة
حاسوب

Figures 2.31, 2.32 and 2.33. The final type of questions in the post-test (word association).

The immediate post-test (follow-up test 1) and the delayed post-test (follow-up test 2) were identical.

During the developmental stage, versions of the software were shown to experts in the field of educational technology for appropriate revision upon the validation of the software until the final version had been completed. The feedback proved useful, and was applied to the software; it then became functional with no errors. The average number of slides proved to be 90 per lesson. Each lesson was then designed in all three of the teaching methods (MDI, GUI, and ZUI), and thus the total number of slides used in each lesson to develop the three methods is 270 slides. The total number of slides for the 6 lessons in the first version of CAVA, for all three representations, was 1620 slides.

2.2.2.4 Implement

The fourth phase of the ADDIE model of instructional design is the implementation phase. In the implementation stage, learners were registered, and then exposed to pre-course interaction with the software. The researcher himself was a facilitator, and thus aware of key learning outcomes, delivery methods, and testing strategies. The software was delivered to the learners on a 13-inch MacBook laptop, which was then connected to a 22-inch Samsung screen for the researcher to monitor the learning process.

The learners were prepared and trained on the use of the new software. The learners could have chosen to take and complete a course guided by the software itself. CAVA is computer-control software that controls the pace for all learners in a guided plot using control buttons, to ensure a similar learning path for all. Research (for example; Gay, 1986 and Young, 1996) shows that less proficient learners learn less in learner-controlled mode than with program controls.

To implement CAVA, the personalization principal (Clark & Mayer, 2011) was applied through the use of both conversational styles and virtual coaching. Instead of relying on formal instructions, the developer used a conversational style to provide instructions. The use of first and second person language was found to be more effective when learning in computer based environments than an impersonal style (Moreno & Mayer, 2004). The instructions were delivered to the learners via a polite conversation with an on-screen pedagogical agent called Cava, which performs the role of guiding participants through the program and providing consistent feedback and hints (see figure 2.34.).

Hello. I'm Cava.

I will be happy to tell you about my life in the Arabic Language.

It's nice to be with you. We are going to discover these topics:

Figure 2.34. Agent Cava is providing polite and clear instructions.

2.2.2.5 Evaluate

The fifth and last phase of ADDIE is the evaluation phase. This evaluation phase is an interactive process that can be conducted alongside other stages, to produce the finished version of the software. The software is primarily evaluated against the objectives for which it was designed. The activities in this phase involve both formative refinements and a summative evaluation of the software in light of its effectiveness at achieving instructional goals.

Formative evaluation

To ensure a successful design of the current study, a formative pilot study was conducted alongside the other ADDIE phases. It involved using CAVA when teaching vocabulary to assess the effectiveness of the software. This formative pilot study was conducted in the spring term over a period of three weeks, starting from Monday 9th February 2015. The following subsections describe this formative piloting studies in detail.

Undergraduate students at the University of York were contacted by email, and then invited to participate in testing a new multimedia representation. They were informed that the experiment would involve interacting with a computer based instructional program, and answering an achievement test. In addition, all the participants knew that the experiment would take place over one day, and last an 1 and a half hours. The participants were compensated £15 in thanks for their time. To encourage them to do their best in both sessions, the participants were informed that there would be an extra £10 for the person who achieved the highest score in the CAVA achievement test. To be eligible the participants needed to be undergraduate university students, 18-23 years old, native monolingual English speakers, and able to give informed consent. Four L1 English native speakers (three male, one female), age (18 – 23 years old) chose to participate in this formative pilot study.

All sessions were conducted in a private study room in the University of York library (see Figure 2.35.). The researcher met each participant individually to introduce them to the experiment, describing the main sessions clearly. The participants were subsequently informed that the research would focus on assessing the acquisition of vocabulary in a multimedia CALL environment, but they were not given any details regarding the specific research question. The researcher then asked the participants to read and sign an informed consent form. The researcher's personal laptop (which hosts CAVA as the apparatus of the study and an Arabic/English keyboard) was used during all parts of the experiment.

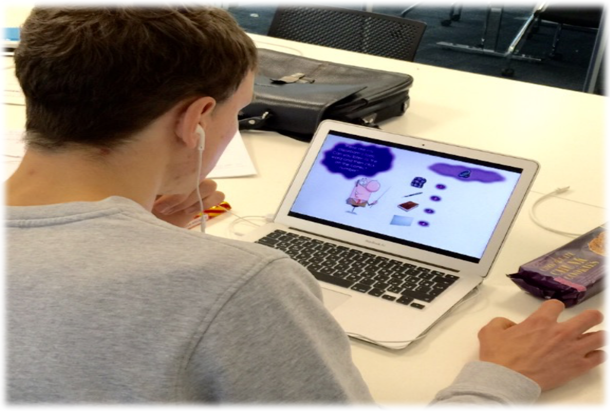


Figure 2.35. A private room at the University of York library was booked to meet each participant.

In the meeting, the participants were exposed to CAVA and completed four of its phases (i.e. study phase 1; practice phase; study phase 2; and the immediate post-test). The learning process in CAVA is computer controlled, with the learner receiving instruction from Cava on each software slide. The researcher was also available at all times to monitor the process, and provide the participants with any necessary assistance. This session lasted approximately one and a half hours.

Feedback on the development of CAVA

During the study, the researcher observed and recorded some errors in the software design, such as with some of the hyperlinks. In addition, he observed a number of typographical errors to be corrected. After exposure to CAVA, the participants were asked to give feedback on the design elements of the program. Furthermore, they were asked to indicate whether they thought the pictures served as adequate pictorial representations of the target words. The interview results indicated that the four participants generally thought all the pictures afforded good visual representations of the target words. The participants also indicated that they had completed the experiment comfortably, and that the time provided was sufficient. In addition to editing all the errors, orthographically similar words, which can affect learnability (Nation, 2013), were changed after the formative pilot study. For instance, all the participants in this study achieved low scores in one of the lessons (the family), because it contained items such as والدي *(my father)* and والدتي *(my mother)*, which were similar and overly challenging. As a result, the researcher, who designed CAVA, omitted and changed the lesson. Furthermore, two more lessons were added to CAVA, to create six instead of the original four to enrich the experiment.

Summative evaluation

A summative pilot study was conducted at the culmination of the ADDIE process. This study was conducted over six weeks in the spring term of the academic year 2015. It ran from Monday 25th of May to Friday 26th of June. The purpose of this summative evaluation was not only to assess CAVA, but also to pilot the methods of data collection to be used in the main experiment for the current thesis. To do so, data will be collected from CAVA post-tests for comparison to the learner's performance in the Automated Working Memory Assessment (AWMA). The following subsections describe summative piloting in detail.

The participants in this study were undergraduate students at the University of York, and they were invited to take part in this study according to the same format as was applied in the previous formative pilot study. Five different L1 English monolingual speakers, age (18 – 23 years old, two female and three male) also participated in the summative pilot study. All the sessions were conducted at the same place, following the same procedure. The main modifications, in addition to the revision of CAVA, were the addition of two lessons, and a delayed post-test, as well as the AWMA. The participants were informed that the experiment would involve a test of working memory capacity. In addition, all the participants knew that the experiment would take place over two days: 1 and a half hours each day. The participants were compensated with £30. To encourage them to do their best in both sessions, the participants were informed that there would be an extra £10 for the person who achieved the highest score in the CAVA achievement test, and another £10 for the participant with the highest score in AWMA, the working memory test.

The AWMA session took place one day before the CAVA, the researcher assessed the participant's working memory capacity using the AWMA. This is composed of 12 tests, focused on verbal, visual and spatial abilities. In each subtest of AWMA, all the participants received identical instructions, which were provided by the program to explain how to respond to the questions posed on the test. This session took around one hour to complete.

In the subsequent session, the participant was exposed to CAVA, and completed all five phases (study phase 1, practice phase, study phase 2, the immediate post-test, and the delayed post-test). These sessions lasted for around one hour.

To assess the use of CAVA when tackling the research questions for the main experiment, the results of both the immediate and delayed post-test were analysed and correlated with the working memory test scores, as presented in the following subsections.

Experimental Variables and Statistical Tests

Two independent variables with different levels were considered in this pilot study: multimedia representational types and individual differences in working memory. See table 2.9 for more detail regarding these variables.

Table 2.9. Independent variables of the experiment.

Independent variable	Type	Levels	Measure
Multimedia representation	Within-subjects	Three levels: MDI, GUI, ZUI	Nominal
Individual differences in working memory	Within-subjects	Three levels: Verbal, Visual, Spatial	Nominal

The dependent variable in this study was vocabulary learning, which was operationalized according to the scores that the students achieved in the immediate and delayed post-tests. Each learner completed the same test, and their scores in this experiment ranged from 0 – 39 points.

The study used inference tests to establish relevant effects, in order to address the two research questions shown in Table 2.10. The level of alpha was set to .05, to determine the statistical significance of all the tests. The inferential results for the first research question were answered by conducting a one-way within-subjects (repeated measures) ANOVA, with multimedia representation being the within-subject factor. A Pearson Correlation was also conducted to answer the second question posed in the study. Both tests were conducted using SPSS.

Table 2.10. Research questions and the inference test of the study.

Research question	Effect tested	Inference test
1- What is the effect of multimedia representation types on vocabulary acquisition?	Main effect multimedia representation types	ANOVA ($\alpha = .05$)
2- What is the effect of working memory abilities on vocabulary acquisition?	Main effect of working memory abilities	Pearson Correlation

The results of the study were presented in two following subsections, corresponding to the two research topics. The first section investigates vocabulary acquisition by displaying the results for the main effects of multimedia representation. The second section explores the results for the main effect of working memory abilities by examining multimedia representation, together with individual differences in working memory, as a way to investigate the interaction between representation and working memory in vocabulary acquisition.

CAVA results

The current research focussed on the effectiveness of multimedia representations in CALL environments for second language vocabulary acquisition. It examined both the short-term and long-term effects of a number of different interactive user interfaces (i.e. MDI, GUI and ZUI), which influence L2 depth of processing and vocabulary acquisition. Firstly, there is a discussion of the immediate post-test findings, followed by details of delayed post-test results.

Immediate Post-test

Table 2.11 shows the results of the immediate post-test for the five participants, in relation to the main effect of the multimedia representation type.

Table 2.11. Immediate post-test results for the three multimedia representation types.

Multimedia representations	Rank	Mean	Maximum score	Percentage	SD
MDI	1	17	26	65.3 %	4.8
GUI	3	14.8	26	56.9 %	7.1
ZUI	2	15.2	26	58.4 %	4.6
All multimedia representations		47	78	60.2 %	

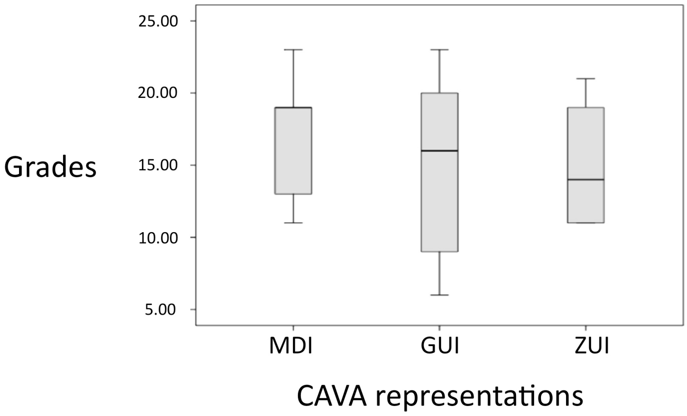


Figure 2.36. The difference in the means for the three CAVA representations in the immediate post-test.

Descriptive statistics of the immediate post-tests results

Table 2.11 reveals that the learners achieved an average score of 47 out of 78, representing a vocabulary acquisition score with a mean rate of 60.2%. Figure 2.36. displays multimedia representation (MDI) (i.e. a form of verbal based L2-L1 user interface), which resulted in the highest mean score, with 17 out of 26 (26 being the maximum possible points for the thirteen questions on MDI words), representing an average of 65.3 %. This was subsequently confirmed as the most effective support for vocabulary acquisition in the immediate post-test. The second most effective Multimedia representation was ZUI, which contained a layout with a zoomable interface, and a mean of 58.4%, closely followed by GUI (graphic based L2-picture user interface) with a mean of 56.9%. The raw scores for the five participants can be seen in table 2.12 below.

Table 2.12. Raw scores showing the results for the five participants in the immediate post-test.

Participant	MDI	GUI	ZUI
s1	58.97	58.97	53.85
s2	33.33	23.08	28.21
s3	48.72	51.28	48.72
s4	28.21	15.38	28.21
s5	48.72	41.03	35.9

Inferential statistics of the immediate post-tests results

The immediate post-test examined the primary influence of multimedia representation in the formal hypothesis, when tested with an alpha level of .05. The difference proved insignificant (F (2, 12) = 0.21, p = 0.8). Post-hoc multiple pairwise comparisons revealed that the multimedia representation of MDI proved more effective than the multimedia representations for GUI and ZUI for the acquisition of vocabulary (See Table 2.13).

Table 2.13. Pairwise comparisons for multimedia representation on the immediate post-test.

Multimedia representation I	Multimedia representation J	Mean difference (I – J)	Standard error	Sig.
MDI	GUI	2.2	3.5	0.81
	ZUI	1.8	3.5	0.87
GUI	ZUI	- 0.4	3.5	0.99

Delayed Post-test

All the participants confirmed that they had not deliberately studied the target vocabulary between the completion of the immediate post-test and the delayed post-test.

Descriptive statistics of the delayed post-tests results

Table 2.14 demonstrates the results for the delayed post-test, including that mean word retention was 41% (i.e. 32 / 78 points), which proved lower than the results for the immediate post-test (i.e. 60.2%, or 47 / 78 points).

Table 2.14. Delayed post-test results for the three multimedia representation types.

Multimedia representations	Rank	Mean	Maximum score	Minimum score	SD
MDI	2	10.4	26	40 %	2.7
GUI	3	9.8	26	37.6 %	3.7
ZUI	1	11.8	26	45.3 %	2.1
All Multimedia representations		32	78	41 %	

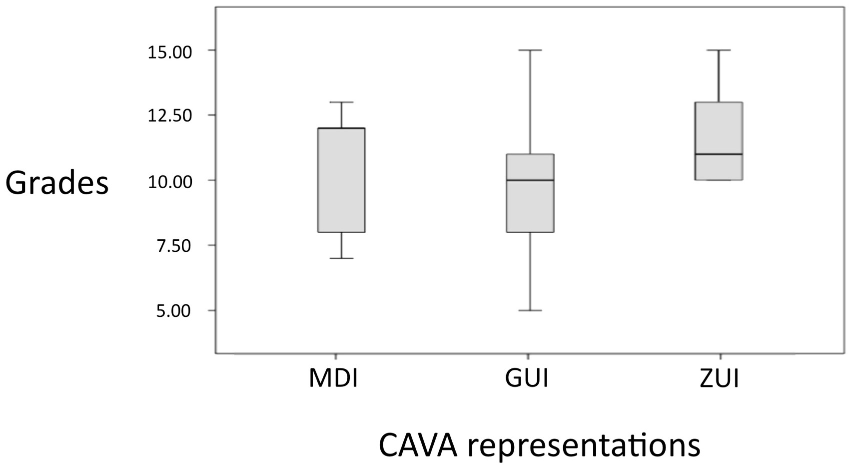


Figure 2.37. The difference in the mean for the three CAVA representations in the delayed post-test.

In relation to the research question posed in the current study, the difference in vocabulary acquisition efficiency among the three multimedia representations on the delayed post-test was less pronounced than that on the immediate post-test. However, the multimedia representation based on zoomable user interactivity (i.e. ZUI) was found to be the most effective multimedia representation, with a mean word retention of 45.3%. This was in contrast to the multimedia representations for GUI (40%) and MDI (37%). Figure 2.37. shows the difference of two points (7.7%) in the mean vocabulary recall scores between the most effective multimedia representation (ZUI: mean 11.8 points) and the least effective (GUI: mean 9.8 points). The raw scores for the participants in the delayed post-test can be seen in table 2.15.

Table 2.15. Raw scores for the five participants' results in the delayed post-test.

Participant	MDI	GUI	ZUI
S1	30.77	28.21	28.21
S2	20.51	25.64	25.64
S3	30.77	38.46	33.33
S4	17.95	12.82	25.64
S5	33.33	20.51	38.46

Inferential statistics for the delayed post-tests results

Multimedia representation types impacted equally on learners' scores in the delayed post-test and the immediate post-test, and were established with formal hypothesis testing, as having an alpha level of .05. However, there also remained a lack of any significant differences in the effectiveness of multimedia representations effecting the acquisition of second language vocabulary (F (2, 147) = 0.6, p = 0.5). Post-hoc multiple pairwise comparisons revealed that, over the long-term, multimedia representations of ZUI tended to be more effective tools for vocabulary acquisition than the multimedia representations MDI and GUI (see Table 2.16).

Table 2.16. Pairwise comparisons for multimedia representation in the delayed post-test.

Multimedia representation I	Multimedia representation J	Mean difference (I – J)	Standard error	Sig.
MDI	GUI	0.6	1.85	0.94
	ZUI	- 1.4	1.85	0.73
GUI	ZUI	- 2	1.85	0.54

AWMA results

The second research question posed in the study investigates whether working memory abilities effect vocabulary acquisition.

Descriptive statistics

The hypothesis posed for the second research question in the study suggests that learners with high-ability in verbal, visual, and spatial working memory will retrieve more Arabic words than low-ability learners when required to select and process verbal (MDI), visual (GUI), and spatial (ZUI) multimedia representations when using CAVA. Tables 2.17 and 2.18 present a correlation matrix of the participants' scores in the twelve AWMA subtests detailing short-term and working memory abilities, and their scores in CAVA's immediate and delayed post-tests respectively.

Table 2.17. A correlation matrix of twelve short-term and working memory test scores and the learners' scores in the immediate post-test for the second pilot study.

Immediate		Digit recall	Word recall	Non word recall	Listen-ing recall	Count-ing recall	Back wards Digit recall	Dot matrix	Mazes memory	Block recall	Odd one out	Mister X	Spatial recall
MDI	N	5	5	5	5	5	5	5	5	5	5	5	5
	Pearson	.84	.87	.57	.73	.53	.86	.5	.018	.42	.69	-.73	-.08
	Sig.	.071	.054	.31	.15	.35	.05	.38	.97	.47	.19	.15	.89
GUI	N	5	5	5	5	5	5	5	5	5	5	5	5
	Pearson	**.91***	**.94***	.57	.84	.37	.82	.6	.18	.56	.76	.68	.008
	Sig.	.028	.015	.31	.07	.53	.08	.27	.76	.32	.13	.2	.99
ZUI	N	5	5	5	5	5	5	5	5	5	5	5	5
	Pearson	**.98****	**.99****	.72	**.94***	.26	.68	.76	.41	.51	.68	.67	-.012
	Sig.	.003	.001	.170	.014	.67	.2	.31	.49	.37	.2	.21	.98

Table 2.18. A correlation matrix of twelve short-term and working memory test scores and the learners' scores in the delayed post-test for the second pilot study.

Delayed		Digit recall	Word recall	Non word recall	Listen-ing recall	Count-ing recall	Back wards Digit recall	Dot matrix	Mazes memory	Block recall	Odd one out	Mister X	Spatial recall
MDI	N	5	5	5	5	5	5	5	5	5	5	5	5
	Pearson	.62	.72	.19	.54	.59	**.98****	.40	-.24	.50	.86	-.41	.24
	Sig.	.26	.16	.75	.34	.29	.003	.49	.69	.38	.06	.48	.68
GUI	N	5	5	5	5	5	5	5	5	5	5	5	5
	Pearson	.76	.76	.20	.79	-.34	.38	.42	.55	**.95***	.69	-.34	.2
	Sig.	.13	.12	.73	.1	.57	.52	.47	.33	.01	.19	.57	.74
ZUI	N	5	5	5	5	5	5	5	5	5	5	5	5
	Pearson	.12	.30	-.33	.09	.57	.85	.16	-.56	.38	.8	.19	.6
	Sig.	.84	.62	.57	.88	.31	.06	.79	.32	.52	.1	.84	.28

The correlation matrix indicates that there are four significant correlations between the memory tests scores and the learners' scores in the immediate post-test, at the level of 0.05. The digit recall and the word recall tests were found to correlate significantly with the learners' CAVA scores for ZUI representation. Nevertheless, other significant correlations appear between GUI and digit recall, as well as word recall. In addition, listening recall processing is highly correlated with ZUI.

In the delayed post-test, there are three significant correlations, as shown in table 2.18, between the AWMA tests scores and the results for the participants in CAVA delayed post-test. Backwards digit recall is correlated with MDI significantly, and block recall with GUI. The odd one out processing test is also correlated with MDI.

To conclude, CAVA has been shown to be a user-friendly platform, delivering thorough and detailed instructions to participants. The pilot study also confirms that the content and information is presented in a format that participants recognize as logical and intuitive. Both the formative and summative evaluative pilot studies were critical to allow revision and an update of the initial design of CAVA to better meet the needs of foreign language learners at beginner level. As a result, the participants in the main study will not experience any difficulties interacting with all the phases of CAVA.

After introducing the instruments that were designed and selected, and explaining how they were used to assess learners' achievements and cognitive abilities, and to deliver and communicate L2 words, the following subsections of this chapter present a description of the methods used in the first study. It describes the participants, then introduces the experimental design before presenting the intervention. Finally, the process of collecting, handling, and analysing the data will be explained.

2.2.3 Participants

The first main study participants in this research were fifty monolingual British undergraduate students at the University of York (of which 30 were females and 20 males). None of the participants had previously learned Arabic and knew nothing about it. One participant was eliminated because he did not attend the delayed post-test. In addition, all the final participants had avoided studying the target vocabulary between the immediate post-test and the delayed post-test. The participants' ages ranged between 18 to 23 years old, with a mean of 20. Regarding their chief characteristics, the participants were homogeneous. They were all undergraduates at the University of York and either in their late teens or early twenties. The participants also have the same language background, were all British and had met the undergraduate entry requirements set by the University. None of the participants had previous knowledge of Arabic, and none had ever been to an Arabic-speaking country. The participants were also homogeneous regarding their degree of knowledge and proficiency required to use the computer, with all indicating that they were very comfortable using the software.

2.2.4 Design of the experiment

The experiments investigated the efficacy of three multimedia representations (i.e. verbal-based (L2-L1: MDI), visual-based GI (L2-Picture: GUI), and spatial-based (L2-Context: ZUI), for vocabulary acquisition, while controlling for working memory abilities and variations among learners. For this, the CAVA software presented the participants with 24 words, which were divided into six lessons (four words in each lesson). The study aims to answer the following research questions:

1- What is the effect of the three different interactive representation types; namely verbal-based (L2-L1: MDI), visual-based (L2-Picture: GUI) and spatial-based (L2-Context: ZUI), on second language vocabulary acquisition?

2- What is the effect of individual differences on verbal and visuospatial working memory and of short-term memory capacity on the initial uptake of brand new vocabulary?

Two independent variables set at different levels were considered in this study: multimedia representation type and individual differences in working memory. See table 2.19 for more details concerning these variables.

Table 2.19. Independent variables in the experiment.

Independent variable	Type	Levels	Measure
Multimedia representation	Within-subjects	Three levels: MDI, GUI, ZUI	Nominal
Individual differences in working memory	Within-subjects	Four levels: Verbal working memory, Visuospatial working memory, Verbal short-term memory, Visuospatial short-term memory,	Nominal

The dependent variable in this study was vocabulary acquisition. The students' acquisition of the target words was operationalized based on the scores that they received in the immediate and delayed post-tests. The tests were identical, and the scores for each ranged from 0 – 39 points per participant.

To answer the research questions using CAVA, it is important to ensure the distribution of the three representations so that all the words are balanced, not only in the study phases, but also in the practice exercises phase. There are three main objectives here: first, all learners must be exposed to all three multimedia representations, to allow a fair comparison of the efficiency of each representation for each learner. Second, each multimedia representation must be assessed with all the target words, to control for any potentially confounding effect from the learning burden imposed on the words through representational efficiency. Third, each representation must be displayed to learners with different working memory capacity, so as to permit an analysis of the role of working memory relative to the effectiveness of the representation. In order to attain these objectives, a Latin square within-subjects experimental design was used herein. It is a form of a partial counterbalancing design that is essential to control for the effects of practice and fatigue (see table 2.20).

Table 2.20. Example of a Latin square.

A	B	C
B	C	A
C	A	B

In a Latin Square experimental design, the measurements under any condition are repeated on the other conditions for each participant. In the current study, the experimental design of the

Latin Square was employed to present each treatment (multimedia representation) exactly once in each column, and on each row to control for two sources of extraneous variation (word and learner) (see Table 2.21).

Table 2.21. Multimedia representation distribution in the Latin Square design of the study.

Lesson	Exposure Group 1 (CAVA 1)	Exposure Group 2 (CAVA 2)	Exposure Group 3 (CAVA 3)
Office	ZUI	MDI	GUI
Food	GUI	ZUI	MDI
Occupations	MDI	GUI	ZUI
Clothes	ZUI	MDI	GUI
Kitchen	GUI	ZUI	MDI
Class	MDI	GUI	ZUI

In the Latin square design of the current study, each lesson was studied in all multimedia representations by different participants, each of whom varied in their working memory abilities. For each participant, the multimedia representation of the word was consistent across the CAVA phases (study phases and practice phase). Moreover, 24 words were split thematically into 6 lessons in the Latin square design. The overall learning burden for all the lessons is approximately similar, since the linguistic features of the chosen words were matched.

2.2.5 Intervention

In the current study, there were three exposure groups, combining the lessons and multimedia representations. The participants with different working memory abilities were divided randomly into the groups. The presence of three exposure groups, were formed from a combination of the lessons and multimedia representations, into which the participants were divided in a random manner. The participants in Group 1 were exposed to four words in each of the following: in lesson 1 (Office) by means of the zoomable representation (ZUI); in lesson 2 (Food) by means of graphical representation (GUI); and in lesson 3 (Occupations) by means of a menu driven representation (MDI). They were then exposed to four words in lesson 4 (Clothes) via the zoomable representation (ZUI), four words in lesson 5 (Kitchen) via the graphical representation (GUI), and four words in lesson 6 (Class) via the menu driven representation (MDI).

In Group 2, the participants were exposed to four words in lesson 1 (Office) via the menu driven representation (MDI), in lesson 2 (Food) via the zoomable representation (ZUI), and in lesson 3 (Occupations) via the graphical representation (GUI). These were then followed by exposure to four words in lesson 4 (Office) via the menu driven representation (MDI), four words in lesson 5 (Food) via the zoomable representation (ZUI), and four words in lesson 6 (Occupations) via the graphical representation (GUI).

Finally, in Group 3, the participants were exposed to four words in lesson 1 (Office) via the graphical representation (GUI), four words in lesson 2 (Food) via the menu driven representation (MDI), and four words in lesson 3 (Occupations) via the zoomable representation (ZUI). This was followed by exposure to four words in lesson 4 (Office) via the graphical representation (GUI), four words in lesson 5 (Food) via the menu driven representation (MDI), and four words in lesson 6 (Occupations) via the zoomable representation (ZUI).

2.2.6 Procedure

After making all the required editions to the study apparatus, based on the results from the subsequent pilot studies, the initial main study was conducted. It consisted of two parts. The first part comprised of a collection of memory data over a six week period in the summer term of the academic year 2015/2016, beginning Wednesday 4th of May 2016, and ending on Friday 10th of June. The second part of the study involved collecting data from the vocabulary learning program. It took place in the autumn term of the academic year 2016/2017, beginning on Tuesday, 11th of October 2016, with a learning phase, followed by an immediate post-test ending on Wednesday 30th of November after the delayed post-test. The participants involved in this study were also undergraduate students at the University of York, who had been contacted to participate in this study as described in reference to the first pilot study. Three meetings were required with each participant to conduct the research experiment. At the first meeting, the researcher introduced the participant to the experiment offering a clear explanation. The participant was also informed that the research related to assessing working memory abilities and vocabulary acquisition, without giving any details about specific research questions. Then, the researcher asked the participants to read and sign an ethical consent form, before testing the participants' working memory capacity using sections 1 through 12 of the Automated Working Memory Assessment (AWMA) (Alloway, 2007). In each subtest of the AWMA all the participants received identical instructions, which were provided by the program and related to how to respond to the questions put on the test. Each participant was tested individually in a quiet room at the University of York, and the test was completed in a single session and lasted from 45 to 1:30 minutes. All twelve subtests of the AWMA were conducted, so as to fully assess the participants' memory abilities. These measures were employed to test different memory components. Six of the subtests were used to assess storage-plus-processing components, and were called working memory tasks. Of these, three subtests investigated the verbal memory span and the other three were designed to measure visuospatial abilities. The remainder of the twelve subtests of the AWMA involved storage-only factors, and short-term memory tasks. Among these, three tasks focused on verbal ability and three on visuospatial ability. The AWMA test was conducted individually for each participant, and delivered through a computer with 22-inch display, screen resolution set to 600 x 480 pixels. The researcher himself tested all participants. The researcher sat in the same room and used a linked laptop to monitor the testing process. All twelve subtests of the long form of AWMA were automatically administered in a fixed sequence, to avoid any variation in task demands and to reduce fatigue.

The AWMA is fully automatic, and was designed using Barland's C11 Builder 5 (2004). The pictures used in some of the AWMA subtests were created using Microsoft Powerpoint, and the audio files were recorded with a minidisk player and edited using GoldWave (2004). The researcher reminded the participants of the overall aim of the test and then allowed the individuals to start the AWMA program. In each subtest of AWMA, oral instructions were presented while the computer screen was blank, and then followed by practice trials. After this, the main trials for each test were displayed in a series of blocks with six trials in each block. The participants' responses were recorded by the researcher on his laptop, using the right arrow key (>) or the left arrow key on the keyboard (<) for correct and wrong responses, respectively. A correct answer was automatically credited with a score of 1, and when a participant gives four correct answers out of six in a block, they are automatically moved to the following block. Yet if the participant makes three or more errors in a block of trials, the subtest was automatically ended, and the program returned the participant to the main screen. This is then considered as the end of the subtest, and the number of accurate answers then provides the test score. At the end of the program, the data was presented in a Word file to be downloaded onto the researcher's computer. The data was also transferred into SPSS for further analysis.

At the second meeting, the participant was exposed to CAVA and completed its two sessions. Each session consists of three lessons, including two study phases and the practice phase for each lesson. The participant also completed the immediate post-test, and in the third meeting, the delayed CAVA post-test.

The immediate post-test (follow-up test 1)

An immediate post-test was conducted after completing the second study phase. Once the participant completed the immediate post-test, they were informed of their post-test score by email on the same day, giving no details of the correct answers, as a way to avoid post-treatment learning. All the learners were given their participation costs (8 pounds per hour), and the best students received a reward (10 pounds).

Delayed post-test (follow-up test 2)

A month after the immediate post-test (follow-up test 1), the learners participated in a delayed post-test. The delayed post-test was identical to the immediate post-test. The participants were not informed that there would be a delayed post-test, they were just asked to meet the researcher for an interview. Once the participant arrived, they were given the post-test and received another reward.

All the sessions were conducted at the Research Centre of Social Sciences (ReCSS) building at the University of York, following the same procedure as in the two pilot studies. The main modification to the experimental protocol, in addition to editing and revising CAVA, involved applying the (advanced) Latin square to eliminate the recency effect of the last lesson. The students received eight pounds per hour, even if they did not attend all the sessions. Each participant earned their reward for participating, and all were informed of this before taking part. They also received soft copies of their working memory test analysis and their CAVA post-test scores.

2.2.7 Analysis

The immediate and delayed post-test CAVA scores were addressed to tackle the research questions. Both tests were identical and consisted of receptive recall, and recognition of multiple-choice questions, in addition to one productive recall question to which the participants responded by typing the correct word. When scoring this question, the participants received 0 points, 0.5 points or 1 point for each word. 0 points were received if the participant did not write the word down, or if there were more than two errors in the string of items. An error counted as any omission or substitution of a single letter in the target word. An example of a 0 point response would be when a participant writes **(فكيت)** for the target response **(مكتب)** (desk). If the participant writes the target word with one or two changes, they receive a score of 0.5 points. For example, the student response is **(مكيب)** for the target response **(مكتب)** *(desk)*. If the response is identical to the target item, it is counted as correct and receives 1 point.

Relevant effects had been tested, and inference tests used to address the two research questions, as shown in Table 2.22.

Table 2.22. Research questions and the inference test for the first study.

Research question	Effect tested	Inference test
1- What is the effect of three different interactive representation types; namely verbal-based (L2-L1: MDI), visual-based (L2-Picture: GUI) and spatial-based (L2-Context: ZUI) on second language vocabulary acquisition?	Main effect multimedia representation types	ANOVA (α = .05)
2- What is the effect of individual differences in verbal, visuospatial working memory and short-term memory capacity on the initial uptake of a brand new L2 vocabulary?	Main effect of working memory abilities	Pearson Correlation

The level for alpha was set to .05, to determine the statistical significance of all the tests. The inferential results for the first research question posed in the study were answered by conducting a one way within-subjects (repeated measures) ANOVA, with multimedia representation as the within-subject factor using SPSS. To answer the second question, a Pearson Correlation was conducted using SPSS to determine if there is a correlation between any of the twelve subtests of AWMA, and the CAVA results on both a short- long- term basis.

2.2.8 Summary

This section presented the methodology used in the first study to answer the first two of the research questions. Overall, the choice of the research design and methodology for this study took account of similar work conducted by other researchers in the field. A mixed methods design was adopted to tackle the different requirements associated with the research questions. The AWMA test was chosen as a readily available and reliable means of identifying Cognitive abilities. The second main apparatus targeted to generate comparative data (CAVA) was designed by the researcher and contained three multimedia representations, which were tested and compared. The main aim in employing different output measures (immediate, and delayed post-tests) along with AWMA, is to explore the effect of educational software as used in the study on a short- and long-term basis. The next section presents the results from the first study.

2.3 Results

The aim of this section is to present answers to the research questions listed previously in the methodology section. The results of the data analysis included as the first main study component of the research will be presented, followed by a discussion section covering related topics. The primary purpose of this study is to examine how different multimedia vocabulary CALL representations could affect the learning process for novel L2 words. This section also describes a quantitative analysis of 50 participants involved in the first main study. It includes the participants' immediate and post-test results for the first version of the CAVA software (CAVA-1), followed by the results of the Automated Working Memory Assessment (AWMA). When tackling the research questions for the present study, the results of both the immediate and delayed CAVA post-tests will be correlated with the AWMA scores. The results are presented in the two following subsections, corresponding to the two topics of the study. The first section investigates vocabulary acquisition, showing the results related to the main effect of multimedia representation. The second section explores the results for the main effect of memory abilities; examining multimedia representation together with individual differences in working memory, as a way to investigate the interactive effect of multimedia representation and memory on vocabulary acquisition.

2.3.1 CAVA-1 results

The effectiveness of multimedia representations in CALL environments for second language vocabulary acquisition were investigated by examining the main effect from multimedia representations, across the three different interactive user interfaces (i.e. MDI, GUI, ZUI), on vocabulary acquisition.

2.3.1.1 Immediate Post-test

Figure 2.38. and table 2.23 show the main effect from multimedia representations on the immediate post-test for the 50 study participants. Each participant studied 4 words in each of the six lessons, culminating in a total of 24 words.

Table 2.23. Immediate post-test results for the three multimedia representation types.

Multimedia representations	Rank	Mean	Maximum score	Percentage	SD	Skewness	Kurtosis
MDI	3	15.46	26	59.46 %	18.6	-.023	-.706
GUI	2	16.2	26	62.31 %	17.79	-.394	-.773
ZUI	1	16.73	26	64.35 %	17	-.609	-.054
All multimedia representations		48.39	78	62.03 %	17.83		

Figure 2.38. A histogram for the CAVA-1 immediate post-test.

Descriptive statistics

The learners, as shown in Table 2.23 above, achieved a mean vocabulary acquisition score of 62.03% (48.39 out of 78). The use of ZUI multimedia representation, comprising a zoomable interface and layout showing an overview of all the target words in the lesson, resulted in the highest mean score on the immediate post-test: a mean of 64.35 % (16.73 out of 26). Thus, it was found to best support vocabulary acquisition in the immediate post-test. The second best multimedia representation was GUI (graphic based L2-picture user interface), with a mean of 62.31 %, followed by (MDI), a type of verbal based L2-L1 user interfaces with 59.46%.

The following table (2.24) as well as figure (2.39) show the distribution of the scores for the three multimedia representations (MDI, GUI, and ZUI) in the immediate post-test. The histograms provide evidence that the data was normally distributed in all cases, as was also indicated by Skewness and Kurtosis values (values of Skewness and Kurtosis within ± 2, indicate normal distribution (George & Mallery, 2010).

Table 2.24. Descriptive statistics for the scores in MDI, GUI, and on the ZUI immediate post-test.

	MDI	GUI	ZUI
N	50	50	50
Std. Error of Mean	2.63613	2.51670	2.41334
Std. Deviation	18.64023	17.79575	17.06489
Variance	347.458	316.689	291.210
Skewness	-.023	-.394	-.609
Std. Error of Skewness	.337	.337	.337
Kurtosis	-.706	-.773	-.054
Std. Error of Kurtosis	.662	.662	.662
Range	75.00	67.31	73.08
Minimum	19.23	23.08	19.23
Maximum	94.23	90.38	92.31

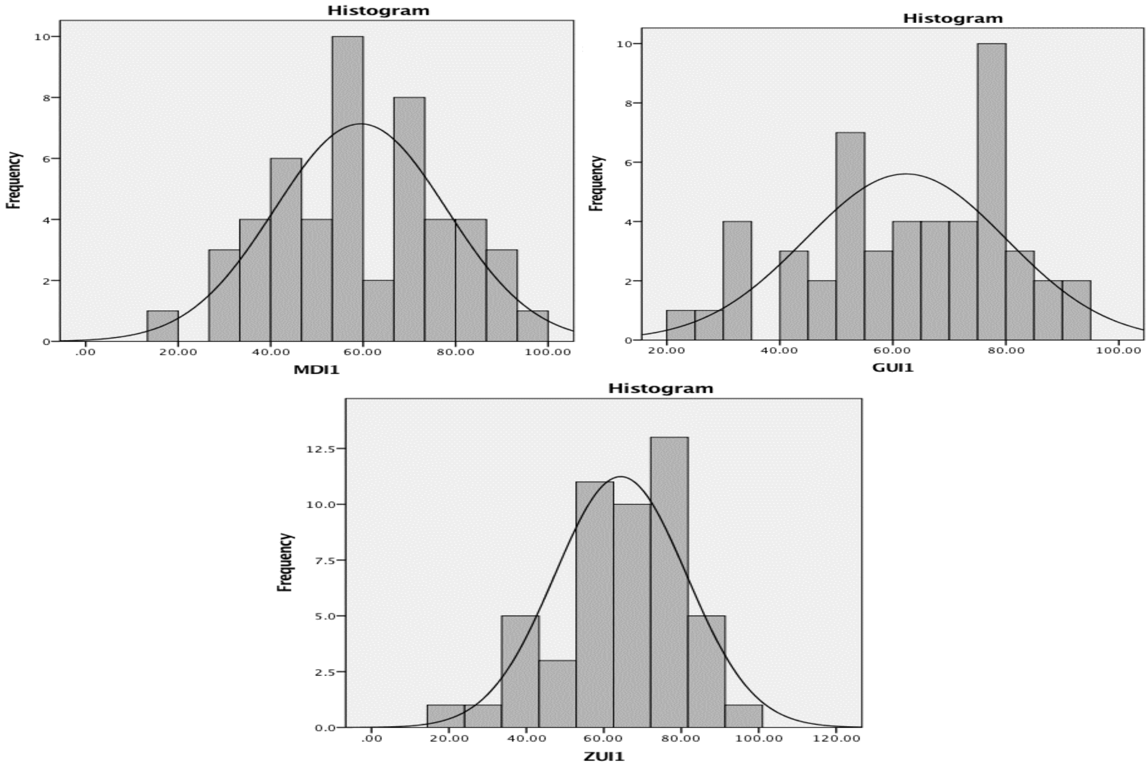


Figure 2.39. Frequency histograms showing the distribution of the participants' immediate post-test scores in the three multimedia representations.

Inferential statistics

On the immediate CAVA-1 post-test, a repeated measures ANOVA was conducted, and a main effect from multimedia representation was tested in formal hypothesis testing at an alpha level of .05. Since a repeated measures ANOVA was used, Mauchly's test of sphericity was conducted to evaluate the assumption of sphericity; in particular the assumption that the relationship between the three multimedia representations is similar (Field, 2012). Table 2.25 presents the results, indicating that the assumption of sphericity had not been violated: there was no statistically significant difference here $\chi2$ (2) = .175, p = .91.

Table 2.25. Mauchly’s test of sphericity.

Mauchly’s test of sphericity			Epsilon		
Approx. Chi- square	Df	Sig	Greenhouse-geisser	Huynh-feldt	Lower-bound
.175	2	.91	.99	1	.5

Moreover, in line with the outcome of Mauchly’s test of sphericity, the results for the Huynh-Feldt correction, as shown in Table 2.26, revealed no statistically significant difference in the participants’ vocabulary acquisition via the three multimedia representations, *F* (2) = 2.2, *p* = .109. This means the variances in the differences between all combinations of the three groups of multimedia representations are equal.

Table 2.26. Test of the within subjects effect (Hyunh-Feldt) to correct sphericity.

Hyunh-feldt	df	F	Sig
	2	2.2	.109

Nevertheless, the effectiveness of the multimedia representation types on the acquisition of second language vocabulary does not differ significantly. The Bonferroni post-hoc multiple pairwise comparison results were in line with the descriptive statistics. The results suggest a trend towards multimedia representation ZUI being more effective for vocabulary acquisition than the multimedia representations MDI and GUI. The differences, however, were not statistically significant F (2, 147) = 0.9, p = 0.3 > .05 (See Table 2.27).

Table 2.27. Results of repeated measures ANOVA (Pairwise), comparing the scores for the three conditions, MDI, GUI, and ZUI on the CAVA-1 immediate post-test.

Multimedia representation (I)	Multimedia representation (J)	Mean difference (I – J)	Standard error	Sig.
ZUI	MDI	4.88	3.5	0.17
	GUI	2	3.5	0.57
GUI	MDI	2.88	3.5	0.41

2.3.1.2 Delayed Post-test

To ensure that all the participants had not deliberately studied the target vocabulary between the CAVA-1 immediate and delayed post-tests, they were not told if there would be any more tests following the immediate post-test.

Descriptive statistics

CAVA-1 delayed the post-test results for the experiment, which are shown in Table 2.28. The mean word retention on the delayed post-test (37%, or 28.79 / 78 points) was lower than that for the immediate post-test (62.03%, or 48.39 / 78 points).

Table 2.28. Delayed post-test results for the three multimedia representation types.

Multimedia representations	Rank	Mean	Maximum score	Percentage	SD	Skewness	Kurtosis
MDI	2	9.5	26	36.57 %	11	559	.007
GUI	1	9.9	26	38.46 %	13.8	1.062	2.051
ZUI	3	9.39	26	36.15 %	13.4	.363	.414
All multimedia representations		28.79	78	37 %	12.7		

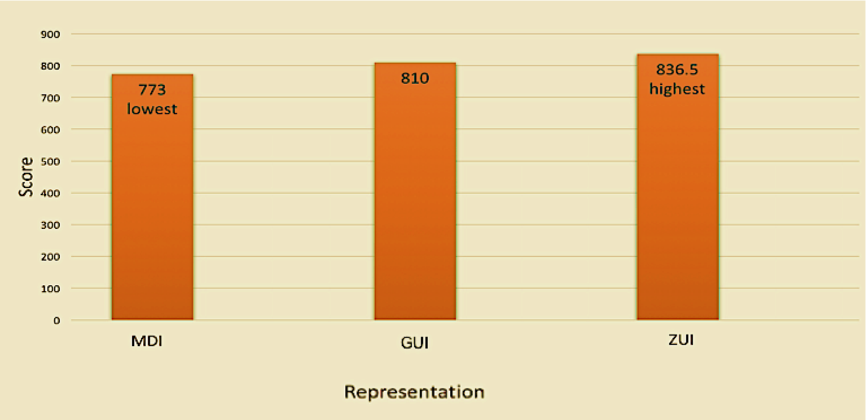


Figure 2.40. A histogram of the CAVA-1 delayed post-test.

With regard to the first research question for the study, the difference in vocabulary acquisition efficiency among the three multimedia representations for the delayed post-test were less pronounced than for the immediate post-test. However, the multimedia representation that presented the word with a picture (GUI) seemed the most effective multimedia representation, with a mean word retention of 38.46%, whereas the multimedia representations MDI and ZUI were less effective, with a mean word retention of 36.57% and 36.15%, respectively (see also figure 2.40). The difference in mean vocabulary recall scores between the most effective multimedia representation (GUI: mean 9.9 points) and the least effective (ZUI: mean 9.39 points) was 0.61 points (2.3 %).

The following table (2.29) and figure (2.41) visually display the distribution of scores in the three multimedia representations (MDI, GUI, and ZUI), for the delayed post-test. The histograms provide evidence that the data was normally distributed in all cases, as supported by Skew and Kurtosis values, as shown in the table.

Table 2.29. Descriptive statistics for the scores in MDI, GUI, and the ZUI delayed post-test.

	MDI	GUI	ZUI
N	50	50	50
Std. Error of Mean	1.56713	1.95959	1.90256
Std. Deviation	11.08127	13.85640	13.45313
Variance	122.795	192.000	180.987
Skewness	.559	1.062	.363
Std. Error of Skewness	.337	.337	.337
Kurtosis	.007	2.051	.414
Std. Error of Kurtosis	.662	.662	.662
Range	48.08	69.23	65.38
Minimum	19.23	19.23	11.54
Maximum	67.31	88.46	76.92

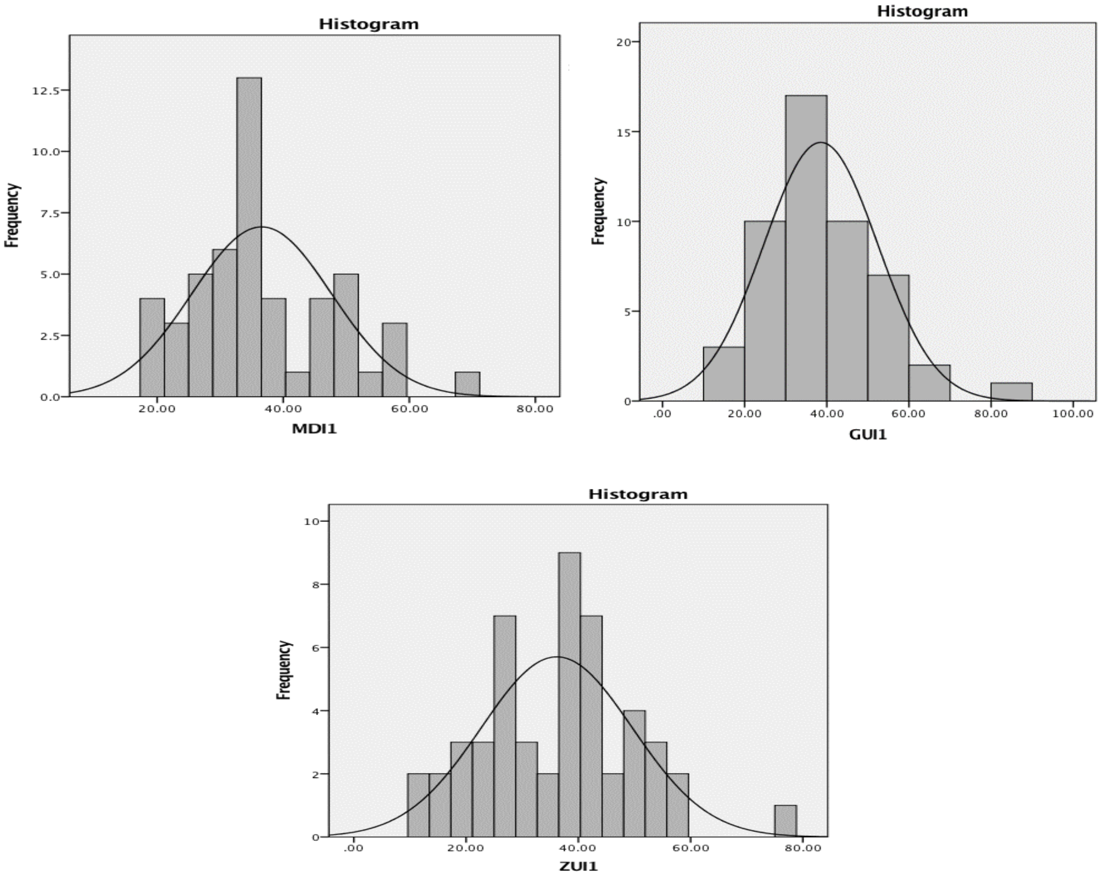


Figure 2.41. Frequency histograms showing the distribution of the participants' delayed post-test scores in the three multimedia representations.

Inferential statistics

The effect of multimedia representation type on learners' scores was again notable in the delayed post-test, as it was for the immediate post-test. It was tested in formal hypothesis testing with an alpha level of .05. Since a repeated measures ANOVA was used, Mauchly's test of sphericity was again employed to test the assumption that the relationship between the three multimedia representations would be similar (Field, 2012). Table 2.30 presents the results, indicating that the assumption of sphericity would not be violated; moreover, there was no statistically significant difference, $\chi 2$ (2) = 4.89, *p* = .086.

Table 2.30. Mauchly's test of sphericity.

Mauchly's test			Epsilon		
Approx. Chi- square	Df	Sig	Greenhouse- geisser	Huynh-feldt	Lower- bound
4.89	2	.086	.91	.94	.5

Moreover, the results of the Hyunh-feldt agreed with the results for Mauchly's test of sphericity. Table 2.31 shows there was no statistically significant difference in the participants' vocabulary acquisition across the three multimedia representations, *F* (1.8) = .88, *p* = .4. This means that the variances in the differences between all combinations of the three groups across the multimedia representations proved equal.

Table 2.31. Test of within subjects effect (Hyunh-Feldt) to correct sphericity.

Hyunh-feldt	df	F	Sig
	1.8	.88	.4

However, the effectiveness of the multimedia representations when detailing the acquisition of second language vocabulary still shows no significant difference. The results for the Bonferroni post-hoc multiple pairwise comparisons were in line with the descriptive statistics. The results suggest a trend towards the multimedia representation of GUI being more effective for long term vocabulary acquisition than for the multimedia representations MDI and ZUI. The differences, however, were not found to be statistically significant F (2, 147) = 0.4, p = 0.6 > .05 (See Table 2.32).

Table 2.32. Results of repeated measures ANOVA (Pairwise) comparing the scores for the three conditions, MDI, GUI, and ZUI on the delayed post-test.

Multimedia representation (I)	Multimedia representation (J)	Mean difference (I – J)	Standard error	Sig.
ZUI	MDI	-0.42	2.57	0.87
	GUI	-2.3	2.57	0.37
GUI	MDI	1.88	2.57	0.46

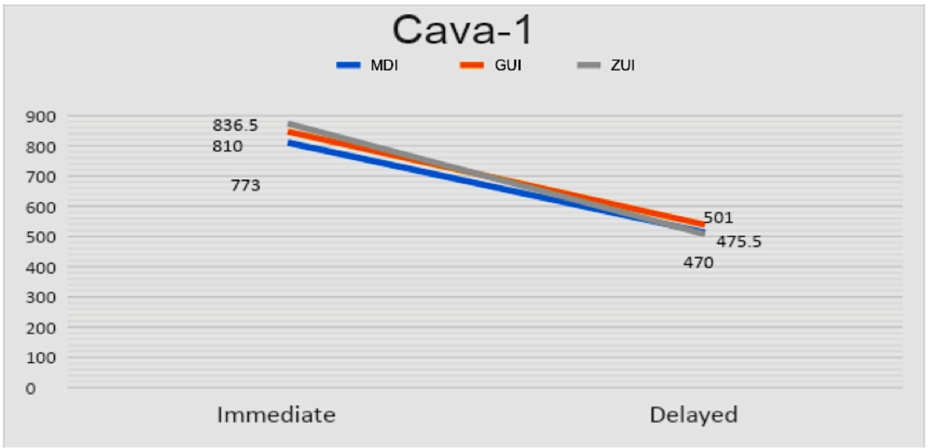


Figure 2.42. A line chart showing the CAVA-1 results for all three representations in both the immediate and delayed post-tests.

Figure 2.43. A bar chart showing the CAVA-1 results for all three representations in both the immediate and delayed post-tests.

2.3.1.3 Summary of findings pertaining to research question 1

The results of the analysis to address the first research question of the first main study, as shown in figures 2.42 and 2.43 above, revealed that ZUI condition scores tend to produce the highest mean in comparison with the other two conditions in the immediate post-test. However, the difference was not statistically significant F (2, 147) = 0.9, p = 0.3 > .05. The results for the delayed post-test revealed a dramatic change, and the ZUI condition led to the lowest mean vocabulary acquisition scores. However, no statistically significant difference emerged between the scores of the conditions at F (2, 147) = 0.4, p = 0.6. Several potential interpretations will be examined in the discussion section to explain why the participants achieved the scores they did for all types of representations. The following section presents the results for the correlation analysis of memory abilities of the participants with their CAVA-1 scores, to assist in interpreting how the participants interacted with the multimedia representations in the current vocabulary CALL environment.

2.3.2 AWMA results

The second research question of this study investigates whether short-term and working memory abilities influence vocabulary acquisition. The hypothesis posed in relation to the second research question stated that learners with high-ability in verbal short-term memory, visuospatial short-term memory, verbal working memory, or visuospatial working memory, will be more likely to retrieve more Arabic words than low-ability learners when required to select and process any (MDI), (GUI), and (ZUI) multimedia representations when using CAVA. Tables 2.33 and 2.34 present a correlation matrix of the participants' scores in the AWMA's twelve subtests of short-term and working memory abilities, and in relation to the participants' scores in CAVA-1's immediate and delayed post-tests respectively.

Table 2.33. A correlation matrix of twelve short-term and working memory test scores and the learners' scores in the CAVA-1's immediate post-test.

Immediate		Digit recall	Word recall	Non word recall	Listen-ing recall	Count-ing recall	Back wards Digit recall	Dot matrix	Mazes memory	Block recall	Odd one out	Mister X	Spatial recall
MDI	N	50	50	50	50	50	50	50	50	50	50	50	50
	Pearson	.165	.059	-.059	**.309***	.189	**.336***	.138	.255	.209	**.303***	.064	**.331***
	Sig.	.252	.683	.682	.029	.188	.017	.339	.074	.144	.032	.659	.019
GUI	N	50	50	50	50	50	50	50	50	50	50	50	50
	Pearson	.074	.077	.068	.259	.100	.169	.176	.263	.081	**.368****	-.007	.069
	Sig.	.608	.596	.639	.070	.491	.240	.222	.066	.576	.009	.963	.633
ZUI	N	50	50	50	50	50	50	50	50	50	50	50	50
	Pearson	**.337***	.236	.237	**.365****	.143	**.372****	**.286***	.262	.115	**.392****	.079	.253
	Sig.	.017	.099	.097	.009	.320	.008	.044	.066	.426	.005	.585	.076

Table 2.34. A correlation matrix of twelve short-term and working memory test scores and the learners' scores in the CAVA-1's delayed post-test.

Delayed		Digit recall	Word recall	Non word recall	Listen-ing recall	Count-ing recall	Back-wards Digit recall	Dot matrix	Mazes memory	Block recall	Odd One out	Mister X	Spatial recall
MDI	N	50	50	50	50	50	50	50	50	50	50	50	50
	Pearson	-.004	.002	-.148	.078	.057	.211	.001	.172	.221	.155	.103	.159
	Sig.	.981	.991	.306	.588	.692	.142	.994	.234	.123	.283	.478	.270
GUI	N	50	50	50	50	50	50	50	50	50	50	50	50
	Pearson	-.152	-.100	-.276	-.021	-.059	-.038	.032	**.451****	.146	.121	-.044	.024
	Sig.	.292	.488	.052	.883	.682	.796	.823	.001	.312	.402	.759	.870
ZUI	N	50	50	50	50	50	50	50	50	50	50	50	50
	Pearson	.052	.123	-.052	.010	.033	.068	.065	.098	-.011	-.048	.178	.146
	Sig.	.718	.394	.717	.947	.819	.637	.655	.500	.938	.741	.217	.313

2.3.2.1 Correlational Analyses

The tables above show several AWMA subtests' results were significantly correlated with their CAVA-1 vocabulary acquisition scores in the immediate post-test, based on all the different teaching methods. The most striking test is the 'Odd one out', which is correlated with all the representation types. ZUI had the highest number of correlated memory tests: eight out of the AWMA twelve tests. In addition to the odd-one out, the ZUI vocabulary acquisition scores correlated with digit recall, listening recall, backward digital recall, and a dot matrix. GUI, on the other hand, was only found to be highly correlated with odd-one out task. Nevertheless, other significant correlations were apparent between the menu driven user interface, MDI, and the listening recall, counting recall, and spatial recall, in addition to the odd-one out. In the delayed post-test, however, the only correlation exhibited related to GUI, which appears to be significantly correlated with the mazes memory task. In the following sections, additional details outlining the results of each subtest of AWMA will be presented.

2.3.2.2 Analysis of mean scores for the four AWMA components and subtests

An overview of the differences of the mean scores for the four components of the AWMA will be presented in this section of the analysis. The differences will be analysed, and judgements made regarding their significance. The following table (2.35) shows the mean scores for the four components, and is followed by tables 2.36 and 2.37, which show the correlations between the AWMA measures (Verbal Short-Term Memory (VerbalS), Verbal Working Memory (VsS), Visuospatial Short-Term Memory (VerbalW), Visuospatial Working Memory (VsW), and learning performance in the CAVA-1 immediate and delayed post-tests across the three multimedia representations: MDI, GUI, and ZUI.

Table 2.35. Mean (by group) of the scores for the four components.

Group	N	Minimum	Maximum	Mean	Std. Deviation
Verbal short-term memory	50	66	136	107	15.57
Visuospatial short-term memory	50	68	141	103	14.1
Verbal working memory	50	60	148	100	15.3
Visuospatial working memory	50	79	137	108	15.3

Table 2.36. Pearson's Rank Correlations between the AWMA measures and learning performance in the CAVA-1 immediate post-test.

		VerbalS	VsS	VerbalW	VsW
MDI	Pearson	.166	.135	**.308***	**.301***
	Sig.	.249	.348	.029	.033
	N	50	50	50	50
GUI	Pearson	.072	.176	.258	**.371****
	Sig.	.621	.222	.070	.008
	N	50	50	50	50
ZUI	Pearson	**.338***	**.281***	**.369****	**.387****
	Sig.	.017	.048	.008	.006
	N	50	50	50	50

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Table 2.37. Pearson's Rank Correlations between the AWMA measures and learning performance in the CAVA-1 delayed post-test.

		VerbalS	VsS	VerbalW	VsW
MDI	Pearson	-.004	.001	.078	.155
	Sig.	.981	.994	.588	.283
	N	50	50	50	50
GUI	Pearson	-.152	.032	-.021	.121
	Sig.	.292	.823	.883	.402
	N	50	50	50	50
ZUI	Pearson	.052	.065	.010	-.048
	Sig.	.718	.655	.947	.741
	N	50	50	50	50

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Verbal short-term memory subtests

The digit recall task:

In this task the participant hears a sequence of digits, and is asked to recall each sequence in the correct order.

Table 2.38. Mean of the scores for the digit recall task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Digit recall	50	73	126	100.9	15.1

Table 2.39. Correlation between the digit recall task score and the immediate and delayed CAVA-1 post-test under the three conditions.

		Digit recall (Immediate post-test)	Digit recall (delayed post-test)
MDI	N	50	50
	Pearson Correlation	.165	-.004
	Sig. (2-tailed)	.252	.981
GUI	N	50	50
	Pearson Correlation	.074	-.152
	Sig. (2-tailed)	.608	.292
ZUI	N	50	50
	Pearson Correlation	**.337***	.052
	Sig. (2-tailed)	.017	.718

As table 2.38 shows, the standard deviation for the mean of the Scores for the digit recall task is 15.1, shows huge differences between the participants' verbal short-term abilities. However, table 2.39 shows no Pearson correlation pertaining to the digit recall task with any of the teaching methods, even in the immediate or delayed post-tests. The exception is the ZUI immediate post-test, and this does show a significant positive correlation at p = .017, with r = 0.337. The possible reasons for this exception will be provided in the discussion section.

The word recall task:

In this task, the participant hears a sequence of words, and is then asked to recall each sequence in the correct order.

Table 2.40. Mean of the scores for the word recall task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Word recall	50	66	131	104.7	15.7

Table 2.41. Correlation between the word recall task scores and the immediate and delayed CAVA-1 post-test in its three conditions.

		Word recall (immediate post-test)	Word recall (delayed post-test)
MDI	N	50	50
	Pearson Correlation	.059	.002
	Sig. (2-tailed)	.683	.991
GUI	N	50	50
	Pearson Correlation	.077	-.100
	Sig. (2-tailed)	.596	.488
ZUI	N	50	50
	Pearson Correlation	.236	.123
	Sig. (2-tailed)	.099	.394

Table 2.40 shows the Pearson correlation between the participants' scores on the word recall task, and their results in CAVA-1, in each of its three multimedia representations. The results indicate that the role of the word recall ability as a verbal short-term span working is almost similar when applied to learn about the different representations. However, as shown in table 2.41, the relationship between word recall and ZUI approaches a significant level of correlation (r= 0.236, p= .099).

The non-word recall task:

In this task the participant hears a sequence of non-words and is asked to recall each sequence in the correct order.

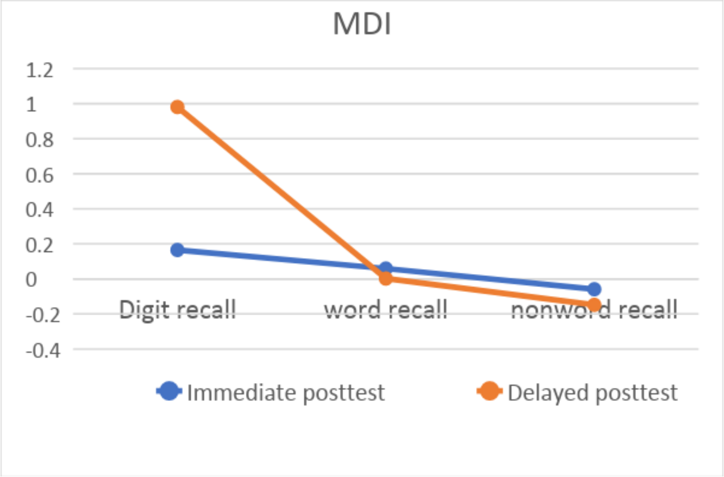
Table 2.42. Mean of the scores for the non-word recall task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Non-word recall	50	91	136	115.3	12

Table 2.43. Correlation between the non-word recall task score and the immediate and delayed CAVA-1 post-test in its three conditions.

		Non-word recall (immediate post-test)	Non-word recall (delayed post-test)
MDI	N	50	50
	Pearson Correlation	-.059	-.148
	Sig. (2-tailed)	.682	.306
GUI	N	50	50
	Pearson Correlation	.068	-.276
	Sig. (2-tailed)	.639	.052
ZUI	N	50	50
	Pearson Correlation	.237	-.052
	Sig. (2-tailed)	.097	.717

Table 2.42 shows the Pearson correlation between the participants' scores for the non-word recall task and their results in CAVA-1 in each of the three multimedia representations. The results show the role of the non-word recall ability as a verbal short-term span, which works differently when learning of the different representations. Table 2.43 shows the relationship between non-word recall and the immediate ZUI borderlines significant trend (r= 0, p= .097). Figure 2.44 illustrates the correlations of the three subtests of the verbal short-term memory, with the participants' vocabulary recall scores presented in the immediate and delayed CAVA-1 post-tests of all three representations.



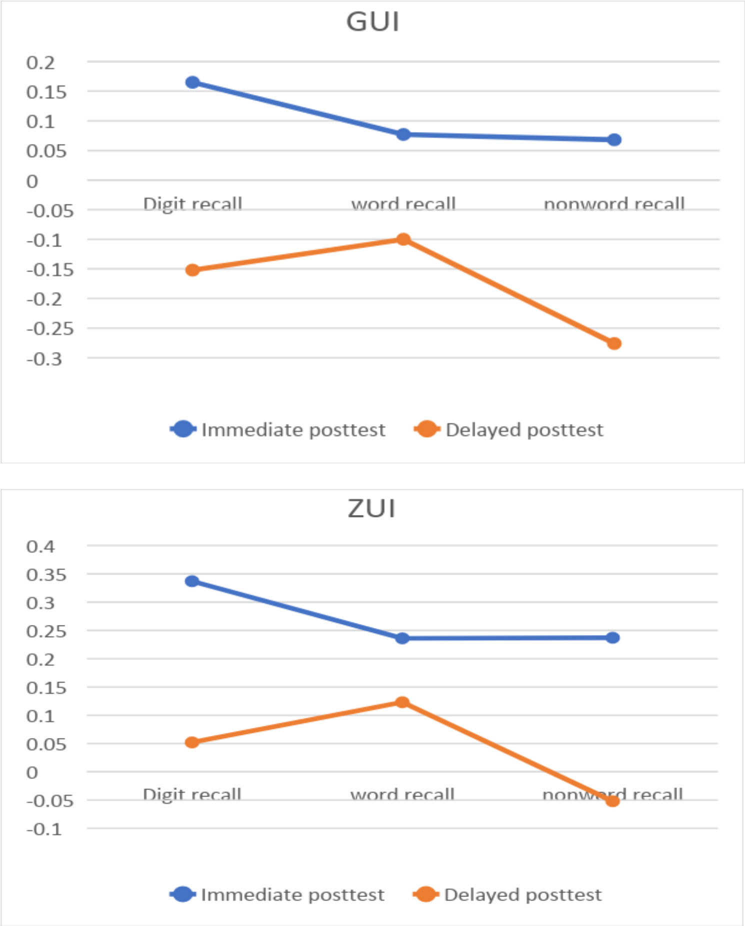


Figure 2.44. Correlations of the three subtests of the verbal short-term memory with the immediate and delayed CAVA-1 post-tests for all three representations.

Visuospatial short-term memory tests

The dot matrix task:

Each participant is presented with a red dot positioned in a series of 4 * 4 matrices for 2 seconds. They are then asked to remember the position, and indicate its place by pointing to the squares on the computer screen for the examiner to score.

Table 2.44. Mean of the scores for the dot matrix task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Dot matrix	50	77	141	107.6	14

Table 2.45. Correlation between the dot matrix task score, and the immediate and delayed CAVA-1 post-test in its three conditions.

		Dot Matrix (immediate post-test)	Dot matrix (delayed post-test)
MDI	N	50	50
	Pearson Correlation	.138	.001
	Sig. (2-tailed)	.339	.994
GUI	N	50	50
	Pearson Correlation	.176	.032
	Sig. (2-tailed)	.222	.823
ZUI	N	50	50
	Pearson Correlation	**.286***	.065
	Sig. (2-tailed)	.044	.655

Table 2.44 shows the Pearson correlation between the participants' scores in the dot matrix task, and their results in CAVA-1 in each of three multimedia representations. The results demonstrate that the ability to complete a dot matrix task, as a visuospatial short-term memory test, differs when learning different representations. Table 2.45 shows the relationship between the dot matrix and the ZUI immediate post-test scores, as these are statistically significant (r= 0.286, p= .044).

The mazes memory task:

In this task, the participant is presented with a maze. For three seconds the participant is shown a red line indicating the correct path. After three seconds the line disappears, and the participant is asked to trace the same path on the blank maze.

Table 2.46. Mean of the scores for the mazes memory task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Mazes memory	50	82	134	105.2	12.7

Table 2.47. Correlation between the mazes memory task score and the immediate and delayed CAVA-1 post-tests in its three conditions.

		Mazes memory (immediate post-test)	Mazes memory (delayed post-test)
MDI	N	50	50
	Pearson Correlation	.255	.172
	Sig. (2-tailed)	.074	.234
GUI	N	50	50
	Pearson Correlation	.263	**.451****
	Sig. (2-tailed)	.066	.001
ZUI	N	50	50
	Pearson Correlation	.262	.098
	Sig. (2-tailed)	.066	.500

Table 2.46 shows the Pearson correlation between the participants' scores in the mazes memory task, and their results in CAVA-1 for each of the three multimedia representations. The findings reveal the role of the ability to complete the mazes memory task, as a visuospatial short-term memory test, varies when learning different representations. Table 2.47 shows that the relationship between the mazes memory and the GUI delayed post-test score is statistically significant (r= 0.451, p= .001).

The block recall task:

In this task, the participant is shown a video of a series of blocks being tapped. They have to recall the sequence in the correct order.

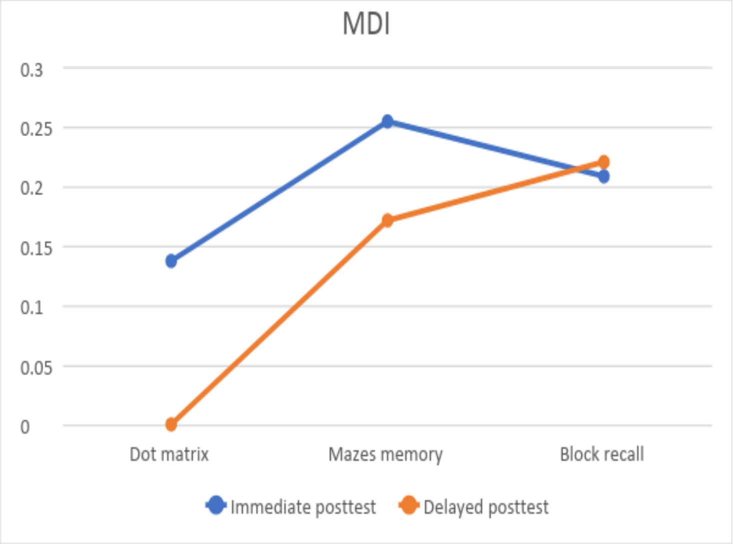
Table 2.48. Mean of the Scores for the block recall task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Block recall	50	68	121	97	13.5

Table 2.49. Correlation between the block recall task score and the immediate and delayed CAVA-1 post-test in its three conditions.

		Block recall (immediate post-test)	Block recall (delayed post-test)
MDI	N	50	50
	Pearson Correlation	.209	.221
	Sig. (2-tailed)	.144	.123
GUI	N	50	50
	Pearson Correlation	.081	.146
	Sig. (2-tailed)	.576	.312
ZUI	N	50	50
	Pearson Correlation	.115	-.011
	Sig. (2-tailed)	.426	.938

Table 2.48 shows the correlation for the block recall for all participants with the CAVA-1 scores. There is no evidence of a statistically significant correlation between the short-term span and other methods of teaching used (see table 2.49). Figure 2.45 illustrates the correlations of the three subtests of visuospatial short-term memory, with the participants' vocabulary recall scores in the immediate and delayed CAVA-1 post-tests for all three representations.



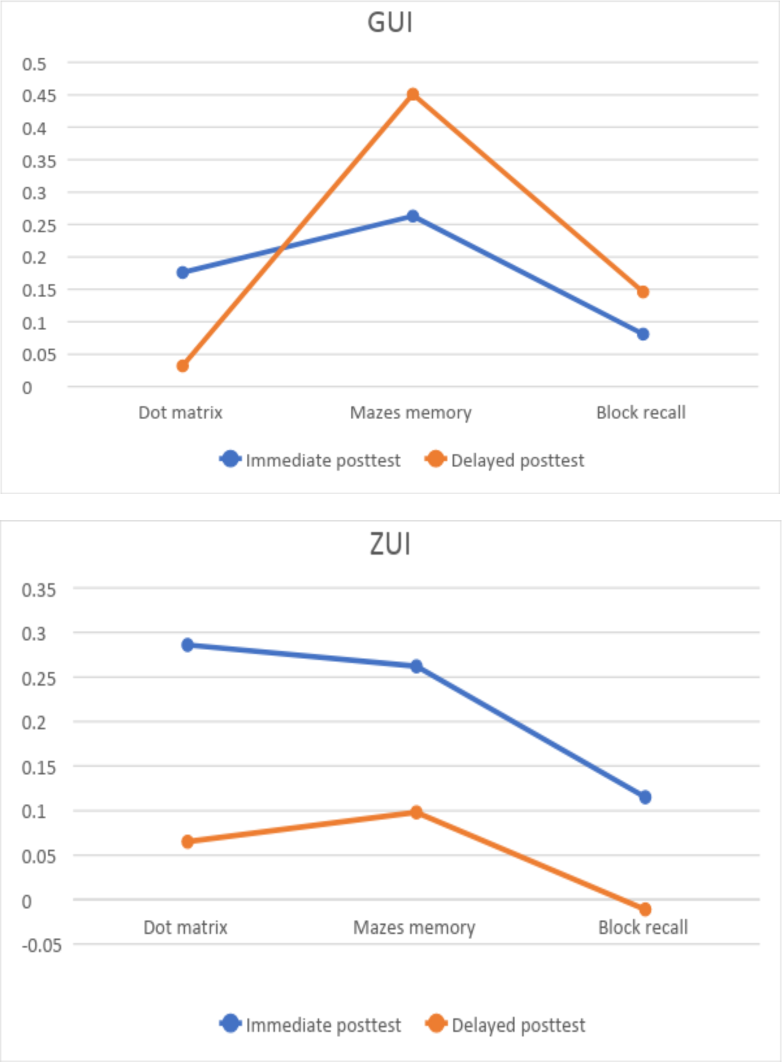


Figure 2.45. Correlations for the three subtests detailing visuospatial short-term memory with the immediate and delayed CAVA-1 post-tests for all three representations.

Verbal working memory tests

Listening recall:

The participant hears a series of spoken sentences and then has to judge if the sentence is *true* or *false*. Finally, they are asked to recall the last words of all the sentences sequentially.

Table 2.50. Mean of the scores for the listening recall task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Listening recall	50	77	129	100	15

Table 2.51. Correlation between the listening recall task score and the immediate and delayed CAVA-1 post-test in its three conditions.

Listening recall		immediate post-test	delayed post-test
MDI	N	50	50
	Pearson Correlation	.309*	.078
	Sig. (2-tailed)	.029	.588
GUI	N	50	50
	Pearson Correlation	.259	-.021
	Sig. (2-tailed)	.070	.883
ZUI	N	50	50
	Pearson Correlation	**.365****	.010
	Sig. (2-tailed)	.009	.947

While table 2.50 shows no correlation between the listening recall and other representations on a long term basis (in the delayed post-test), it is apparent from table 2.51 that listening recall has a fairly strong significant correlation with ZUI's immediate post-test scores (r = .365, *p* = .009), and positive significantly lower correlations for MDI and CUI.

The counting recall task:

In this task, the participant sees a visual array of blue triangles and red circles. The participant is required to count the red circles and then recall how many there are.

Table 2.52. Mean of the scores for the counting recall task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Counting recall	50	60	148	100.4	15.9

Table 2.53. Correlation between the counting recall task score and the immediate and delayed CAVA-1 post-tests in all three conditions.

Counting recall		Immediate post-test	Delayed post-test
MDI	N	50	50
	Pearson Correlation	.189	.057
	Sig. (2-tailed)	.188	.692
GUI	N	50	50
	Pearson Correlation	.100	-.059
	Sig. (2-tailed)	.491	.682
ZUI	N	50	50
	Pearson Correlation	.143	.033
	Sig. (2-tailed)	.320	.819

The results in tables 2.52 and 2.53 show no correlation between the counting recall task scores for any of the teaching representations, indicating that a good score in CAVA-1 would not necessarily mean a high score in counting recall.

The backwards digit recall task:

In this task, the participant hears a sequence of spoken digits and is required to recall them in reverse order.

Table 2.54. Mean of the scores for the backwards digit recall task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Backwards digit recall	50	70	130	99.6	15.2

Table 2.55. Correlation between the backwards digit recall task score and the immediate and delayed CAVA-1 post-test in its three conditions.

Backwards Digit recall		Immediate post-test	Delayed post-test
MDI	N	50	50
	Pearson Correlation	**.336***	.211
	Sig. (2-tailed)	.017	.142
GUI	N	50	50
	Pearson Correlation	.169	-.038
	Sig. (2-tailed)	.240	.796
ZUI	N	50	50
	Pearson Correlation	**.372****	.068
	Sig. (2-tailed)	.008	.637

The Pearson correlation as shown in table 2.55 indicates a strong positive correlation, depicting a closer relationship between the backward digit recall and ZUI immediate scores than in MDI: R = 0.372, *p* = .008. An average positive correlation between MDI and the backward digit recall R=0.336, *p* = .017 also exists. Figure 2.46 shows the correlations for the three subtests of verbal working memory, with the participants' vocabulary recall scores in the immediate and delayed CAVA-1 post-tests in all three representations.

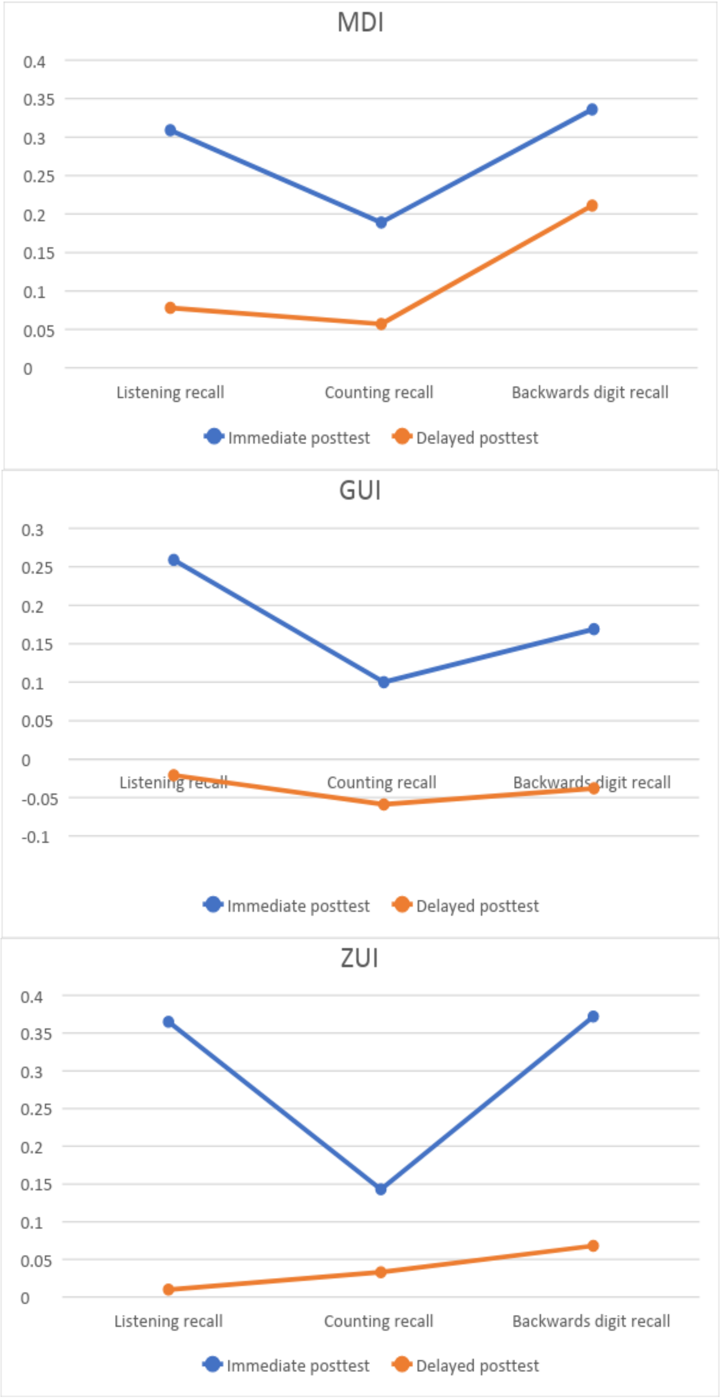


Figure 2.46. Correlations of the three subtests of the verbal working memory with the immediate and delayed CAVA-1 post-tests for all three representations.

Visuospatial working memory tests

The odd-one-out task:

In this task, the participant is shown three shapes in a row. They are then asked to identify the odd-shape-out. Then, the shapes disappear and the participant is required to recall the position of the odd-one-out.

Table 2.56. Mean of the scores for the odd-one-out task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Odd one out	50	82	131	112.6	13

Table 2.57. Correlation between the odd-one-out task score and the immediate and delayed CAVA-1 post-test in its three conditions.

Odd one out		Immediate post-test	Delayed post-test
MDI	N	50	50
	Pearson Correlation	**.303***	.155
	Sig. (2-tailed)	.032	.283
GUI	N	50	50
	Pearson Correlation	**.368****	.121
	Sig. (2-tailed)	.009	.402
ZUI	N	50	50
	Pearson Correlation	**.392****	-.048
	Sig. (2-tailed)	.005	.741

As mentioned above, the odd one out is the only task found to have significant positive Pearson correlations with all three representations in their short-term effect when shown in the immediate post-test. R =0.303, *p = .032* < 0.05 for MDI learners' scores. This positive correlation is statistically significant. R= 0.368, *P = .009* < 0.01 for the GUI learners' scores, and the correlation on odd one out scores, showing a positive and relatively strong correlation, indicating a statistically significant correlation. The highest statistically significant correlation was obtained for ZUI scores of r = 0.392 at *p = .005* (see tables 2.56 and 2.57).

The Mister X task:

The participant is shown a picture of two differently coloured 'Mr. X' figures and asked to judge whether the blue Mr. X is holding the ball in the same hand as the yellow Mr. X. The blue Mr. X is rotated. In addition, the participant needs to remember the location of the ball in Mr. X's hand, in a single sequence.

Table 2.58. Mean of the scores for the Mr. X task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Mister X	50	82	136	104.9	16.7

Table 2.59. Correlation between the Mr. X task score and the immediate and delayed CAVA-1 post-test in all three conditions.

Mister X		Immediate post-test	Delayed post-test
MDI	N	50	50
	Pearson Correlation	.064	.103
	Sig. (2-tailed)	.659	.478
GUI	N	50	50
	Pearson Correlation	-.007	-.044
	Sig. (2-tailed)	.963	.759
ZUI	N	50	50
	Pearson Correlation	.079	.178
	Sig. (2-tailed)	.585	.217

Tables 2.58 and 2.59 above show no significant correlations between the Mr X scores and any of the representations, neither in the immediate or the delayed post-test scores.

The spatial span:

The participant is presented with a picture of two arbitrary shapes, with a red dot present on the shape on the right. The participant has to identify whether the shape on the left is the same as, or opposite to, the shape on the right.

Table 2.60. Mean of the scores for the spatial span task.

Test	N	Minimum	Maximum	Mean	Std. Deviation
Spatial recall	50	79	137	107.29	15.2

Table 2.61. Correlation between the spatial span task score and the immediate and delayed CAVA-1 post-tests under the three conditions.

Spatial recall		Immediate post-test	Delayed post-test
MDI	N	50	50
	Pearson Correlation	.331*	.159
	Sig. (2-tailed)	.019	.270
GUI	N	50	50
	Pearson Correlation	.069	.024
	Sig. (2-tailed)	.633	.870
ZUI	N	50	50
	Pearson Correlation	.253	.146
	Sig. (2-tailed)	.076	.313

Table 2.61 shows MDI learners (R=0.331, $p = .019 < 0.05$) a positive significant correlation. Figure 2.47 shows the correlations of the three subtests of visuospatial working memory, with the participants' vocabulary recall scores in the immediate and delayed CAVA-1 post-tests for all three representations.

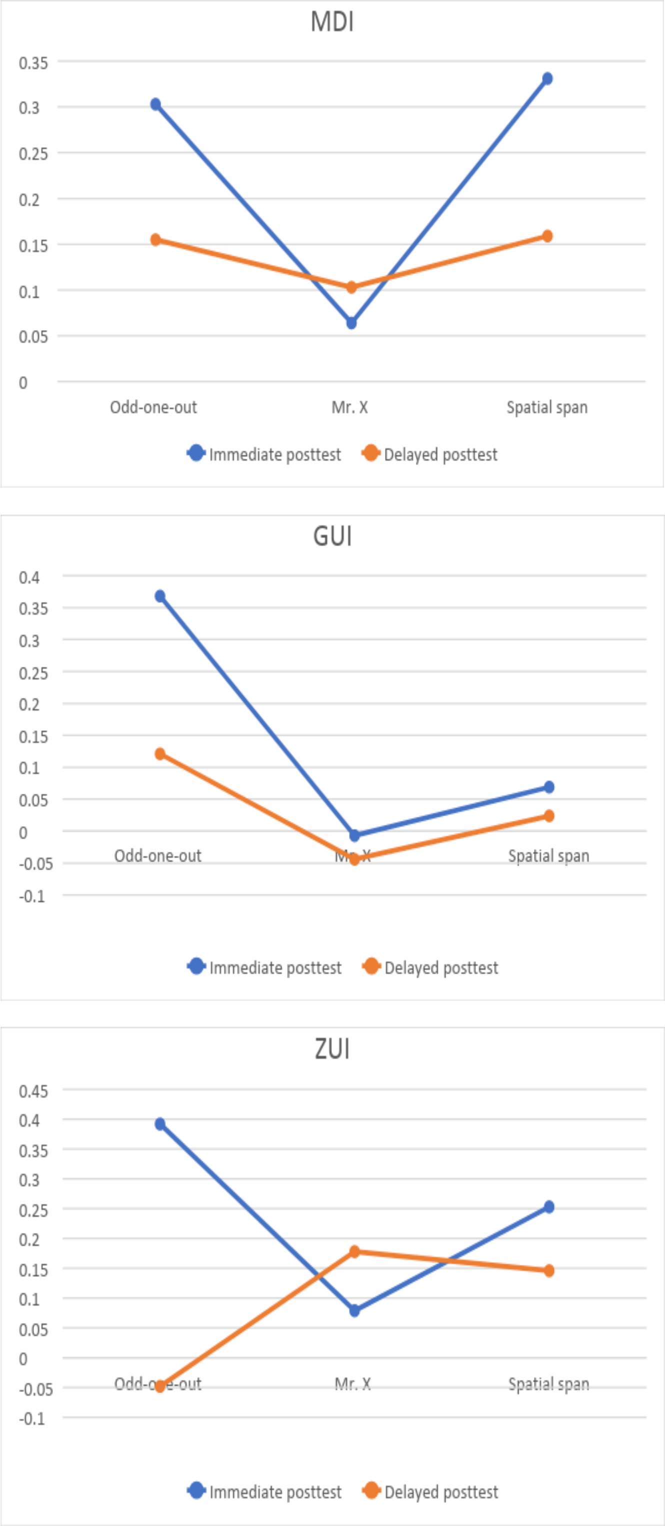


Figure 2.47. Correlations of the three subtests for the visuospatial working memory, with the immediate and delayed CAVA-1 post-tests in its three representations.

2.3.2.3 Summary of findings pertaining to research question 2

This second question posed in the first main study investigated the participants' individual differences in memory when using the AWMA. It sought to summarize the participants' abilities and role in L2 word learning when using three main memory components; namely verbal short-term memory, verbal working memory, visuospatial short-term memory, and visuospatial working memory. These four components were compared to the vocabulary recall scores resulting from the three different types of representation, and Pearson correlation tests applied. In addition, a test of the correlation between each subtest of the four components was conducted. The results revealed that the odd-one-out is correlated with all the representation types, and the immediate ZUI post-test was correlated with the highest number of memory tests, i.e. eight out of the twelve AWMA tests. In both cases the correlation with the ZUI immediate post-test was statistically significant ($p < .05$). In the delayed post-test; however, the correlation between AWMA and all the multimedia representations reduced sharply to a level where no connection seems to be present. Obviously, these results are not definite, and more research is still essential. Before discussing the several possible interpretations of these results, it is important to consider the data in the second study, using the Eyelink Data Viewer. This would involve measuring pupil size, as a proxy for mental effort, and the second study would assist in establishing more practical and theoretical insights into the use of multimedia to teach L2 vocabulary, as well as its impact on vocabulary learning and lexical access.

To sum up, this section presented the first two questions and analysed the collected data. The questions investigated two fundamental areas that when combined offer a comprehensible picture of the process of vocabulary learning in CALL environments. These areas are categorized according to: a) participants' CAVA-1 post-test scores, and b) participants' AWMA scores. In particular, the analysis performed to answer research question one revealed the participants' scores on the ZUI tended towards a more significant effect than the scores on the GUI and MDI in the immediate post-test. For the delayed post-test, the CUI scores were the highest, and there were no statistically significant differences between the three teaching methods in L2 vocabulary learning either. Subsequently, the results from research question two showed that performance in most of the AWMA subtests correlated with ZUI scores in the immediate post-test, and the relationships were statistically significant as a consequence.

2.4 Discussion

Several valuable insights were revealed by the current study, illuminating how L2 learners interact with the three multimedia learning interfaces evaluated during the initial stages of learning novel L2 vocabulary. The study employed three types of L2 word teaching methods in multimedia CALL environments, namely MDI, GUI, and ZUI. The insights gained will be discussed in this section, evaluating each research question in turn.

2.4.1 Discussion about research question 1

> What is the effect of three different interactive multimedia representation types; namely verbal-based (L2-L1: MDI), visual-based (L2-Picture: GUI) and visuospatial-based (L2-Context: ZUI) on second language vocabulary acquisition?

After testing three different interactive representation types in this first study conducted for the current research, the results revealed L2 vocabulary learners typically achieve higher scores in ZUI than GUI and MDI in an immediate post-test. The pattern of results differed in the delayed post-test: GUI was found to be the highest, and was then followed by MDI and ZUI respectively. However, the difference between the conditions was not statistically significant over either the short- or long-term. In the following subsections, the results obtained from the immediate and delayed post-tests will be discussed prior to presenting a general discussion of the first research question.

CAVA-1 immediate post-test

The analysis of the immediate post-test suggested that including visuospatial design elements in multimedia learning might result in the employment of the visuospatial sketchpad for working memory (Baddeley, 2012), which would then contribute positively to the learners' retrieval of L2 words in the short-term. This idea was advocated as a result of the learners achieving higher scores for ZUI compared than for GUI and MDI. Given that the CAVA-1 post-tests were designed to assess the efficiency of L2 word learning, it can be inferred that ZUI was largely superior to GUI and MDI, in the immediate post-test. The multi-channel combination of presentations (verbal and visuospatial) in ZUI, intended to communicate L2 words resulted in enhanced learner performance in the immediate post-test. The combination of multiple channels of presentations proved effective, which is in line with assertions in multimedia learning effectiveness literature; namely, the Cognitive Theory of Multimedia

Learning (Mayer, 2005) and Dual-Code Theory (Paivio, 1991). It is suggested that the influence of ZUI and GUI proved greater, as it delivered the meaning for L2 words in two modes (verbal and visual), as proven in various studies (e.g. Jones & Plass, 2002; Jones, 2009) designed to investigate multimedia instruction in CALL environments. ZUI and GUI displayed visual information about concrete objects that are functionally similar to what is seen through the mind's eye. This seemed to facilitate the process of mental imagery, which is thought to be strongly analogous to visual perception, although not identical to it (Farah, 1988; Finke, 1989; Jolicoeur & Kosslyn, 1985). As a result, ZUI and GUI support better encoding and retrieval of target L2 words in the immediate post-test than MDI. Moreover, the distinctive features of visual materials used in ZUI and GUI contributed additional crucial aspects of word knowledge in the target language, and are considered a fundamental quality of language learning (Nation, 2013). The CAVA-1 post-test used in this study was designed to test learners' ability to acquire different levels of word meaning, including word associations, to enhance L2 form-meaning mapping. This finding reflects the positive multimedia environment, with interactive annotations that are designed to deliver essential information about word knowledge, to teach the target language.

The analysis of the CAVA-1 immediate post-test also revealed another insight of importance for the design of future vocabulary CALL software. MDI was found to be less effective as a mode of facilitating learning, and of recalling a target L2 word to answer the CAVA-1 post-test questions. The low vocabulary recall scores that learners using MDI achieved might imply that using only written words as a stimuli limits this teaching method. This finding echoes previous research, which has highlighted that less information can be delivered through the written form of the word, as this method principally relies on the phonological loop as a verbal channel. Teaching L2 meaning through an L1 translation equivalent seemed inadequate to communicate other aspects of the word knowledge when compared with ZUI and GUI, such as ZUI's visuospatial word associations. Providing extra information about the concrete objects taught in the current study via ZUI and GUI's was found to be more effective. The absence of such additional information in MDI might weaken the encoding process, and thereby reduce the retrieval of L2 words learnt through MDI in the immediate post-test.

CAVA-1 delayed post-test

In the 1-month delayed post-test, the use of different visual representations of CUI, as a multimedia learning method in CAVA-1 indicated efficient long-term vocabulary learning. ZUI scores fell dramatically, and the multimedia representation was allocated last place among the three e-learning methods in the delayed post-test. The difficulties that faced learners in the long-term retrieval of ZUI words was shown by the lowest CAVA-1 delayed post-test scores. Conversely, GUI proved to be the most effective teaching method over the long-term, followed by MDI. The CAVA-1 delayed post-test analyses thus highlighted the long-term effects of using an illustrated picture in a GUI design. This was inferred from the CAVA-1 delayed post-test results that revealed higher learning efficiency for GUI over MDI and ZUI. This finding demonstrated that, over the long term, second language vocabulary learners might be better able to learn L2 concrete words via pictures compared with visuospatial representations or L1 translation equivalents. The dual coding (Paivio, 1991) benefit of GUI over MDI when learning target L2 words lasted longer as it affected its short-term basis.

General discussion about research question 1

The effect when using various multimedia representations to communicate L2 words was present for both short- and long-term vocabulary learning. Nevertheless, the differences between the three teaching methods were not significant. This could be as a result of two main reasons. First, the number of the target L2 words in CAVA-1 is 24 and this might need to be increased in order to reach larger differences in the learners' achievements. Second, CAVA-1 seemed to provide rich contents in all its three teaching methods with three learning phases including practice exercises. This way of presenting the target words in CAVA-1 might decrease the differences between the impact of MDI, GUI and ZUI for vocabulary learning. However, the spatial and visual design elements of ZUI and GUI potentially explain the priority of both multimedia representation types when compared to the verbal nature of MDI for short- and long-term word learning. This priority was generally associated with including more visuospatial information about target concrete objects, which allowed the inclusion of more aspects associated with target word knowledge as delivered through the two modes (verbal and visuospatial). These kinds of information functioned as cumulative information about the target word, which employed multiple working memory components. However, some of this information, namely ZUI spatial based stimuli was quickly forgotten.

It seems that the design elements of ZUI played an essential role in attracting the attentions of learners, as the visual element appealed. The ability to see the whole overview of a shot, including all the target words and their relationship as a lexical set visuospatially seemed to facilitate the learning process efficiently for the learner. Additional information, such as spatial (positional) relationship (e.g. above, behind, below, beside) is concretely represented in ZUI. This kind of representation apparently employs Baddeley's (1986) Central Executive (CE), which is an attentional controlling system, associated with the two CE slave systems; the phonological loop and the visuospatial sketchpad, which provide short-term storage and permit manipulation of information. Learners might find it easier to utilize ZUI when its attributes reflect what the learner might observe in real-life situations. ZUI learners were able to repeatedly grasp the majority of the aspects of the view presented (e.g. the living room) individually for each object by zooming in for a closer look, or as a more holistic overview when zooming out to see the big picture, and this might deepen the process of developing word knowledge.

The L2 vocabulary test used in this study was designed to test the learners' ability to learn different aspects of the word's meaning (translation equivalent, illustrated picture, and visuospatial representation) and to explore their impact on L2 word form-meaning mapping. As the ZUI learners' performance as indicated by the scores in the CAVA-1 immediate post-test proved to be the best, its visuospatial component should be included as part of the construct of L2 multimedia learning. The distinctive features of the visuospatial materials used in ZUI reflected crucial aspects of word knowledge in the target language, and could be considered a fundamental part of vocabulary CALL. On the other hand, GUI and MDI, which are commonly used when teaching L2 words to beginners, by showing pictures or translation equivalents, limited the use of spatial design elements that can be effectively employed for learning and processing the word. However, GUI has an advantage over MDI, as it employs visual materials along with verbal ones. Although the difference is not significant, it is likely that this dual coding of information has an impact on both short- and long-term word learning. Such a conclusion coincides with dual coding theory. In support of Paivio dual coding theory (1991), the results from the first question in the current research correspond with the idea that each system can receive its own load independently from the other system, and how the stimulus we get from one system effects the other. The theory suggests it is possible to increase our cumulative capacity by splitting information across cognitive systems, since we can also employ them simultaneously to our benefit, as a way to provide extra information without overloading any one system.

Another possible explanation of the current findings derives from mental imagery theory. The use of concrete nouns minimized extraneous differences between the three Cava conditions (MDI, GUI, and ZUI), and avoided potential confounders from abstract words. In fact, concrete nouns are probably more readily represented in pictures, and the mental images associated with items that are represented in the mind deliver similar copies of external physical images (Kosslyn, 2006). Therefore, mental imagery might have an impact on the retrieval of both ZUI and GUI words.

Mental imagery theory is supported by the functional equivalence hypothesis, which represents the knowledge of concrete objects. It proposes that mental images that are associated with concrete objects, although they form inside the brain, are functionally similar to what we perceive through our sense organs. Thus, mental imagery and visual perception are strongly analogous, although not identical (Farah, 1988; Finke, 1989; Jolicoeur & Kosslyn, 1985; Rumelhart & Norman, 1988). A further important principle of visual imagery, as it relates to the current discussion, was suggested by Finke (1989), and could explain ZUI high vocabulary recall scores in the CAVA-1 immediate post-test. It relates to the representation of spatial information, as a mental visual image. The principle states that the mental spatial organization of elements resembles the arrangements in the actual object (Finke, 1989). Therefore, mental imagery might be effective for both retrieving memories about the physical characteristics of ZUI and GUI words that participants learned, and for recreating spatial relationship among ZUI objects. Another possible explanation supporting the high short-term impact of ZUI, is that the functional equivalence hypothesis also underlies the idea of zooming in on our mental images, just as we do with our representations of precepts. Capacity for mental zooming might be facilitated be ZUI interface transitions, leading to deeper encoding of L2 target words when learned by ZUI.

Moreover, the present findings coincide with Mayer's (2005) cognitive theory regarding two principles of multimedia learning; namely the multimedia principle, and the spatial contiguity principle. The multimedia principal supports the use of GUI. It says that words and pictures are better than pictures or words on their own, and we analyse verbal and visual communication differently. The spatial contiguity principle was reflected in the ZUI advantage over GUI when enhancing the effectiveness of the learning material, although only on a short-term basis.

It is likely that, in the early stages of second language vocabulary learning, participants proved to be better at processing pictures and texts, as a means to acquiring word meaning. Moreover, from GUI and MDI learners' results in particular, it can be argued that both immediate and delayed post-tests results reveal the importance of dual coding learning. The present findings and analyses supporting this pattern further demonstrate the significance that dual coding has on long-term learning. That is, the fact that the learners in the present study achieved better results in GUI than MDI not only shows that effective vocabulary CALL intervention involves pictorial annotations, but also, and more importantly, that visual materials seemed to take centre stage in learner cognition.

Such a conclusion regarding effective dual coding coincides with previous research focused on vocabulary CALL, in both incidental and intentional programs. These studies have demonstrated that communicating L2 word meaning through both verbal and visual channels effect learners' acquisition of new second language vocabulary in multimedia environments. Akbulut (2007), for instance, investigated the effect of presenting verbal annotations (written definitions) and multimedia annotations (written definition with illustrated picture or video) as a way to teach English vocabulary to Turkish students incidentally. The findings revealed that the learners who had access to the multimedia glosses (definitions along with picture or video) obtained significantly better learning outcomes on both immediate and delayed vocabulary post-tests than verbal only learners. These results are similar to Al-Seghayer's (2001) study, as they found that on both recognition and production vocabulary tests, adding a video or a picture to a definition is more effective than producing a definition in isolation. Corresponding to these findings, Kost et al. (1999) found that presenting pictorial and verbal annotations to deliver L2 English words for German students led to better vocabulary acquisition in both immediate and delayed post-tests than showing just one. Furthermore, the video group significantly outperformed other groups. Similar results were also obtained by Lin and Tseng (2012) when teaching difficult English words to a total of 88 Taiwanese junior high school. Yoshii and Flaitz (2002) found that the combination of the two annotations, i.e. picture and text annotations group, performed better than the other two groups in immediate as well as delayed post-tests, when conducting the vocabulary test for ESL students at beginner and intermediate levels. Yeh and Wang (2003) examined the impact of three conditions of EFL vocabulary teaching methods on vocabulary acquisition for college learners in Taiwan. The researchers compared the presentation of text annotation glosses only, focusing on texts with pictures, and texts with pictures and sounds. Text-plus-picture annotations were found to be significantly better than no-picture (text- only) glosses in Yoshii's (2006) study.

The support of my findings did not only arise from research into incidental vocabulary learning, but also the combination of both verbal and pictorial annotation, which proved to be more effective in intentional vocabulary learning in multimedia CALL environments. For example, Dubois and Vial (2000) found that French learners of Russian vocabulary, who had learned using verbal and pictorial annotations (translation plus illustrating image), outperformed those who had learned with only verbal annotations (translation); i.e. word memorization. Fagehi (2013) had 53 non-native English participants who studied 21 terms via three modes; text only, text and video, and text and video plus question (7 items for each mode). In a post comprehension test, the findings showed that teaching via Text plus Video and Text plus Video plus Question were both significantly more effective than teaching using Text-only. In a Web-based intentional vocabulary learning program, Kim and Gilman (2008) exposed 172 middle school Korean students to one of three multimedia conditions, designed to learn English vocabulary; visual text, spoken text, and graphics. Again the results revealed superiority for conditions with graphics added to the text over text only methods. Finally, Rimrott (2010) found that only in the immediate, but not in the delayed post-tests, illustrated clusters provided with pictures were significantly more effective than clusters without pictures, for both concrete and abstract L2 words.

The current results for the first question of this research were found to be consistent with the above-mentioned studies: texts with picture annotations, as used in GUI, were effective for vocabulary learning as reflected in immediate and delayed post-tests.

2.4.2 Discussion about research question 2

> What is the effect of individual differences on verbal and visuospatial working memory, and of short-term memory capacity on the initial uptake of a new vocabulary?

The second research question posed as part of the first study explores whether individual differences in cognitive abilities, specifically working memory and short-term memory spans, influence the process of foreign language vocabulary acquisition. The researcher applied the Pearson correlation coefficient to uncover any correlations between learners' scores in CAVA-1 immediate and delayed post-tests via all three conditions (MDI, GUI, ZUI), and participants' cognitive abilities as achieved in the AWMA on the other. As stated in the results section, the AWMA revealed high level differences among the 50 participants in relation to their four major components; verbal short-term memory; visuospatial short-term memory; verbal working memory; and visuospatial working memory. The results of the Pearson coefficient showed significant correlations between some of the AWMA components and the subtests under the CAVA-1 conditions. In the following subsections, the results of the AWMA correlations will be discussed alongside the CAVA-1 immediate and delayed post-tests, before a general discussion of the second research question is presented.

AWMA correlations with the CAVA-1 immediate post-test

The CAVA-1 immediate post-test analyses showed that using different multimedia representations indicated a relationship with learners' cognitive abilities. As stated in detail in the results section, several AWMA results showed significant correlations with the immediate CAVA-1 post-test in its three multimedia representation methods. The participants' performance using some of the CAVA representation methods were found to correlate with AWMA more than others. Interestingly, the odd one out test, which is a visuospatial working memory task, was correlated with all representation types. In addition to the odd one out, another seven AWMA tasks were correlated significantly with the participants' sores in ZUI; i.e. the teaching method that had the highest number of correlations with AWMA in CAVA-1 immediate post-test scores. These seven tasks were digit recall, listening recall, backward digital recall, and dot matrix. The second number of AWMA correlation with CAVA-1 immediate post-test was with the learners' scores in MDI, which were found to correlate significantly with listening recall, counting recall, and spatial recall. GUI came last when the participants' vocabulary recall was correlated only with the odd one out in the immediate post-test. Given that the Pearson correlation is used in this study to reflect the strength of the correlations, the learners' individual differences across the memory components can be inferred to have had a strong impact on L2 vocabulary acquisition. These results further support the idea of Baddeley's (2012) recent model of the multiple structure of working memory. Correspondingly, the current findings parallel previous research that proved the important role of individual differences in memory in language learning.

It can also be demonstrated that ZUI design elements interacted with more AWMA components on a short-term basis than GUI and then MDI. It is likely that the additional visuospatial stimuli used by ZUI served as extra cues providing further information to be utilized to aid subsequent short term recognition. The L2 word, as a target memory, is supposed to link to cues in associative connections, which form a specific memory trace. Such connected information can serve as retrieval cues that spread the activation to the target memory, which is to be recovered. As more information was available in ZUI than GUI and MDI, there was potential to create more associative connections with the target L2 word.

AWMA correlations with CAVA-1 delayed post-test

In the CAVA-1 delayed post-test, only GUI appeared to have a significant correlation with one of the visuospatial short-term memory tasks; the mazes memory. The other most obvious finding to emerge from the analysis was that all the other AWMA subtests' correlations declined with all of CAVA multimedia representations. That includes ZUI, which had several significant correlations with most of the AWMA subtests in the immediate post-test, and was not found to make any correlations in the delayed post-test with AWMA. The long-term effect of using the illustrated picture in a GUI design can be inferred form the high correlation with the mazes memory task, which assesses visuospatial ability. This can be supported by the CAVA-1 delayed post-test results, which revealed higher learning efficiency for GUI over MDI and ZUI. This finding demonstrated that second language learners with a high ability in visual memory will learn the L2 concrete words via pictures better when compared with the translation equivalent condition only. Correspondingly, the dramatic changes in ZUI's correlations with AWMA subtests over time supported the results of the first question of the current thesis, revealing a rapid decay of ZUI representation elements despite the short-term advantage of this. This analyses thus highlighted memory decay and a forgetting curve, Ebbinghaus (1913), but in different ways and reasons for each of the CAVA multimedia representations. Long-term retrieval of ZUI words seemed to have been adversely affected by the swift decay of spatial memory (Fagan et al., 2013), whereas the dual coding (Paivio, 1991) benefit of GUI over ZUI and MDI influenced the long-term retention of the target L2 word.

AWMA correlations with the CAVA-1 immediate post-test

The CAVA-1 immediate post-test analyses showed that different multimedia representations indicated a relationship with learners' cognitive abilities. As stated in detail in the results section, several AWMA results showed significant correlations with the immediate CAVA-1 post-test in its three multimedia representation methods. The participants' performance using some of the CAVA representation methods were found to correlate with AWMA more than others. Interestingly, the odd one out test, which is a visuospatial working memory task, was correlated with all representation types. In addition to the odd one out, another seven AWMA tasks were correlated significantly with the participants' sores in ZUI; i.e. the teaching method that had the highest number of correlations with AWMA in CAVA-1 immediate post-test scores. These seven tasks were digit recall, listening recall, backward digital recall, and dot matrix. The second number of AWMA correlation with CAVA-1 immediate post-test was with the learners' scores in MDI, which were found to correlate significantly with listening recall, counting recall, and spatial recall. GUI came last when the participants' vocabulary recall was correlated only with the odd one out in the immediate post-test. Given that the Pearson correlation is used in this study to reflect the strength of the correlations, the learners' individual differences across the memory components can be inferred to have had a strong impact on L2 vocabulary acquisition. These results further support the idea of Baddeley's (2012) recent model of the multiple structure of working memory. Correspondingly, the current findings parallel previous research that proved the important role of individual differences in memory in language learning.

It can also be demonstrated that ZUI design elements interacted with more AWMA components on a short-term basis than GUI and then MDI. It is likely that the additional visuospatial stimuli used by ZUI served as extra cues providing further information to be utilized to aid subsequent short term recognition. The L2 word, as a target memory, is supposed to link to cues in associative connections, which form a specific memory trace. Such connected information can serve as retrieval cues that spread the activation to the target memory, which is to be recovered. As more information was available in ZUI than GUI and MDI, there was potential to create more associative connections with the target L2 word.

AWMA correlations with CAVA-1 delayed post-test

In the CAVA-1 delayed post-test, only GUI appeared to have a significant correlation with one of the visuospatial short-term memory tasks; the mazes memory. The other most obvious finding to emerge from the analysis was that all the other AWMA subtests' correlations declined with all of CAVA multimedia representations. That includes ZUI, which had several significant correlations with the majority of the AWMA subtests in the immediate post-test, and was not found to make any correlations in the delayed post-test with AWMA. The long-term effect of using the illustrated picture in a GUI design can be inferred form the high correlation with the mazes memory task, which assesses visuospatial ability. This can be supported by the CAVA-1 delayed post-test results, which revealed higher learning efficiency for GUI over MDI and ZUI. This finding demonstrated that second language learners with a high ability in visual memory will learn the L2 concrete words via pictures better when compared with the translation equivalent condition only. Correspondingly, the dramatic changes in ZUI's correlations with AWMA subtests over time supported the results of the first question of the current thesis, revealing a rapid decay of ZUI representation elements despite the short-term advantage of this. This analysis thus highlighted memory decay and a forgetting curve, Ebbinghaus (1913), but in different ways and reasons for each of the CAVA multimedia representations. Long-term retrieval of ZUI words seemed to have been adversely affected by the swift decay of spatial memory (Fagan et al., 2013), whereas the dual coding (Paivio, 1991) benefit of GUI over ZUI and MDI influenced the long-term retention of the target L2 word.

General discussion about research question 2

The relationship between the individual differences in capacity for different aspects of memory and the multimedia representations used in CAVA-1 were described. Students with a better working memory achieved higher scores when learning via ZUI representation. ZUI design elements interacted with almost all the AWMA components in the short-term; however, its elements were found to be associated with a short period of impact, as memory for the ZUI components soon decayed. Using fewer visual design elements than ZUI's, such as providing only an illustrated picture as in GUI design, was found to have a longer lasting effect, as performance associated with GUI learning correlated with mazes memory long-term. This style of teaching L2 concrete words from pictures was more effective for high visual memory learners, in both the short and long term; whereas learners with higher verbal short-term memory abilities scored more in MDI.

The current results suggest the multicomponent model of working memory offers a better explanation detailing learners' performance on immediate and delayed post-tests than other theories, which assume a unitary-store model of memory (e.g. Nairne, 1990; Cowan, 1999). M-WM is supported by evidence accumulated from diverse sources, which reveal that distinct cognitive functions are performed by different systems within the brain in a highly independent way (Baddeley, 2012). However, these components apparently have remote links to one another (Baddeley & Hitch, 1974).

The current findings support Logie's (1995) view that there are two separate subcomponents of visuospatial working memory: a passive visual store (visual cache) for processing visual information, and an active rehearsal process (inner scribe), for rehearsing and retrieving spatial information. Apparently, the visual cache stored details of the visual form and colour of a GUI and ZUI stimuli, whereas the inner scribe processes ZUI spatial and movement related information. The current findings correspond to the assumption that verbal and visual short-term memory systems are discrete, as supported by evidence from neuropsychological studies (e.g. Della Sala & Logie, 2002; Shallice & Warrington, 1970), and behavioural data collected from unimpaired participants (Meiser & Klauer, 1999; Della Sala et al., 1999).

AWMA findings suggest the visuospatial sketchpad was involved in the short-term storage and processing of GUI and ZUI materials. The visuospatial sketchpad is a slave system in the working memory, however, the evidence suggested it is responsible for the encoding and manipulation of temporary GUI and ZUI representations. As the AWMA results revealed, the visuospatial sketchpad plays an important role in learning processes that involve pictorial and visuospatial representations in CAVA, as occur under GUI and ZUI conditions. Moreover, although connected within the sketchpad, the spatial and visual aspects of working memory appeared to be detachable, as the current findings show.

Processing spatial information, which is attributed in memory via cognitive mapping processes (Fagan et al., 2013), seemed to facilitate the creation of mental representations through a series of acquiring, encoding, and then recalling the information provided via ZUI. This method of media representation takes advantage of spatial memory as an important aspect of the memory system. It has not yet been investigated for its role in language learning to the extent that verbal or visual abilities have been. However, previous research found that spatial memory is characterized by its short-term effect (Hole, 1996; Fagan et al., 2013) and can decay faster than any other types of information. Such a conclusion from earlier studies, regarding the fast decay of spatial memory, coincides with current study results in terms of the short-term effect of ZUI in the immediate post-test of CAVA-1, as well as with the AWMA findings, which revealed strong correlations with ZUI over the short-term but not the long-term basis. Support for the fast decay of spatial memory can be inferred as correlations disappeared entirely a month after ZUI-based learning. Memory decay refers to the idea that memory becomes weaker and fades as time progresses. This phenomenon explains the loss of information from verbal and visual working memory (Broadbent, 1958; Baddeley, 1986; Cowan, 1988; Towse, Hitch, & Hutton, 2000). Decay happens when neurons die, tissues change, or associations deteriorate. Mental representations of spatial information are susceptible to decay, especially in the presence of extraneous stimuli (Hole, 1996). Spatial memories, unless they maintained or reinforced, will decay faster than verbal or visual memories (Fagan et al., 2013). Spatial information is attributed in memory via cognitive mapping processes to form mental representations through psychological series of acquiring, encoding, storing, recalling and decoding (Fagan et al., 2013). What might make the current results of this study agree with those observed in memory decay literature is that the types of memory that were probably formed by MDI and GUI contrasted with the ZUI design, and were spatially independent when encoding L2 word knowledge.

In light of the fact that we have access to limited information stored in our memory, additional information in CAVA representations could help provide extra cues utilized to aid subsequent recognition. The ability to succeed in accessing the target memory, i.e. of the L2 word, depends on the retrieval process as a cognitive control of memory, which is contributed to by the left ventro-lateral prefrontal cortex (VLPFC), as found in Badre and Wanger's (2007) fMRI study. The target memory is connected with cues and associative connections, forming a specific memory trace. All the connected information can serve as retrieval cues, which spread the activation to the target memory for recovery. The spreading activation begins from the cues to the target memory and involves associations. When the activation level of the target memory is increased, it then becomes accessible. Seeing or even thinking about associative cues raise the activation level of the memory trace and then the activation is spread, via the associations, to reach the target. The amount of activation will be greater when the associative connections are stronger.

Sometimes we do not succeed in retrieving memories because there are factors determining retrieval success. One of the important factors related to the current study, is the cue-target associative strength. The strength of association between the cue and the target helps to increase the amount of activation spread. If these associations are weak, retrieval can fail.

Another important factor in retrieval is the number of cues. When more relevant associative cues have been added and activated during the retrieval process, it is likely to be enhanced. The combination of the cues together might elicit a target memory. The activation of more cues will be spread to the target utilizing different resources, which causes it to activate more rapidly so that it can be recovered more easily. Having two cues was suggested by research on dual coding for its usefulness in the subsequent retrieval process. Rubin and Wallace (1989) found that using semantic and rhyme cues together is significantly better than using only one when retrieving learned words. Moreover, retrieval is likely to be affected by context. Context cues are the retrieval cues that indicate features associated with the circumstances in which the target item has been encoded, such as (for instance) the time and location.

When someone decides to do something, and goes to do it, and then forgets what they are doing, they typically return to where they were and suddenly remember what they want. This is because, when they go back to the first place, they can see the environmental context cues that were encoded along with the target memory, and those cues help with retrieval. These context-dependent memory cues are encoded with spatiotemporal, physiological, mood, or cognitive context cues, and make it is faster and easier to recall the target memory when these cues are present. Previous research has found that spatial memory is characterized by a short-term effect, and can decay faster than any other type of information. This finding is consistent with the current study's results in terms of the short-term effects of ZUI in the immediate post-test of CAVA-1.

What are the mechanisms that underlie forgetting? Since retrieval breakdown is a type of forgetting, it is essential to understand the mechanisms that underlie forgetting. There is no doubt, for many people, that when time goes by, forgetting happens and increases. Hermann Ebbinghaus (1913) conducted a study to investigate the forgetting of nonsense syllables. He tested himself on 169 lists of thirteen syllables. The forgetting curve that he presented (see figure 2.48) shows that forgetting happens rapidly at the beginning, slowing down gradually as time progresses. This result was later repeated under different learning conditions.

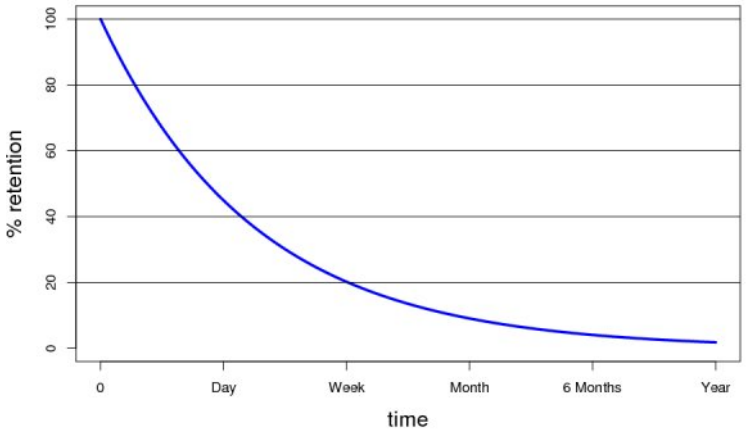


Figure 2.48. The forgetting curve depicting the relationship between memory and time. (Source: Ebbinghaus, 1913)

Forgetting can be a measure of effective learning, as it takes longer when something is appropriately learned. The newly learned information will be vulnerable until it is consolidated. Consolidation is a time-dependent process that effectively cements the new memory traces making them more permanent. There are two types of consolidation: synaptic consolidation, which takes from hours to days (Dudai, 2004) and requires biological changes in the connections between neurons, and system consolidation, which takes years to complete and occurs when memory retrieval becomes hippocampus independent reliant on the cortex (Squire, 1992; Dudai, 2004).

The GUI and MDI correlations with AWMA also agree with those in previous studies, which highlighted the contribution of cognitive abilities to multimedia learning. There is no study, as far as the researcher knows, that has empirically examined how working memory affects intentional multimedia vocabulary learning. Furthermore, few studies have discussed individual differences in cognitive abilities in the context of other types of multimedia learning, such as incidental vocabulary learning. Several studies (e.g. Acha, 2009; Plass et al., 1998, 2003; Jones, 2009) have correlated between incidental vocabulary multimedia learning and memory abilities, whereas no intentional vocabulary experiment has yet been found to do so. My findings seem to be consistent with Plass et al.’s (1998) findings, which revealed that when learning L2 English words, verbalizers tend to benefit from verbal annotations, as indicated by vocabulary acquisition post-tests, whereas visualizers tend to benefit from visual annotations. However, the method conducted in Plass et al.’s study to classify the participants into verbalizers or visualizers did not use a cognitive measuring task, although in fact it was by watching the tendencies of the participants in choosing verbal or visual options.

There were strong correlations between highly visuospatial and verbal students with CAVA-1 results that emerged in the current study, and these are in line with those in previous studies. For example, in a more recent study by Plass, Chun, Mayer and Leutner (2003), English-speaking college students' cognitive abilities (verbal and spatial) were measured prior to them learning German vocabulary incidentally while reading stories in a multimedia environment. The recall of word translation was worse with low spatial and low verbal students when they were exposed to both verbal and visual annotations than was the case in other conditions, where there is only one kind of information; verbal or visual annotations. The researchers used the Card Rotation Test to measure spatial ability, which, they argued, is closer than visual ability in terms of integrating the used pictures mentally with the provided verbal information to create a mental representation of the target word. Verbal ability, on the other hand, was measured using the Vocabulary Test from The Nelson–Denny Reading Test (NDRT). Another example that is in agreement with the current findings is a study that used verbal and spatial ability tests from the Evaluación Factorial de Aptitudes Intellectuals. Acha (2006) applied a study with 135 Spanish children learning English words in a multimedia program, and revealed results that were consistent with difficulties in adding pictures designed for learners with low cognitive abilities.

On the other hand, the correlations of individual differences in visuospatial abilities with CAVA-1 representations, as revealed by the present study, contradict the results of Jones's (2009) study, finding little difference between high- and low-spatial-ability learners when interacting with verbal or visual stimuli. Jones used the span tasks taken from the Kit of Factor-Referenced Cognitive Tests (Harman, Ekstrom, & French, 1976) to measure the spatial and verbal abilities of 171 English students, and then classified them as low verbal, high verbal, low spatial, and high spatial learners. The participants were exposed to a multimedia program to learn L2 French words while listening. The researcher also found that high-verbal-ability learners outperformed low-verbal-ability learners, in terms of their learning outcomes, when only pictorial annotations of target words were presented.

Overall, the findings of the second research question established different kinds of learning, resulting from the use of various vocabulary teaching methods. The use of ZUI, to teach other aspects of word knowledge, created an environment that facilitated an appropriate depth of knowledge. Schmitt (2014) defined depth of word knowledge as it related to familiarity with individual lexical aspects, i.e. associations and collocations. A number of researchers identified the contextualized knowledge of the word as more problematic to explain, as it has a tendency to be acquired implicitly. However, ZUI's interface explicitly provided the participants with this knowledge in terms of its visuospatial representations. Consequently, a higher number of aspects of word knowledge could be delivered via ZUI (i.e. visuospatial associated terms of the target word), and resulted in an increased depth of processing of the word meaning. This created a longer lasting memory trace concerning the word in question, thereby facilitating the retention of a greater number of words in the delayed post-test than delivered via GUI or MDI. The zoomable interface then tended to be more effective for the long-term storage and retrieval of lexis. Although ZUI was found to exert a relatively beneficial effect over the short term, the shallow processing appeared insufficient to enable the participants to consolidate form–meaning links. Thus, the majority of these connections were broken, resulting in the participants being unable to retrieve the words during the delayed post-test. In contrast, the dual coding effect of GUI lasted longer. When it came to using MDI, this helped the participants to acquire the meanings of items from a word list using a translation equivalent. The results showed that this mode of presenting the words (i.e. isolated from any kind of contextual elaboration) tended to limit students from forming small degrees of association between meaning and form.

2.4.3 Summary

The first study in the current thesis investigated how different types of multimedia interfaces interact with learners' verbal and visuospatial abilities, affecting their vocabulary learning, causing various results. In addition to the agreement present in most of the intentional vocabulary CALL studies mentioned above, the present study's findings echo the majority of previous studies that have investigated and discussed the impact of individual differences in working memory, specifically Baddeley's M-WM model, on multimedia learning in general. Former studies that concur with the current study, in terms of confirming the association between multimedia representations and Baddeley's theory of working memory, took two key model assumptions into consideration, when discussing their findings. The two assumptions included the separation of verbal and visuospatial working memory subcomponents, and the limited capacity for each of these subcomponents to process information in parallel with another subsystem.

In Gyselinck et al.'s (2000) study, learners with high spatial abilities outperformed low spatial ability learners when presented with pictures in a follow up study (2002). Again, the high spatial ability participants relied more on visuospatial sketchpads when studying using pictures.

Gyselinck et al. (2002) found that the spatial tapping task did not interfere with text processing, indicating that the visuospatial sketchpad is not involved in text manipulation. However, the results revealed the involvement of the phonological loop in text processing, although not in picture processing. Another study (Pazzaglia et al., 2008) further supported the idea of the contribution of working memory components in the context of multimedia learning.

The impact of individual differences in working memory on multimedia learning, which also emerged in the present study, correspond to more recent studies (e.g. Austin, 2009; Doolittle et al., 2009; Doolittle & Mariano, 2008; Sanchez & Wiley, 2006). The results of the current study further confirmed the findings of previous research, which revealed associations between the phonological loop and the external verbal stimuli. In MDI information processing, the phonological loop and central executive measures were found to have contributed to the L2 learning process. MDI is a method of presenting lists of bilingual vocabulary, which shows the target L2 word with its L1 translation equivalent. It is apparent that this teaching method is verbally based, and that the information presented in this way is probably going to be manipulated in the phonological loop; i.e. the verbal channel of the working memory. The statistically significant correlations of MDI with the listening recall task, correspond with the findings in a growing body of literature. The listening recall and counting recall tasks are both verbal tasks, and my findings echo previous studies, indicating that highly verbal learners achieve more when learning through representations that are only verbal. Few experiments have examined the effect of the phonological loop and the visuospatial sketchpad span using only simple span tests (e.g. Gyselinck et al. 2000; Pazzaglia et al., 2008). Moreover, in addition to using simple span tasks, which arguably provide less clear-cut evidence than complex ones (Schüler, Scheiter, & van Genuchten, 2011), these two studies revealed different results related to the involvement of PL and VSSP capacity in multimedia learning, as can be explained by the dependent variables in each study. Information processing in ZUI and GUI, on the other hand, relied more on the visuospatial sketchpad, as well as the central executive. Thus, their span measures were found to be associated with the retrieval process of target L2 words. The correlations of the GUI representations with the AWMA were found to be logical and consistent with what has been reported in most previous research. GUI is a media representation that delivers target L2 words by showing pictures that illustrate the meaning of the words. Use of pictures allows the visual memory to be stimulated and employed. As this representation is visually based, it makes sense that learners' performance on the recall CAVA-1 post-test is significantly correlated with the odd one out task. Odd one out is an AWMA task measuring the visual working memory. In the multimedia research literature, the use of pictures was shown to have a greater effect with high visual learners.

It has also been reported in previous research that spatial memory is characterized by its short-term effect, as it decays faster than other types of information. These findings are consistent with the current study results in terms of the short-term effect of ZUI in the immediate post-test of CAVA-1, and also with the AWMA findings. These revealed a strong correlation with ZUI on a short-term but not a long-term basis. That supports the rapid decay of spatial memory, since the AWMA correlations declined sharply a month after exposure to ZUI. Although there are some empirical studies that have used reliable assessment tasks to measure the phonological loop, visuospatial sketchpad, and central executive correlations with multimedia learning, to date, there is no study that has yet investigated the contribution of the episodic buffer in multimedia learning (Schüler et al., 2011). This should be taken into consideration here, and will be discussed in the conclusion chapter. Overall, the present study gives resounding support for the diverse results in the literature designed to research the involvement of working memory components in multimedia learning.

CHAPTER THREE: SECOND STUDY

This chapter presents the second study, which replicates the first. CAVA with its three multimedia representations will again be used to communicate novel L2 Arabic words via its three interfaces; MDI, GUI, and ZUI. However, the target words in the second version of CAVA (CAVA-2) were increased from 24 in CAVA-1 to 108 words, while the practice exercises in the software were reduced by focusing more on introducing more target words than on the number of practice exercises. Moreover, instead of using CAVA post-tests, three measures will be employed in the second study when employing a translation recognition task, focusing on response time, accuracy, and pupillometry. This will open up new windows to investigate the impact of different types of user interfaces as used in the current computer-assisted vocabulary learning environment.

The subjects used in this second study will be taken from the pool of participants who were engaged in the first study. The use of a group of participants who participated in the first study will provide the researcher with the opportunity to eliminate any external variables by better obtaining greater precision when considering individuals' characteristics. This will also help when achieving the aim of linking and comparing the results from the two studies in order to better achieve the overall goal set out in the current research. The first study revealed significant correlations between the learners' memory abilities, and their scores on the CAVA-1 post-tests across the three multimedia representations. Thus, the use of the translation recognition task might potentially reveal about the learning process that took place in the first study, and explain the mechanism behind the findings that arose.

In contrast to previous CALL research, the second study in this thesis will employ pupillometry, in addition to response time and accuracy, to achieve a better exploration of the relationship between working memory and L2 learning, using different learning interfaces. The learners are still in the earliest stages of developing their second language, and thus, the process of L2 word-to-meaning mappings might be more clearly shown, and the impact of the learning method on the mapping, as well as accessing the L2 meaning representations will be investigated. For example, ZUI learners might a priori be more sensitive to meaning than MDI learners, which might then impact their L2 word-to-concept mappings differently.

Overall, the second study will investigate three issues that should be taken into account when designing vocabulary CALL for second language learners: access to the mental representations of L2 words, the effect of the first language in the early stages of L2 word learning, and the relationship between working memory and vocabulary learning. These issues are investigated in the context of three multimedia environments that employ three different interfaces, which have thus far been little compared or investigated. This might afford information about SLA and processing, specifically highlighting the impact of CALL interventions on accessing conceptual and lexical representations in the bilingual lexicon. The assumption underlying the structure of the developing bilingual lexicon and lexical processing can be reflected in the results from the translation recognition task. This can be seen in terms of the accuracy when rejecting word pairs, and the time it takes to make this decision. Accordingly, the degree to which meaning relatedness and form relatedness cause interference offer critical empirical evidence to test the predictions of the bilingual models of lexical encoding and retrieval. This is also evidenced in the extent to which this interference is modulated by multimedia representation as a teaching method. Furthermore, the use of pupillometry should yield novel insights into the demands of working memory, as well as the different types of user interfaces employed in the current thesis.

3.1 Research questions

The second study aims to contribute to developing our understanding of the vocabulary acquisition process, in turn, the field of SLA: specifically, how multimedia CALL environments enhance vocabulary learning for learners in the initial phase of the learning process. To achieve this, the study answers the following principal research questions:

1- What is the impact of the three different interactive multimedia representations types; (MDI, GUI, and ZUI) on response accuracy, response time, and pupil responses for L2 word retrieval?

2- What kind of performances (accuracy, response time, pupil size) can be observed across the three multimedia representations (MDI, GUI, ZUI), as indicated by the EyeLink Data Viewer? Is there a correlation between these performances and differences in the participants' individual memories?

3.2 Specific research questions of the second experiment

1- Does lexical information in the L1 remain active during L2 word processing?

2- Is the activation of lexical information in the L1 for the translation equivalent, different among MDI, GUI and ZUI learners of L2 words?

3- Does access to the meaning of L2 words increase with use of more visual based teaching methods, such as GUI and ZUI?

4- Do pupillary responses differ across multimedia representations (MDI, GUI, and ZUI) when retrieving L2 words?

5- Is there a relationship between working memory capacity and pupillary responses during L2 word access?

It can be hypothesized that the participants in the first condition; i.e. those who learned via the MDI representation, which presents the L2-L1 list of words, might first access L1 lexical connections to gain access to the concept. They may process كرة, and then directly access the translation equivalent *ball*, which then activates conceptual information. Participants who studied using GUI and/or ZUI, which are visually based representations showing the picture to illustrate the target Arabic word, might be able to access the concept directly without the L1 lexical connection. Therefore, MDI learners might also be expected to spend more time rejecting a word pair when the distracter is form relates to the L1 translation equivalent that that had to be activated in order to access the concept. While GUI and ZUI learners might face more difficulties rejecting a word pair that includes meaning related distractors. High ability working memory participants are expected to experience reduced pupil sizes while accessing L2 words. However, this might be differ across the three CALL representations.

In the next section, an overview of Language production and models of lexical access will be highlighted. Then, theoretical models detailing word representation and processing in the bilingual lexicon will be presented. After which, the use of pupillometry in word learning will be reviewed.

3.3 Literature review

Information is temporarily stored and manipulated in the working memory before then being transferred to the long-term memory for more permanent storage. A huge amount of information is stored in the long-term memory. Types of information may include details of events attended, the names of different cities, or even information about how to drive a car. It is thought that much of this information is held in the long-term memory in the form of schemas of knowledge, and schemas are believed to be used during language comprehension. Eysenck and Keane (2010, p. 639) define schemas as “well-integrated packets of knowledge about the world, events, or people, stored in long-term memory." Philosophers have classified this kind of knowledge into two main types: declarative and procedural. Memory recollection is not involved in procedural memory, also known as implicit memory. Its role can be detected when someone learns to ride a bike; for instance, the improvement in their performance can be observed as a form of behaviour, and they will then be able to recollect what they have learned unconsciously during cycling (Baddeley et al., 2015). By contrast, declarative memory, which is sometimes referred to as explicit memory, is the storage of memories that can be declared through conscious remembering of events or facts. Tulving (1972) first made the distinction between two kinds of declarative memory: episodic and semantic memory. Episodic memory refers to storage of specific events and episodes that happened in a particular place and at a particular time, such as what one had for breakfast today, whereas semantic memory stores general knowledge about people, events, objects, or word meanings, with no link to a specific time or place, like when one remembers that the French expression *le petit déjeuner* means breakfast. Kormos (2011) argues that the semantic memory includes the mental lexicon, stating that, "Semantic memory includes linguistic and non-linguistic concepts as well as meaning-related memory traces associated with these concepts, while the episodic memory is the store of temporally organized events or episodes experienced in one's life" (Kormos, 2011, p. 41).

The mental lexicon is the memory system within which an enormous vocabulary is stored throughout a person's lifetime (Hulstijn, 2000). This store holds many thousands of words, each connected with other related words in lexical networks, but its exact composition is not yet clear (Schmitt, 2010). Although it stores a huge number of words, the speaker can access a single required word easily. Thus, the L1 lexicon must be systematically organized and flexible. In addition, it is thought that its organization affects the development of L2 vocabulary acquisition (Pavičić, 2008). Knowledge of another language makes the study of the mental lexicon even more complicated; it also necessitates studying not only the organization of the L2 lexicon, but also its relationship with the L1 lexicon (Pavičić, 2008). Researchers have investigated the nature of the mental lexicon structure using different techniques, such as analysing the communication strategies used by second language learners, studying the tip-of-the-tongue phenomenon, and the behaviour of aphasia patients (Aitchison, 1990). Although these studies have contributed to describing the organization of the L1 mental lexicon, their results appear to be based on assumptions about these speakers' behaviours. It has commonly been assumed that there is a space for each word in the mental lexicon, which can be represented as a three-dimensional model of a phonological network crossed by an orthographical network and criss-crossed by a semantic network (McCarthy, 1994). Yet, these links are weak and can be broken. This explains why, in some cases, a speaker can no longer pronounce a word even though they remember some of its other features, like the meaning or the part of speech. This can also be seen as evidence supporting the productive and receptive division of vocabulary (McCarthy, 1994). Drawing on this research on the mental lexicon, in this chapter an overview will be presented of the theoretical models of language production as well as word representation and processing in SLA, focusing on the structure of the bilingual lexicon.

3.3.1 Language production models and the role of the mental lexicon

L1 lexical development has been investigated more extensively than the lexical development of L2. One commonly held view of first language processing comes from Levelt's model of language production (Levelt, 1989). Levelt studied the levels of language structure and processing, and the resultant model details the stages of language comprehension and production. According to this model, the Conceptualizer, the Formulator, and the Articulator are the three stages leading to the production of language. The intentions and ideas of the speaker, which involve nonverbal messages, are generated in the Conceptualizer. The Formulator then receives these pre-verbal messages and transforms them into a phonetic plan by encoding the message grammatically and phonologically. As a result, the lemma part of the lexicon is accessed by the grammatical encoder to select the appropriate lexical item. The lexical item, which Levelt refers to as the lemma, contains the concept as well as the grammatical and morphological features. Lemmas are 'words to represent a concept' but, in addition to the concept, they also have semantic qualities, such as word type (noun, verb, adjective etc.) and the type of register from which these words originate (slang, formal etc.). Moreover, lemmas contain information about how the word can be combined with other words. Then, the phonological encoder accesses the form part of the lexicon to determine the appropriate phonological features of the word. This phonetic plan is the output from the Formulator, which is then changed to actual speaking by the Articulator. In the comprehension process, the same stages are followed in reverse.

Lexical activation precedes the selection of the appropriate word to express the intended message. Activation takes place in the conceptualization phase to encode the concept. In the conceptualization stage, when the speaker thinks about the concept SHEEP, that particular concept will be activated. Related concepts, such as LION and BEAR, will also be activated, but at lower levels than SHEEP and, therefore, SHEEP will be selected. The activation moves from the conceptual to the lexical level. The lemma *sheep* will be activated along with its semantically related lemmas, also known as its *semantic cohort*, such as *lion* and *bear,* in the mental lexicon. The lemma *sheep* receives the highest activation, so is selected. This selected target item is then phonologically encoded and subsequently used for constructing the articulatory plan (Levelt et al., 1991). Levelt's model stresses the significance of the semantic properties of the lemmas during the activation and selection processes in the Formulation phase.

The process of comprehension and production in Levelt's (1989) model is linked to the study of the lexicon. The mental lexicon is considered a central component of the model, which is divided into two parts: lemmas and forms. Therefore, the ease of activation and the subsequent selection of the word could be affected by how the word was acquired.

In an attempt to explain the relationships between L2 and L1 concepts more clearly, and to clarify the difference between word meanings and concepts, Kormos (2006) proposed a bilingual model that incorporates psychological processes of language production. Kormos combined the three knowledge stores of Levelt's (1999) latest blueprint – knowledge of the external and internal world, the mental lexicon, and the syllabary – with Tulving's (1972) long-term memory subcomponents: episodic memory, and semantic memory. Kormos argued that the bilingual mental lexicon consists of a hierarchical structure of three levels: conceptual, lemma and lexeme level (see Figure 3.1.). According to Kormos, L1 or L2 lexical items are stored in the same lexicon in the long-term semantic memory and has information spread across these three levels. The conceptual level encompasses the semantic information of the lexical item, whereas the lemma and lexeme levels contain the syntactic and morpho-phonological information, respectively. The activation flow in Kormos' model is similar to what was proposed by Levelt in its direction from the conceptual level to the lemma, and then to the lexeme level. The conceptual level in Kormos' model is seen as a store of both L1 and L2 concepts, which can be identical, shared, or even separated. Each concept at this level has a network of memory traces, which carry its distinct semantic information. When a concept is activated in the semantic memory its semantically related concepts will also be activated. For example, when the concept CHILD is activated, related concepts such as MOTHER, PLAY, and LOVE are also activated in both languages. When a selection of the intended concept is made, activation is spread to the lemma level for activating all semantically and phonologically related L1 and L2 words that can be used to convey the chosen concept. The appropriate word will be chosen, in the selected language, for lexical encoding at the lexeme level of the metal lexicon to finally be pronounced by the speaker.

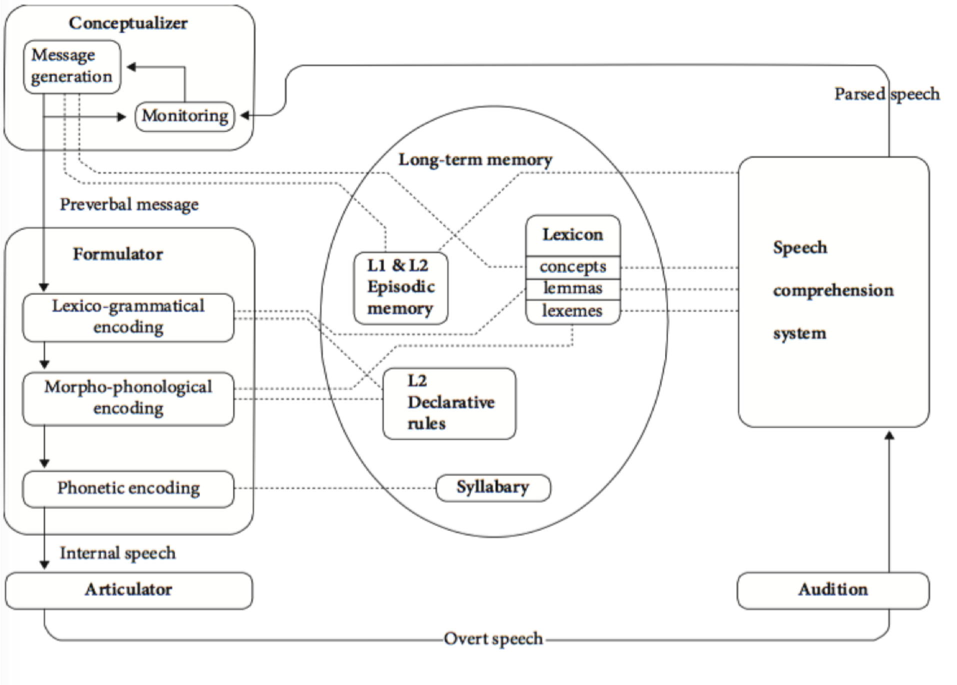


Figure 3.1. Language production model in bilinguals. (Source: Kormos, 2006, p.168)

Parallel activation in an integrated L1/L2 lexicon has been documented in previous research on word recognition (e.g. Dijkstra & Van Heuven, 2002; Van Heuven, Dijkstra, & Grainger, 1998), which led to the proposal of the Bilingual Interaction Activation (BAI) models. The fact that activation is sent to both L1 and L2 lemmas in the bilingual lexicon is evidenced in a number of studies (e.g. Colomé, 2001; Hermans, Bongaerts, & Schreuder, 1998; Poulisse & Bongaerts, 1994; Poulisse, 1999). Parallel activation research and second language production models and their adaptation to model the bilingual mental lexicon is essential for the present study, which concentrates on the effect of acquiring vocabulary via different multimedia representations on facilitating the activation and selection of the required L2 word. The focus in the next section will be on the theoretical models of conceptual (meaning) and lexical (word form) representation in the case of second language learners.

3.3.2 Accessing the meaning of the second language word

Early work by Uriel Weinreich (1953) proposed three types of bilingual lexical representations: compound, coordinate, and subordinate modes. In the compound form, L2 and L1 words are linked with similar concepts in the same store, whereas in the coordinate form two slightly different concepts are linked to the equivalent L2 and L1 words. Finally, in the subordinate form of bilingual mental lexicon representation, Weinreich argued that L2 words are not linked directly to the concept but are connected through the L1 word. Potter, Von Eckardt and Feldman (1984) drew on Weinreich's classic view and suggested their *hierarchical model*, claiming that the concept could be connected to the L2 word in two ways. The first is the *concept mediation* alternative, which is similar to the compound representation. The second alternative is *word association,* which is similar to the subordinate representation. Kroll and Stewart (1990, 1994) combined the concepts of mediation and word association in a new model called the revised hierarchical model (RHM) (see Figure 3.2.). This model assumes that the *concept* is linked with the L2 word, and the L1 word with different strengths. In the initial stages of L2 learning, it is assumed that a strong link is formed between the L2 word and its equivalent in L1, which is strongly linked to the concept (the direct connection between the L2 word and the concept is weak).

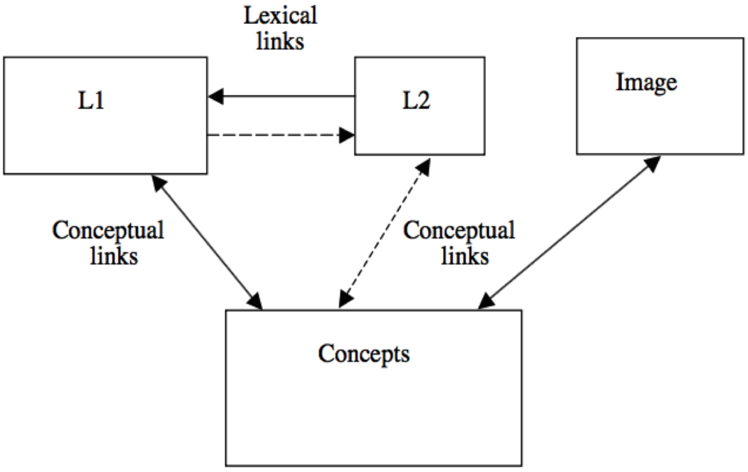


Figure 3.2. The revised hierarchical model of lexical and conceptual representation in bilingual memory. (Source Kroll and Stewart, 1994, p. 158)

The RHM assumption of accessing the L2 word meaning through L1 in the early stages of learning has been supported in previous research (e.g. Dufour & Kroll, 1995; Talamas, Kroll, & Dufour, 1999). It has also been found that this connection between the L2 word and the concept becomes stronger with the development of language proficiency (Potter et al., 1984; Chen & Leung, 1989; Kroll & Curley, 1988). However, other previous researchers have found that L2 learners are sensitive to conceptual information from the early stages of acquisition (Altarriba & Mathis, 1997; De Groot & Poot, 1997). Still other researchers have argued that the integration of L2 words into mental lexicon depends on different factors, such as the word type and grammatical class (Van Hell & De Groot, 1998), the degree of proficiency (Sunderman & Kroll, 2006), and the learning method by which the L2 lexical item was acquired (Comesaña, Perea, Piñeiro, & Fraga, 2009; Poarch, Van Hell, & Kroll, 2015). Kormos (2006,

p. 77) stated, “hierarchical models would deserve more careful consideration under what circumstances strong associative relationships between L1 and L2 lexical items are established. It might be supposed that this is not just the function of proficiency, but might depend on the context of acquisition, the methods of teaching and individual learning strategies.”

Overall, it can be concluded from the above discussion of the existing literature, and drawing on the findings of the bilingual mental lexicon experimental studies mentioned, that the majority of previous studies have focused on exploring how second language learners and bilinguals access L2 word meaning. Among these studies, little research has taken into consideration the impact of the teaching method on the mapping of L2 word meaning, and no previous studies have used a translation recognition task to explore how multimedia impact on lexical access or the word meaning representations. By contrast, in this study, the earliest stages of L2 word form meaning mappings in CALL environments will be investigated using a translation recognition task. To achieve this, the predictions of the revised hierarchical model (Kroll & Stewart, 1994) will be tested to explore how learners of different multimedia representations (MDI, GUI, ZUI) map novel words to concepts. MDI will be compared with GUI and ZUI to explore whether the L2 word link with its L1 equivalent is stronger than its connection with the concept (an illustrated picture) during the initial stages of learning a second language, as assumed in this model, or not. As such, the study will not only address this language interaction between L1 and L2, but also consider the way technology impacts the integration of novel words into the second language learner’s mental lexicon. Researching this will shed light on the impact of computer-assisted vocabulary learning during first exposure to a foreign language.

3.3.3 Pupillometry

Pupillometry is the study of changes in the diameter of the pupil, which has been used in psychology for the last 60 years (Hess & Polt, 1960; Kahneman & Beatty, 1966). Pupillometry has continued to contribute essential information to the field of psychology since that time, such as providing an estimation of the intensity of mental activities as well as changes in mental state (Laeng, Sirois, & Gredebäck, 2012). Two muscles, the dilator and constrictor, determine the size of the pupil in the human eye; changes in pupil size can result from either an inhibition of the constrictor or a stimulation of the dilator. The pupil size decreases to around 3mm in standard light conditions (Wyatt, 1995), whereas, in darkness or dim light, the pupil size increases to an average of 7mm (MacLachlan & Howland, 2002). This means that it is possible for the pupil to reach double its normal size. However, the change caused by cognitive demands is limited and rarely greater than 0.5mm (Beatty & Lucero-Wagoner, 2000).

Some researchers (e.g. Janisse, 1977; Beatty, 1982) have found that the dilation of the pupil gradually increases after the onset of stimulus, peaking at 1200ms. These responses are difficult for the individual to consciously control. One of the ways to voluntarily control the pupillary dilation is by imagining a mental image of something that normally induces a pupil dilation (Whipple, Ogden, & Komisaruk, 1992). However, the pupillary responses were found to be not only sensitive to external stimuli, such as ambient light, but also to internal stimuli, such as emotions and thoughts (Goldwater, 1972; Loewenfeld, 1993). Many general factors, such as anxiety and emotional arousal, cause tonic changes in the pupil. These tonic changes are different from the phasic changes, which are called task-evoked pupillary responses (TEPRs) and start at the onset of the stimuli. In his theory of attention, Kahneman (1973) applied the TEPRs and found that this method can be used to demonstrate individual differences in cognitive abilities. Since then, phasic changes have been used to deduce cognitive demand in different tasks, such as executive load or working memory load (e.g. Ahern & Beatty, 1979; Chatham, Frank, & Munakata, 2009; Granholm, Asarnow, Sarkin, & Dykes, 1996; Hyönä, Tommola, & Alaja, 1995; Kahneman & Peavler, 1969; Karatekin, Couperous, & Marcus, 2004; Nuthmann & Van der Meer, 2005; Piquado, Isaacowitz, & Wingfield, 2010; Stanners, Coulter, Sweet, & Murphy, 1979; Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004; Võ et al., 2008), face perception (e.g. Goldinger, He & Papesh, 2009), lexical decision (e.g. Kuchinke, Võ, Hofmann, & Jacobs, 2007), attention allocation (e.g. Karatekin, Couperus, & Marcus, 2004), and general cognitive processing (e.g. Granholm & Verney, 2004). Many of these previous studies have revealed a high correlation between pupillary dilation and cognitive load. As a result, pupillary dilation has been assumed to be a measure of the resource demand during cognitive tasks, and greater pupillary dilation is linked to high cognitive demand (Kahneman, 1973; Porter, Troscianko, & Gilchrist, 2007).

Individual differences can also be indicated by pupillary responses. For example, Just and Carpenter (1995) exposed high and low-spatial ability participants to a spatial task where they were required to judge whether a pair of irregular two-dimensional hexagons showed the same figure or mirror-image figures. The task was designed to increase the capacity utilization as the angular disparity of the degree of variation in the hexagons increased. The increase led to greater pupillary dilation for all participants, with significant changes for individuals with lower spatial abilities. In a study by Ahern and Beatty (1979), college student participants were divided into two groups according to their verbal ability scores. During mental multiplication, the group with an average score of 877 showed greater pupillary dilation than the group with an average score of 1407. These results indicate that pupil size was sensitive to individual differences among the group, as well as to the task demand itself.

Most of the previous studies in pupillometry and language processing have involved demanding post-sentential tasks, finding a connection between linguistic processing demands and pupillary dilation. However, word processing has been found to be affected by different variables that are hought to be associated with pupillary dilation, such as word frequency (Guasch, Ferre, & Haro, 2017; Kuchinke et al., 2007; Papesh & Goldinger, 2008, 2012; Schmidtke, 2014), the kind of pseudowords used in monolingual lexical decision tasks (Briesemeister et al., 2009) or bilingual neighbourhood density (Schmidtke, 2014).

Papesh and Goldinger (2012) investigated findings of memory encoding and retrieving of low- and high-frequency words using pupillometry. The researchers presented auditory words for the participants across study and test phases and identified old/new impact of auditory low- and high-frequency words on the pupils' responses for both encoding and retrieving of words. There was greater pupillary dilation when processing low-frequency words, which required higher cognitive demand.

Investigating the correlation between unexpectedness of words and pupil size grabbed the researchers' attention. Both the process of the word itself and the process driven by the expectedness combine to establish the amount of required cognitive effort (Rayner, 1998; Kliegl, Nuthmann, & Engbert, 2006). In reading studies, a relationship between cognitive load and surprising upcoming words was found, where sentences with low-predictability words required a longer time to be read (Brouwer, Fitz, & Hoeks, 2010; Levy, 2008; Frank & Thompson, 2012). Hale (2001) and Levy (2008) argued that the more surprising the word, the higher the cognitive effort required for processing is.

Recognizing old words was found to be another reason for pupillary dilation in other studies. For example, Fernández, Biondi, Castro and Agamenonni (2016) carried out a study with forty participants reading well-defined sentences, (i.e. regular sentences with less contextual predictability and proverbs). Their analysis of the pupil size behaviour showed greater increases in pupil size when reading proverbs than when reading regular sentences. This result was explained by the recognition of the words embedded in the proverbs, suggesting that pupillary dilation is not only influenced by the cognitive load. In another study of word recognition, Papesh and Goldinger (2012) had participants listen to words in study trials; they tested them by presenting the words along with one of the studied voices, a familiar oppositegender study voice, or a new voice. The results showed that pupillary dilation increased upon recognizing words when the memory was strong and specific (the voices were matched). The encoding process of these words had also caused pupil size increases, indicating that participants' effort during encoding was related to subsequent memory strength. Papesh and Goldinger's (2012) study echoes the findings of Võ et al. (2008), who revealed that participants devoted greater cognitive effort during the encoding and recognizing of old items (reflected in larger pupil size). The association of pupil reactions with memory strength was also found in a study by Kafkas and Montaldi (2011), who showed that the extent of pupillary response at encoding can predict the subsequent recognition of the image, since memory of it seemed to be strengthened.

In a recent study investigating pupil reactions in second language processing, Schmidtke (2014) conducted an experiment to gauge cognitive effort during the retrieval of spoken L2 words. The experiment examined the effects of three variables on word recognition, and pupil responses were recorded as an indication of the cognitive effort associated with the task. The three variables were the word frequency (memory strength), neighbourhood density (ND) (number of competing words), and level of English proficiency as a second language. Native Spanish participants were categorized into two groups according to the period of learning English (early/late) and compared to a monolingual English group. In total, there were 53 participants in all these three groups. In each trial, the participants were presented with four pictures including the target picture and three other distractors, which were not overlapping in shape or meaning with the target. Stimuli, which were spoken English words, were played to participants via headphones. Participants were asked to click on the correct picture out of the four choices, while their pupillary dilations were recorded. The results revealed that both groups of bilingual speakers were more affected by neighbourhood density than monolingual speakers, as they showed delayed pupillary responses in the spoken word recognition task. Importantly, more proficient bilinguals performed as well as monolinguals in terms of the weak effect of the word frequency, compared with the late bilinguals who experienced a larger frequency effect. Interestingly, the two groups of bilinguals had different pupillary responses to the earlier responses by the high-proficiency group, who had showed smaller neighbourhood density and frequency effects.

Hyönä et al. (1995) used pupillometry to investigate the cognitive load associated with word processing in both first and second language. In one of their experiments, three tasks were used: (1) listening to a single word, without any following comprehension test; (2) repeating the words back in the same language while hearing them (speech shadowing); and (3) simultaneous translation into the other language. Finnish native speakers participated in the study and the input language in the tasks (second language) was English. Seventeen subjects, fluent in English with Finnish as their native tongue, participated in the study. All three tasks required the comprehension of English, the foreign and non-dominant language. However, the output language for producing the answers was Finnish. Two sets of spoken L2 words were recorded by native speakers. The words were chosen to be different in terms of their translatability, ranging from words that could be easily translated to an equivalent word, to more difficult ones that must be translated as a phrase. In total, 90 words were chosen from each language and divided into 45 easy words and 45 difficult words, according to their translatability; 15 words from each set were used in each of the three tasks. Participants were exposed to the words as audio stimuli, and their pupillary responses were recorded during the three tasks: saying "yes", repeating the word, and translating the word. The study results showed three interesting points: (1) the pupillary responses varied according to the task difficulty. That is, the greatest pupillary dilation occurred in the translation task, followed by repeating of the words aloud, and the smallest occurred when acknowledging hearing the word. (2) In support of the view that cognitive load is associated with large pupillary dilations, translating difficult words stimulated greater pupil size than translation of easy words. To a lesser extent, the translatability impact was found to influence the other two tasks, shadowing (repeating the word) and listening. The researchers argued that this supports the word association hypothesis of the bilingual lexicon, which suggests that the lexical items in both languages are in close contact with each other. Another finding of this study also supports this notion: latency for the difficult English word was found to be longer in all three tasks, even in listening and repeating. (3) Larger pupil size accompanied the repeating of English words (non-native language) than when repeating Finnish words (native language). The researchers assumed that this was due to the greater cognitive demand involved in processing a foreign language.

To sum up, most of the research in word processing and recognition has revealed different results regarding pupillary responses, which might be due to cognitive demand or memory strength. However, the combination of other measures in the current study, such as AWMA, response time, and accuracy, will help the researcher to achieve a more thorough investigation. The use of pupillometry in the current study, which mainly investigates the impact of the three multimedia representations (MDI, GUI, and ZUI) on L2 word-learning, was motivated by three main reasons. First, there is the question of whether word processing is affected by variables found to be associated with pupillary dilation, such as word frequency (e.g. Guasch, Ferre, & Haro, 2017; Kuchinke et al., 2007) or neighbourhood density (e.g. Schmidtke, 2014). If this was the case, then, by controlling such variables in the current study, it is possible to investigate whether cognitive demand should also be affected by learning method. As such, it is expected that pupillary dilation will differ across CAVA representations as a result of the differences in the design elements used in each of the interfaces. Second, it is of interest how learners with individual differences in working memory manage to perform the translation recognition task in light of the evidence regarding the role of working memory in second language vocabulary acquisition (e.g. Atkins & Baddeley, 1998; Gathercole & Baddeley, 1990; Papagno & Vallar, 1992; Martin & Ellis, 2012), which should impact differently across CAVA interfaces. Finally, in light of the finding that pupillary dilation is also sensitive to behavioural data, such as reaction times (Kuchinke et al., 2007), the combination of pupillometry and two behavioural measures employed in this study – response time and accuracy – should provide insight regarding to what extent CAVA representations can impact the L2 word form-meaning mapping during the early stages of learning, according to the revised hierarchical model (Kroll & Stewart, 1994), and to whether the increase in pupil size upon recognition of the target L2 word is the result of memory strength or of cognitive demand.

3.3.4 Summary

Drawing on research on lexical access and language production, this chapter has presented an overview of the theoretical models of language production as well as word representation and processing in SLA, focusing then on the use of pupillometry in word processing studies. Thus, to date, little research has taken into consideration the impact of teaching method on the mapping of L2 word meaning. In addition, the majority of the research into word processing and recognition has revealed diverse results when testing pupillary responses, which has been explained in reference to either cognitive demand or memory strength.

To address this gap, a number of hypotheses will be tested to explore how learners exposed to different multimedia representations (MDI, GUI, ZUI) map novel words to concepts:

- Encoding of L2 lexis into the bilingual mental lexicon will vary according to the teaching method, rather than the level of second language proficiency.
- Learners' response accuracy and response time for L2 word retrieval will differ significantly relative to different multimedia representations (i.e. MDI, GUI, ZUI).
- High-ability L2 learners with stronger short-term and working memories will recall L2 words faster and more accurately than low-ability learners.

- Learners' mental effort (as revealed by pupillary dilation) will differ significantly when retrieving L2 words, relative to different multimedia teaching representation types (i.e. MDI, GUI, ZUI).
- Low-ability L2 learners with weaker short-term and working memory will show greater pupil dilation (as a physiological indicator of processing load) when recalling L2 words, than high-ability learners.

The methodology employed for this study ensures they can test these hypotheses reflecting on the predictions set out in the revised hierarchical model (Kroll & Stewart, 1994). That is, to explore the interference of form-related word to L1 (translation equivalent competitor) and the meaning-related word on L2 word processing. To achieve this, the experiment includes distractors that are related in form to the translation equivalent of the L2 word, and distractors that are related to the shared concept (meaning). By including these two kinds of distractors, the experiment allows the researcher to explore which type of distractors could cause more prominent interference when retrieving the L2 target word, for novice learners. The influence of the L1 on the L2 during lexical processing is investigated in this study using different teaching methods. The reaction time (RT) measurement, as well as the accuracy of L2 learners' judgments about words, will shed a light on the L1 influence on L2 word processing. Furthermore, the use of pupillometry will help to identify the impact of different multimedia representations on L2 learners' cognitive ability to retrieve the meaning of L2 words, both directly and/or via L1 words. Additionally, the AWMA results might shed more light on the kinds of psychological mechanisms that might be involved in carrying out the translation recognition task.

The current research is, to the best of the researcher's knowledge, the first study of learners who are in their early stage of second language learning in which the combination of three measures – performance accuracy, response time, and pupillometry – relevant to vocabulary learning via multimedia, is used. Given the use, for the first time, of ZUIs in researching second language vocabulary learning, along with other more common multimedia learning methods, the theoretical rationale for combining elicited CAVA-1, CAVA-2, and AWMA results is to find out whether participants' performances will be comparable in the two studies of the current thesis. This, in turn, can provide valuable insights into computer-assisted vocabulary learning.

3.4 Methodology

This section presents a description of the methodology used in the second study. It will first introduce the participants in the experiment, and then describe the design of the study, followed by the interventions. It will then outline the apparatus that were selected and designed, and the development of a second version of CAVA (CAVA-2) to present the participants with 108 words divided into 12 topics (nine words in each topic). Then, the procedure of the study will be presented. Finally, this section also explains the process of collecting, handling, and analysing the data.

3.4.1 Participants

To achieve the main goal of the present research, the participants in the second experiment were drawn from the pool of those who participated in the first experiment. Out of 50 participants in the first study, 12 students (eight females and four males) at the University of York chose to participate in the second study. The remainder of the original 50 participants were not able to participate either because they did not have normal or corrected-to-normal vision, or because they had left the university at the time the second study was conducted. Two of the 12 participants were later excluded due to their poor responses to the translation recognition task. All ten of the remaining participants had already taken the AWAM test and the researcher was in possession of information regarding their cognitive abilities, to be used later in analysing their performance in this study. Regarding the use of the eye tracker, the participants indicating that they were comfortable using the computer while their eye responses were recorded.

3.4.2 Design of the experiment

To answer the research questions, two independent variables were considered in the second experiment: multimedia representation type (three levels) and individual differences in memory abilities. Table 3.1 below provides more detail on these variables.

Table 3.1. Independent variables for the second experiment

Independent variable	**Type**	**Levels**	**Measure**
Multimedia representation	Within-subjects	**Three levels:** MDI, GUI, ZUI.	Nominal
Individual differences in working memory	Within-subjects	**Four levels:** Verbal working memory, Visuospatial working memory, Verbal short-term memory, Visuospatial short-term memory.	Nominal

Three dependent variables were examined in this study: accuracy of vocabulary recognition, response speed, and pupil size. Student acquisition of the target words was operationalized as the scores they achieved in the translation recognition task. A within-subject experimental design was used in this study as a counterbalancing design to control for practice and fatigue effects (see Table 3.2).

Table 3.2. Multimedia representation distribution in the design of the second experiment.

Lesson	Exposure Group 1	Exposure Group 2	Exposure Group 3
1- Dining table	ZUI	MDI	GUI
2- At the surgery	GUI	ZUI	MDI
3- Body parts	MDI	GUI	ZUI
4- My house	ZUI	MDI	GUI
5- An airplane	GUI	ZUI	MDI
6- Living room	MDI	GUI	ZUI
7- On the beach	ZUI	MDI	GUI
8- The head	GUI	ZUI	MDI
9- Shopping	MDI	GUI	ZUI
10- My desk	ZUI	MDI	GUI
11- Basketball	GUI	ZUI	MDI
12- My bedroom	MDI	GUI	ZUI

The Latin square design was applied in the second experiment of the current study to present the topics via three multimedia representations (as shown in Table 3.2), for participants with individual differences in memory abilities. For each participant, 108 words were split thematically into 12 lessons using the Latin square design (see Table 3.3). Correct translations were created for all 108 Arabic words learned in CAVA-2; these Arabic words were taught using three methods (MDI, GUI, and ZUI) as independent variables.

Table 3.3. The 12 nine-word lessons in CAVA-2.

Lessons	Words
طاولة الطعام Dining table	براد (Teapot) , شاي (Tea) , كأس (Mug) , نقانق (Sausages) , بندورة (Tomato) , سكين (Knife) , شوكة (Fork) , صحن (Plate) , خنزير (Bacon).
في العيادة At the surgery	مقياس (Thermometer) , مجس (Stethoscope) , مغذية (Infusion) , علاج (Pills) , إبرة (Syringe)، طاقية (Cap) , مقص (Scissors) , قناع (Mask) , قفاز (Gloves).
أجزاء الجسم Body parts	صدر (Chest) , ركبة (Knee) , رأس (Head) , يد (Hand) , بطن (Stomach) , كتف (Shoulder) , رجل (Leg) , ساعد (Arm) قدم (Foot).
منزلي My house	طبق (Satellite) , سقف (Roof) , شرفة (Balcony) , منور (Skylight), مدخنة (Chimney) , موقف (Garage) , عتبة (Step) , رتاج (Gate) , نافذة (Window).
الطائرة An airplane	جناح (Wing) , جسم (Fuselage) , عجلة (Tire) , موازن (Stabilizer) , أنف (Radom) , محرك (Engine) , مقصورة (Cockpit) , ذيل (Tail) , سلم (Stairs).
غرفة المعيشة Living room	باب (Door) , طاولة (Table) , متكى (Lounger) , كنبة (Sofa) , نبتة (Plant) , درج (Drawer) , حصيرة (Rug) , رائي (Television) , نجفة (Chandelier).
على الشاطئ On the beach	نخلة (Palm tree) , سماء (Sky) , شمس (Sun) , مظلة (Umbrella) , أريكة (Deckchair) , سفينة (Ship) , بحر (Sea) , مزلاج (SUP board) , قارب (Boat).
الرأس The head	ذقن (Chin) , جبين (Forehead) , رقبة (Neck) , عين (Eye) , وجنة (Cheek) , فيه (Mouth) , أذن (Ear) , شعر (Hair) , حاجب (Eyebrow).
التسوق Shopping	سعر (Price) , ماكنة (Card machine) , زبون (Customer) , خزينة (Till) , مال (Money) , كيس (Bag) , محفظة (Wallet) , فاتورة (Receipt) بائع (Cashier).
مكتبي My desk	شاشة (Screen), سماعة (Speakers), دباسة (Stapler), فأرة (Mouse), مذياع (Microphone), مفاتيح (Keyboard) , خرامة (Puncher) , طابعة (Printer), معالج (System Unit).
كرة السلة Basketball	جزمة (Trainers) , كرة (Ball) , ملعب (Court) , شبكة (Net) , سترة (Jersey) , سروال (Shorts) , صفيحة (Board) , شراب (Socks) , إطار (Rim).
غرفة نومي My bedroom	مسند (Bolster) , جارور (Dresser) , سرير (Bed) , مصباح (Lamp) , ستارة (Curtains) , ساعة (Clock) , لحاف (Blanket) , مخدة (Pillow) , لوحة (Picture).

For these 108 pairs of words (Arabic word and its translation), the distracter families for the critical pairs were generated; this will be described in more detail in the instrument section below. In total, 432 distracters (216 related and 216 unrelated) were created for inclusion in the experiment as *No* conditions (not a translation).

3.4.3 Interventions

The second version of CAVA was designed following the same three main criteria as the first version, regardless of the number of topics. The main criteria were: (1) ensuring that each participant was exposed to all three representations (MDI, GUI, and ZUI); (2) each multimedia representation was assessed with all of the target words; and (3) each representation was shown to learners with different working memory abilities.

For this study, there were three exposure groups. The participants who were divided randomly into these groups. Participants in exposure group one were exposed to nine words in each of the following: in lesson one (Dining table) by means of the zoomable representation (ZUI); in lesson two (At the surgery) by means of the graphical representation (GUI); and in lesson three (Body parts) by means of the menu-driven representation (MDI). Then, they were exposed to four words in lesson four (My house) via the zoomable representation (ZUI), four words in lesson five (An airplane) via the graphical representation (GUI), and four words in lesson six (Living room) via the menu-driven representation (MDI). After that, participants were exposed to nine words in each of the following: in lesson seven (On the beach) by means of the zoomable representation (ZUI); in lesson eight (The head) by means of the graphical representation (GUI); and in lesson nine (Shopping) by means of the menu-driven representation (MDI). Then, they were exposed to four words in lesson 10 (My desk) via the zoomable representation (ZUI), four words in lesson 11 (Basketball) via the graphical representation (GUI), and four words in lesson 12 (My bedroom) via the menu-driven representation (MDI).

In exposure group two, participants were exposed to nine words in lesson one (Dining table) via the menu-driven representation (MDI), in lesson two (At the surgery) via the zoomable representation (ZUI), and in lesson three (Body parts) via the graphical representation (GUI). This was followed by exposure to four words in lesson four (My house) via the menu-driven representation (MDI), four words in lesson five (An airplane) via the zoomable representation (ZUI), and four words in lesson six (Living room) via the graphical representation (GUI). After that, participants were exposed to nine words in lesson seven (On the beach) via the menu-driven representation (MDI), in lesson eight (The head) via the zoomable representation (ZUI), and in lesson nine (Shopping) via the graphical representation (GUI). This was followed by exposure to four words in lesson 10 (My desk) via the menu-driven representation (MDI), four words in lesson 11 (Basketball) via the zoomable representation (ZUI), and four words in lesson 12 (My bedroom) via the graphical representation (GUI).

Finally, in exposure group three, participants were exposed to nine words in lesson one (Dining table) via the graphical representation (GUI), four words in lesson two (At the surgery) via the menu-driven representation (MDI), and four words in lesson three (Body parts) via the zoomable representation (ZUI). This was followed by being exposed to four words in lesson four (My house) via the graphical representation (GUI), four words in lesson five (An airplane) via the menu-driven representation (MDI), and four words in lesson six (Living room) via the zoomable representation (ZUI). After that, participants were exposed to nine words in lesson seven (On the beach) via the graphical representation (GUI), four words in lesson eight (The head) via the menu-driven representation (MDI), and four words in lesson nine (Shopping) via the zoomable representation (ZUI). This was followed by being exposed to four words in lesson 10 (My desk) via the graphical representation (GUI), four words in lesson 11 (Basketball) via the menu-driven representation (MDI), and four words in lesson 12 (My bedroom) via the zoomable representation (ZUI).

3.4.4 Apparatus and instruments

The second version of the CAVA program (CAVA-2) was developed by the researcher to teach Arabic as a foreign language to beginner learners. It is an example of intentional vocabulary learning software, and was used to conduct the second experiment of this research and address the main research questions. To answer the questions of the second study, CAVA-2 was developed in the same way as the first version. The number of L2 words was increased in CAVA-2, to 108 Arabic high-frequency nouns, with nine words in each of 12 lessons, divided by theme (see Table 3.3 above). In contrast to the first version of CAVA, where all three of the main aspects of word knowledge were taught, the information provided in CAVA-2 was restricted to only form–meaning mapping without teaching the use of the target words (see Table 3.4). CAVA-2 makes use of three user interfaces (MDI, GUI, and ZUI) to create the form–meaning link in different ways.

Table 3.4. Aspects of word knowledge covered by the information in CAVA-2.

Aspect of word knowledge	Information	Purpose in CAVA
Form	Written form Spoken form	Default information (all) Default information (all)
Meaning	L1 translation Picture Visuospatial association	Representation type (MDI) Representation type (GUI) Representation type (ZUI)

In addition to eliminating the use aspect of the target L2 word, the number of practice exercises used in the first version of CAVA to teach the use of the target word was reduced in CAVA-2. These changes to the CAVA-2 content were intended to focus more on exploring the impact of new variables in the second study (response accuracy, response time, and pupil size). The remainder of the default information (word form) and the representation type information (word meaning) used in CAVA-2 are similar to in the first version of CAVA. However, the presentation duration of each word to be learned in CAVA-2 was set to be 13 seconds, starting once the participant clicks on the selected word.

Figure 3.3 shows a sample of the flow of CAVA-2 in three lessons. The program flow is consistent in all lessons for all three experimental representations. The CAVA-2 program included two phases, one study phase and one practice phase, followed by the immediate post-test (*translation recognition task*).

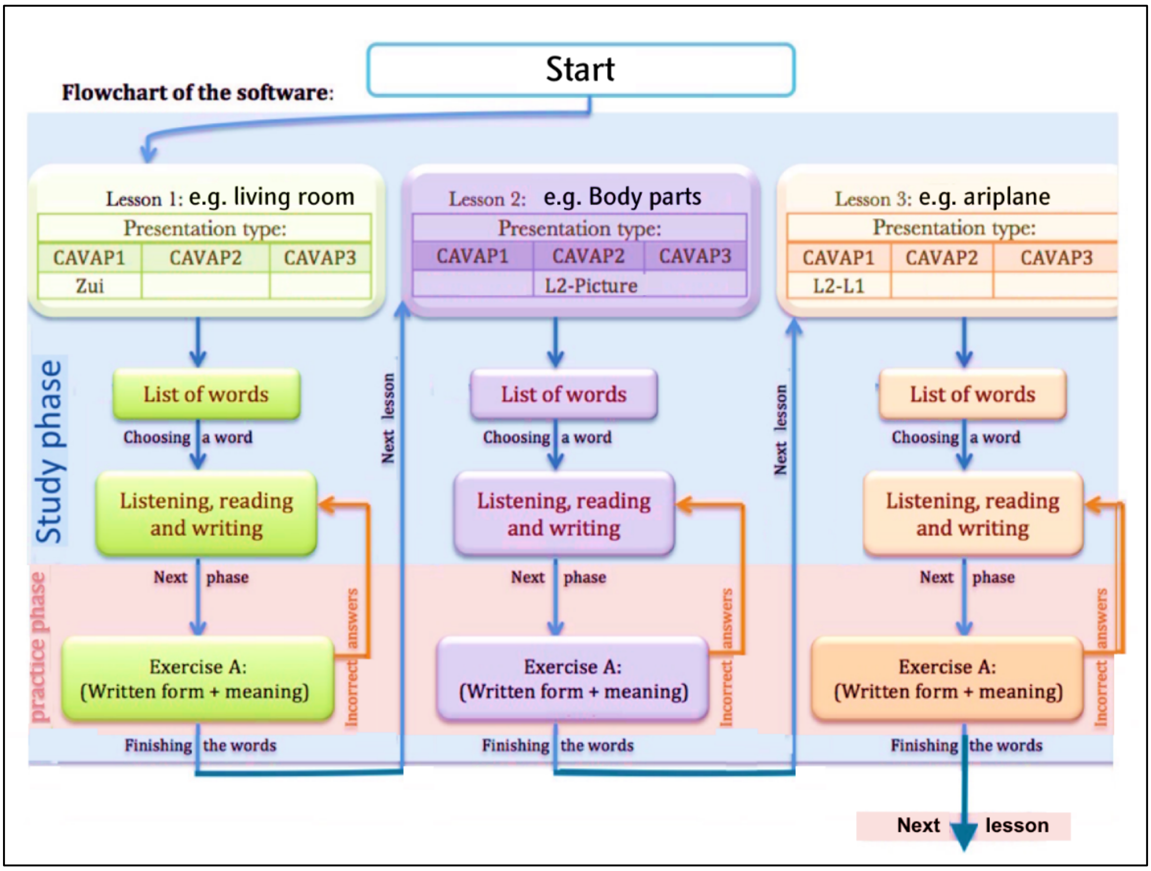


Figure 3.3. CAVA-2 program flow.

In addition to using CAVA-2 as instructional software, and AWMA for measuring participants' cognitive abilities, a translation recognition task was developed and used in this study. For all

108 target L2 words that were taught via the twelve topics of CAVA-2, correct English translation equivalents were paired and included in the task. For each word pair, such as سفينة - *Ship*, two distracters were constructed for the two main conditions of the task: (a) lexically related (written form-related) to the translation equivalent, the second item of the pair, (i.e. an orthographic neighbour in English to the word *ship*), or (b) meaning-related.

Table 3.5 shows how the related word pairs were created in the two critical conditions of the translation recognition task. As shown in the table, the correct translation equivalent of the target Arabic word سفينة is the English word *Ship*. The word *Shield* is orthographically related to the word *Ship*. The similarity in orthography was operationalized as having similar word onset, including first two phonemes of the word. The second condition, as shown in the table, is constructed using the distractor *Train,* which is related in meaning.

Table 3.5. Illustration of materials used in the two conditions for the pair سفينة- *Ship.*

Correct translation pair	Form condition		Meaning condition	
	Related	Unrelated	Related	Unrelated
سفينة - *Ship*	سفينة - *Shield*	سفينة - *Vessel*	سفينة - *Train*	سفينة - *Radio*

In each of the two conditions, two distractors were constructed, form-related and form-unrelated terms associated with the English translation, as well as meaning-related and unrelated terms. Unrelated terms were used to control for the related ones. The following criteria were considered when designing the distractors of the correct word pair (e.g. سفينة- *Ship*):

Form-related terms:

The terms that were chosen to be lexically related to the translation equivalent were required to be from the same part of speech (noun), be high-frequency words, have at least two phonemes in common, and did not need to be the same length as the translation equivalent (e.g. *Ship* and *Shield)*.

Form-unrelated terms:

This was a control condition to contrast with the form-related condition. Words needed to match each other in terms of the part of speech they were taken from (nouns), their length, and frequency, but not relate to each other in form or in meaning (e.g. *Shield* and *Vessel)*.

Meaning-related terms:

Each term semantically related to the correct word pair was required to be a noun taken from the semantic set of the target concept. However, it had to be lexically unrelated to the translation equivalent. Since the students learned the target Arabic words grouped thematically in CAVA-2 over twelve topics, the meaning-related distractors were carefully chosen to be unlearned (e.g. *Ship* and *Train)*.

Meaning-unrelated terms:

This control condition was included in the task to contrast with the meaning-related distractors. The semantically unrelated term matched the related term in the part of speech it was taken from (nouns), the number of letters (length), and frequency, but was unrelated in form and meaning (e.g. *Train* and *Radio)*.

These criteria were applied for each of the distracters for any given item. Most importantly, unrelated distracters matched their respective related terms in word length and frequency.

The goal of generating an individual unrelated distracter for both form-related and meaning-related conditions was to eliminate the impact of the distractors' other lexical properties on translation recognition performance. Matching the frequency and word length of the unrelated distracters to the related terms ensures the difference is only in lexical form for the first condition, and in meaning for the second condition. In such cases, if the response time taken by the participant to judge the form-related pair سفينة – *Shield* was longer than to judge the unrelated form pair سفينة – *Vessel,* this delay might be due to the interference of the lexical form's similarity with the translation equivalent *Ship* and not the other lexical properties of the words *Shield* and *Vessel,* which are both high-frequency nouns of the same length. The second condition, meaning distractors, was designed in the same way. The semantic-related and semantic-unrelated pairs were matched in frequency and word length within the meaning conditions, but were not matched with the form conditions.

3.4.5 Procedure

The second research experiment sessions were conducted during the autumn term of 2017. The process of testing participants lasted one month and took place in the eye tracking laboratory at the University of York. Participants (*n* = 10) were contacted via email to arrange a suitable time for their sessions. Once a participant arrived at the eye tracker laboratory, they were asked to read and sign a consent form (see Appendix 2). Then, the researcher explained the process of studying CAVA-2 and the subsequent post-test. At this meeting, the participant was exposed to the second version of CAVA and completed the study of 108 words over the twelve topics in four sessions. Each session consisted of three lessons, each including one study phase and one practice phase. Then, after a short break, the participant was tested individually after receiving verbal instructions from the researcher in addition to instructions that appeared on the computer screen. The participant was asked to sit in front of the computer screen at a viewing distance of 70cm, with their chin on a chinrest and forehead on a forehead bar. The camera was set up, and eye tracker calibration concluded with a nine-point calibration routine. The laboratory was illuminated only by the computer monitors, to avoid external influence on pupil size. In addition, the researcher made sure that the screen the test was presented on did not vary in luminance, to ensure that there would be no variation in pupil size as a reaction to environment light. Five practice trials were conducted before commencing the main task. At the end of the meeting, all participants were given a reward for participating (£8 per hour) and soft copies of their CAVA-2 and post-test scores.

The eye tracking laboratory is equipped with two main computers. First, the Host PC, a standard workstation that processes and records eye movement data at a rate of 1000 samples per second, and stores them in a data file. This Host PC is linked via a local Ethernet link to a second computer with a display screen, on which the actual experiment is presented to participants. Sitting below the tracked area to capture the participant's pupillary responses, the EyeLink 1000 eye tracker device (SR Research Ltd., 2005–2008) is a video-based desk mounted eye tracker, which uses an infra-red light source with a sampling frequency of 1000 Hz (average accuracy 0.25°– 0.5°, resolution 0.01° RMS).

The EyeLink 1000 is integrated with SR Research Experiment Builder, a software package that is used to design different kinds of research studies. In addition, there is a chinrest and forehead bar in front of the display screen, on which the participant rests their head. The mount is important for minimizing head movements while presenting the task to achieve the best possible accuracy and maintain the viewing distance of 60 cm from the screen (see Figure 3.4.).

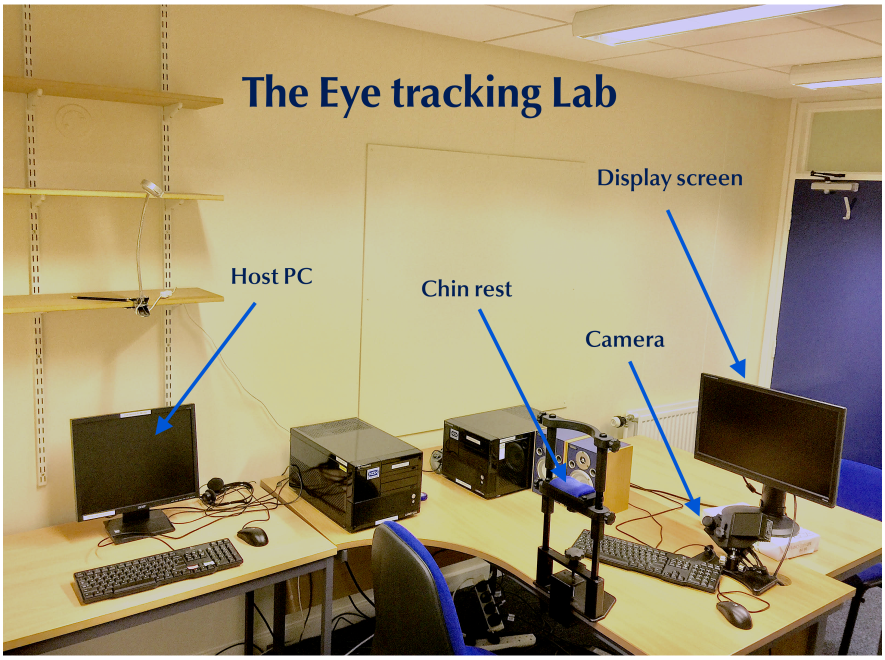


Figure 3.4. The eye tracking laboratory.

To start the procedure of the translation recognition task, the researcher explained in detail what the test was about, how many parts it consisted of, the number of questions, and how to respond to them. Written instructions were also provided on the screen at the beginning of the practice trials, and repeated at the beginning of the main task. Before staring the task, a simple test was conducted to discover which was the participant's dominant eye to be tracked by the camera during the test. The participant was then asked to sit on a comfortable chair, at approximately 60 cm in front of the display screen, placing their chin on the chin rest and forehead on the forehead bar.

The initial phase of the test was the calibration, which was conducted to ensure the accuracy of each participant's eye measurements. The participant was asked to look at the central fixation stimulus, a small black circle, and press the spacebar to confirm they were focusing directly on the target. The circle initially appeared in the centre of the screen; once a point had been confirmed, the circle reappeared in one of eight points distributed across all parts of the screen, as shown in Figure 3.5 below. The participant was required to fixate on the circle at all nine points. Then, the software showed whether the participant had a problem with viewing any part of the screen.

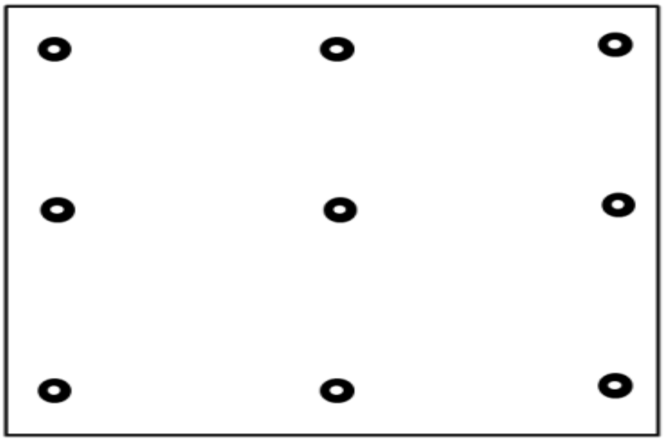
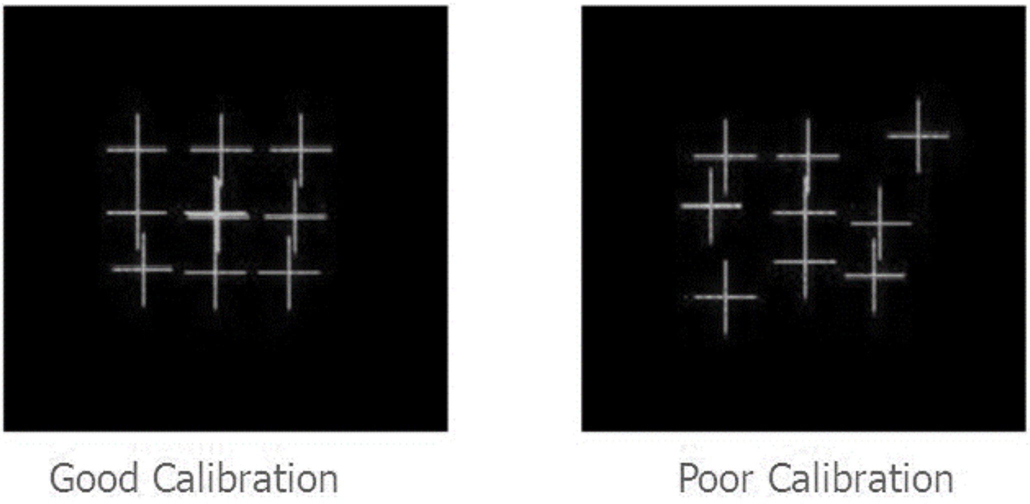


Figure 3.5. Distribution of nine-point calibration.

In order to validate the recorded eye positions, subjects were asked to complete the same task a second time. Calibration was only accepted once the difference between calibration and validation was less than 0.5º overall. The calibration was not accepted if the error was very high for tracking the eye, as this could affect the accuracy of the results. Figure 3.6. shows examples of good and poor calibrations. Once calibration had been successfully completed, the participant proceeded to the next step, the translation recognition task. If calibration was not successful, the participant would be excluded. All twelve participants in the current study completed the calibration successfully; however, two of the twelve participants' responses were excluded from the analysis phase due to poor performances during the task.



Good Calibration

Poor Calibration

Figure 3.6. Examples of good and poor calibrations.

During the task, each participant saw three sets of 180 trials, a total of 540 trials. Of the 540 trials, 108 were *Yes* trials and 432 were *No* trials. Five practice trials were designed and preceded the main task. The task was to judge whether the pairs of words presented were translation equivalents or not. The participant was asked to initiate the trials by pressing any key on a keyboard connected to the computer. The words were presented so that the Arabic (L2) word appeared first, followed by the English (L1) word. The presentation of each pair of words was preceded by a fixation cross shown in the centre of the screen. As illustrated in Figure 3.7., for each trial the fixation cross remained for 500 ms and was then replaced by the Arabic word for 1600 ms. The duration for which the Arabic word was shown was chosen based on what past researchers (e.g. Janisse, 1977; Beatty, 1982) have reported on pupil dilation, specifically that it increases gradually after the onset of stimulus, peaking at 1200 ms. During the Arabic screen presentation, the pupillary responses were recorded. This screen was followed by a brief (400 ms) blank screen, and then the English word appeared in the same position and remained on the screen until the participant pressed either the *Yes* or the *No* button, or for 3000 ms.

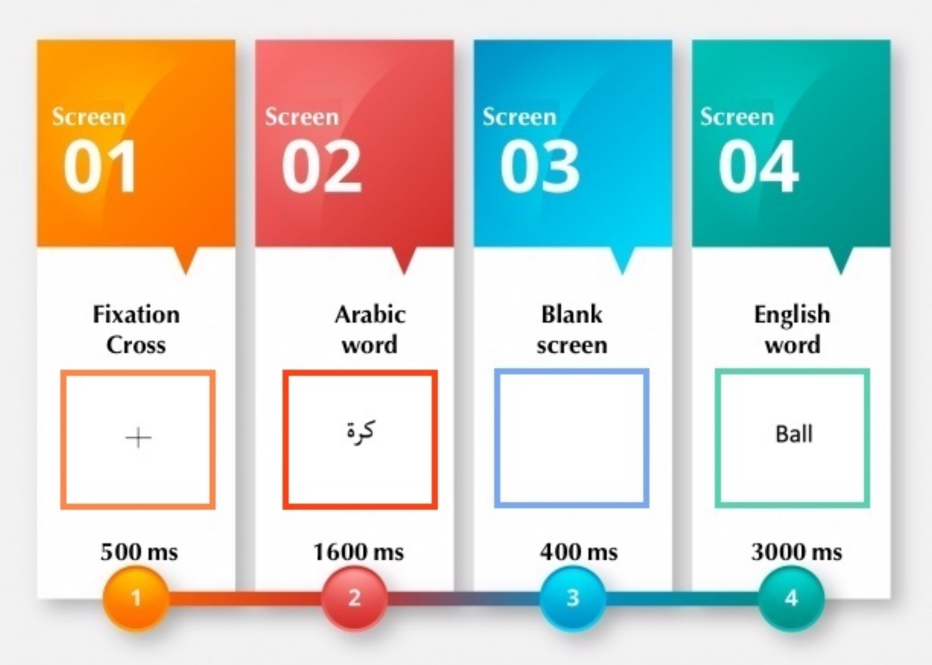


Figure 3.7. Sequence and durations in ms of the screens for each trial in the translation recognition task.

The participant was instructed to respond as accurately and quickly as possible, and, if they were unsure, to guess. If the participant did not make a decision during the presentation of the English word (3000 ms), they were moved to the next trial and their response was marked as incorrect. In addition to accuracy, response time was recorded to the nearest millisecond from the onset of the presentation of the English word, and pupillary responses were also recorded during the task. The presentation order of the word pairs was randomized for each participant.

3.4.6 Analysis

The inferential results for the research questions of the study were answered by conducting a three-way mixed ANOVA with individual differences in memory and multimedia representation as within-subject factors. The research questions of the second study, the relevant effects that were tested, and the inference tests used are summarized in Table 3.6 below. The alpha level to determine statistical significance for all the tests was set to .05.

Table 3.6. Research questions and the inference test in the second study.

Research question	Effect tested	Inference test
1- What is the impact of the three different interactive multimedia representations types; (MDI, GUI, and ZUI) on response accuracy, response time, and pupil responses for L2 word retrieval?	Main effect of multimedia representation types	ANOVA ($\alpha = .05$)
2- What kind of performances (accuracy, response time, pupil size) can be observed across the three multimedia representations (MDI, GUI, ZUI), as indicated by the EyeLink Data Viewer? Is there a correlation between these performances and participants' individual differences in memory?	Main effect of working memory abilities	ANOVA ($\alpha = .05$)

3.4.7 Summary

This section has presented the methodology used in the second study to answer the research questions. The selection of the research design and methodology for this study followed the first study of the current thesis. A mixed methods design was used to address the different requirements of the third and fourth research questions. The previous AWMA results recorded in the first study were used in this study also, and a second version of CAVA was developed by the researcher to generate comparative data. This contained the same three multimedia representations used in the first version but with a larger number of topics and words. The aim of using different output measures in this second study (response time, accuracy, and pupillometry) as well as AWMA was to better investigate the educational software used by providing comparative data on lexical storage and access. The next section will present the results of the second experiment.

3.5 Results

The aim of the second main experiment was to investigate, in greater detail than the first experiment, how the different multimedia vocabulary CALL representations used in CAVA affect the process of learning novel L2 words. The following sections will present the results for 10 participants in a translation recognition task and the correlation to their short-term and working memory abilities.

The research questions of this second study focus on the performances of participants (*n* = 10) during lexical access to respond to a translation recognition task, and compare these performances in the MDI word trials with the GUI word and ZUI word trials. After examining the performances (accuracy, response time, and pupil size) using the EyeLink Data Viewer, the research questions will be answered in two sections: the first will discuss participants' performances in terms of accuracy, response time, and pupil size, comparing these with the participants' scores in the MDI, GUI and ZUI word pairs. Second, the performances (accuracy, response time, and pupil size) will be compared against the participants' individual differences in memory, which were assessed using AWMA.

3.5.1 Translation recognition task results

The first topic of the second study of this research examines the effectiveness of multimedia representations in CALL environments for second language vocabulary acquisition by investigating the main effect of multimedia representations that employ different interactive user interfaces (i.e. MDI, GUI, ZUI) on vocabulary acquisition. This can be reflected by the accuracy of the participants' responses to both the *Yes* and *No* conditions in the translation recognition task.

3.5.1.1 Accuracy

With regard to the main effect of multimedia representations (RQ 3), Table 3.7 shows the accuracy results for the three multimedia representation types in the translation recognition task, for the ten participants of the second study. The table presents the percentage accuracy of translation recognition judgments for true and false translation pairs. Each participant studied nine words in each of the twelve lessons, a total of 108 words.

Table 3.7. Percentage accuracy results in the translation recognition task for the three multimedia representation types.

Average of accuracy	Teaching methods:			
Conditions:	MDI	GUI	ZUI	Grand Mean
Control	56%	32%	48%	45%
Translation equivalent	56%	32%	48%	45%
Critical	96%	96%	97%	96%
Form related	97%	98%	98%	97.7%
Form Unrelated	98%	99%	98%	98.7%
Interference	1%	1%	0%	1%
Meaning related	89%	89%	93%	90%
Meaning Unrelated	99%	97%	98%	98%
Interference	10%	8%	5%	8%
Grand Total	88%	82%	87%	86%

Descriptive statistics

Table 3.7 shows the percentage of participants' judgements that were correct in terms of whether the pairs of words (Arabic and English) were translation equivalents or not. The scores represent vocabulary acquisition, which is reflected in the participants' ability to respond to five types of conditions, divided into two main categories (Yes, and No). The results are displayed separately for the three teaching methods (MDI, GUI, and ZUI). The most striking outcome was that the percentage of correct 'No' answers was significantly higher than correct 'Yes' responses for all three teaching methods. That is, participants were better at correctly rejecting an incorrect translation than they were at accepting a correct one. Looking at this in more detail, participants were more accurate in giving 'Yes' responses to words that they had learned via MDI, (56%), than ZUI and GUI, (48% and 32%, respectively). However, the effectiveness of these three teaching methods was very similar in the 'No' responses (ZUI = 97%, MDI = 96% GUI = 97%). Looking at the 'No' conditions, however, there were other interesting outcomes: the meaning-related condition was the most effective distractor for the participants regardless of the teaching methods used; meaning relatedness caused the highest interference, which resulted in the lowest accuracy, with a mean of 90% compared with 97.7%

in form-related conditions, 98.7% in form-unrelated conditions, and 98% in meaning-unrelated conditions. However, the impact of all the teaching methods was relatively similar in all four ‘No’ conditions.

What is immediately apparent is that the meaning-related false translation pairs produced more interference than form-related pairs in all cases. Furthermore, the magnitude of interference was related to the teaching methods, where MDI yielded the largest meaning-relatedness interference at 10%.

The following tables and figures show the percentage accuracy for the three multimedia representations (MDI, GUI, and ZUI) in the translation recognition task. Figure 3.8. presents a bar graph of the results of the two main categories of distractors (Yes, and No), while graph 3.9. shows the accuracy results for the four critical (No) conditions.

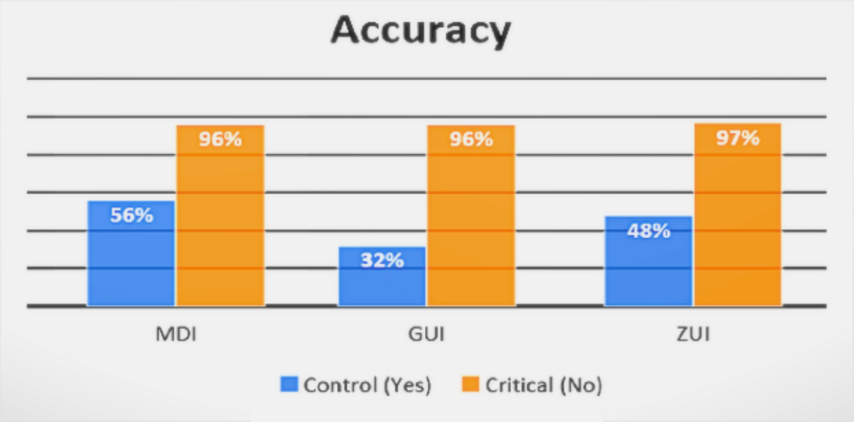


Figure 3.8. A bar chart showing the accuracy means in both the control and critical conditions across the three multimedia representations.

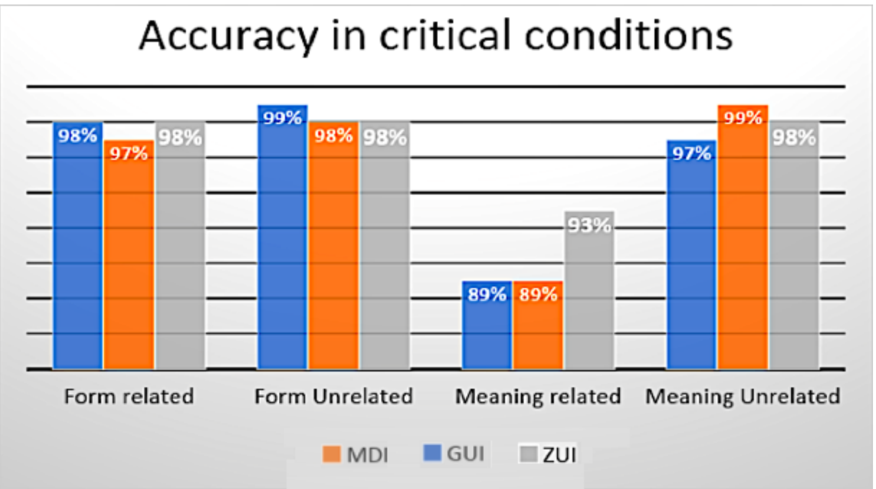


Figure 3.9. A bar chart showing the accuracy means in the four critical conditions across the three multimedia representations.

Inferential statistics

In regard to the accuracy of judging whether or not the word pair was a translation equivalent, a main effect of multimedia representation was tested in the formal hypothesis with an alpha level of .05. However, there was no significant difference in the effectiveness of the three multimedia representation types in the acquisition of second language vocabulary. Post-hoc multiple pairwise comparisons showed that multimedia representations via ZUI are more effective for vocabulary acquisition than GUI and MDI multimedia representations (see Table 3.8).

Table 3.8. Pairwise comparisons for multimedia representation
in the translation recognition task.

Multimedia representation I	Multimedia representation J	Mean difference (I – J)	Standard error	Sig.
ZUI	MDI	4.88	3.5	0.17
	GUI	2	3.5	0.57
GUI	MDI	2.88	3.5	0.41

Participants' overall accuracy in performing the translation recognition task in the 'No' trials was significantly higher than in the 'Yes' trials. Two-way repeated measures ANOVA, with teaching method as the within factor and type of trial as the between factor, were performed on percentage accuracy for 'Yes' and 'No' responses. The results revealed a number of significant interactions between teaching methods and the translation pairs. All analyses were performed using both participants and items as random factors.

Looking at the results of the accuracy measure, this showed that the overall mean of the accuracy for correct translation pairs (i.e. 'Yes' trials) ($M = .46$, $SD = .191$), was lower than the overall mean of accuracy for the incorrect translation pairs (i.e. 'No' trials) ($M = .96$, $SD = .24$), and the difference was statistically significant ($t\ (9.2) = 8.2$, $p = .00 < .05$).

The mean for accuracy in regard to the correct translation pairs was highest for MDI among the different teaching methods ($M = .56$, $SD = .22$), compared with GUI ($M = .32$, $SD = .14$), and ZUI ($M = .47$, $SD = .23$), and the difference between MDI and GUI was significant (t (18)

= 2.8, *p* = .01 *). However, the difference between MDI and ZUI was not significant (*t* (18) = .8, *p* = .42).

Amongst the translation pairs, on the other hand, analyses also revealed statistically significant differences between the overall mean for accuracy between the meaning-related pairs (*M* = .90, *SD* = .05) and the other critical conditions (*M* = .97, *SD* = 24) with a *t*-test result of *t* (18) = 4.1, *p* = .00 < .05.

Furthermore, the percentage of correct responses was also analysed according to the three teaching methods. In each translation pair, paired samples *t*-tests were conducted within each condition to determine how L2 word accessing was affected by the different kinds of relatedness. Looking closely at the meaning-related pairs, the findings reveal other differences based on the type of teaching method. The zoomable interface representation, the visuospatial-based method, caused the least semantic interference compared with the other multimedia representations used. The mean accuracy in rejecting meaning-related pairs (*M* = .92, *SD* = .062) was lower than in rejecting the unrelated pairs (*M* = .97, *SD* = .033) and the difference was significant (*t* (18) = 2.2, *p* = .038 < .05). However, the interference of semantic relatedness was significantly higher in MDI (*t* (18) = 5.8, *p* = .000 < .05) and GUI (*t* (18) = 3.2, *p* = .004 < .05).

3.5.1.2 Response time

The second topic of the second study of this research examines the impact of the three multimedia representations as teaching methods for second language vocabulary acquisition. This was addressed by exploring the main effect of multimedia representations that employ different interactive user interfaces (i.e. MDI, GUI, ZUI) on the speed of the learners' responses in a given task. This can be studied by investigating the latency of participants' responses to both the 'Yes' and 'No' conditions in the translation recognition task.

With regard to the lexical and meaning processing during L2 word recognition, the complete results for all five translation trials are shown in Table 3.9, where response latencies in the translation recognition task are given for each teaching method (MDI, GUI, and ZUI) in each type of translation trial ('Yes' and 'No' conditions).

Table 3.9. Results for accuracy in the translation recognition task for the three multimedia representation types.

Average of RT (ms) 10	Learning Styles:			
Conditions:	MDI	GUI	ZUI	Grand Mean
Control (Yes)	1317	1353	1427	1360
Translation equivalent	1317	1353	1427	1360
Critical (No)	838	849	819	835
Form related	814	843	780	813
Form unrelated	809	819	800	809
Interference	5	24	20	4
Meaning related	946	942	896	928
Meaning unrelated	789	801	807	798
Interference	157	141	89	130
Grand Total	900	891	882	891

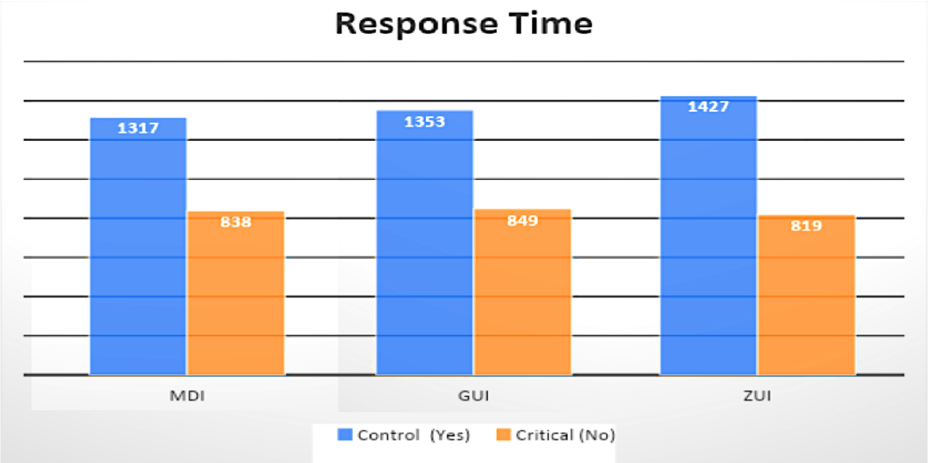


Figure 3.10. A bar chart showing the response time means in both control and critical conditions across the three multimedia representations.

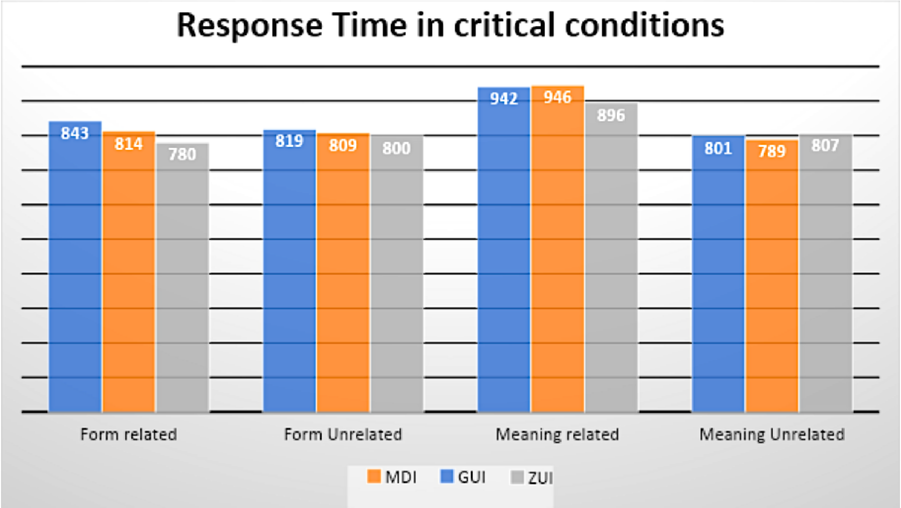


Figure 3.11. A bar chart showing the response time means in the four critical conditions across the three multimedia representations.

Descriptive statistics

Examination of the response time revealed that students took more time to respond correctly in ‘Yes’ conditions (1360 ms) than in ‘No’ conditions (835 ms) (see Table 3.9). The results also showed that MDI was the most effective teaching method in terms of response speed for the correct Yes condition. In judging the English word as the correct translation of the target Arabic word, participants were faster with MDI words than GUI and ZUI words (see Figure 3.10.). Thus, it can be assumed that participants found MDI words the easiest. The mean response time in judging ‘Yes’ conditions for MDI was 1317 ms, followed by GUI at 1353 ms, and then finally ZUI with a mean of 1427 ms.

However, the results of the ‘No’ conditions revealed different order of the teaching method in response time (see Figure 3.11.). Looking at the total mean of correct ‘No’ judgements, the participants were faster in ZUI with 819 ms. This was followed by MDI, and then GUI, at 838 ms and 849 ms, respectively. These results for the four ‘No’ critical conditions affected the overall mean of all the included conditions. In the overall mean of response time, the multimedia representation (ZUI), achieved the fastest mean score, at 915 ms. Thus, it can be assumed that this method best supported vocabulary acquisition in the translation recognition task. The second best multimedia representation was GUI (graphic-based L2–picture user interface), with a mean of 922 ms, followed by MDI, a type of verbal-based L2–L1 user interface, at 948 ms.

Moreover, the results revealed another striking outcome among the ‘No’ conditions. Participants took the longest time to respond in the meaning-related condition than the other three ‘No’ conditions (form-related, form-unrelated, and meaning-unrelated). This was true for all three teaching methods (MDI, GUI, and ZUI), suggesting high interference between the meaning-related and meaning-unrelated conditions. MDI caused the highest semantic interference at 157 ms.

Inferential statistics

In regard to the response time to judge whether the word pair was a translation equivalent or not, a main effect of multimedia representation was tested in the formal hypothesis with an alpha level of .05. The effectiveness of the multimedia representation types in the acquisition of second language vocabulary was not significantly different among the five translation pair types in the response time. Post-hoc multiple pairwise comparisons showed that multimedia representations (ZUI) tended to be more effective for vocabulary acquisition than multimedia representations GUI and MDI (see Table 3.10).

Table 3.10. Pairwise comparisons for multimedia representations in the translation recognition task.

Multimedia representation I	Multimedia representation J	Mean difference (I – J)	Standard error	Sig.
ZUI	MDI	4.88	3.5	0.17
	GUI	2	3.5	0.57
GUI	MDI	2.88	3.5	0.41

The analysis of the response time measure in the translation recognition task revealed statistically significant differences within both the learning method factor and the translation pair, as follows:

Time spent to provide a correct answer for the 'Yes' word pairs (*M* = 1348 *SD* = 132) was longer than for the 'No' pairs (*M* = 837, *SD* = .132). This difference was significant (*t* (18) = 5.3, *p* = .0001).

In terms of the impact of the teaching methods, the mean time to provide a correct response for the control ('Yes') word pairs for words learned via MDI (*M* = 1327, *SD* = 287) was not significantly shorter than for GUI (*M* = 1353, *SD* = 286, p>0.05) or ZUI (*M* = 1427, *SD* = 357, p>0.05), which indicates that MDI resulted in faster correct judgements for the 'Yes' translation pairs.

Although ZUI resulted in slowest performance in the mean response time for the 'Yes' conditions, for only the critical 'No' conditions, participants were faster with ZUI (*M* = 819, *SD* = 124) than with MDI (*M* = 838, *SD* = 134) and GUI (*M* = 849, *SD* = 141). However, the difference was not significant.

For the overall mean of response time in both 'Yes' and 'No' conditions, ZUI (*M* = 882, *SD* = 136) trended toward a significant difference.

The response time results also showed that the overall mean time to judge the meaning-related translation pairs (*M* = 932, *SD* = 136) in all three teaching methods was not longer than for the three other types of 'No' translation pairs, including meaning-unrelated pairs (*M* = 800, *SD* = 145).

3.5.1.3 Pupillometry

The third topic of the second study investigates the effectiveness of multimedia representations in CALL environments for second language vocabulary acquisition by exploring the main effect of multimedia representations that employ different interactive user interfaces (i.e. MDI, GUI, ZUI) on L2 word recognition. This is reflected in the participants' pupillary responses to both the 'Yes' and 'No' conditions in the translation recognition task, which can be used to measure the cognitive demand of word recognition. Using pupil size as a proxy for mental effort helps to investigate the relative mental effort required for the L2 retrieving process learned via ZUI versus GUI and MDI. To extract the specific pupillometry data from the EyeLink Data Viewer software, the researcher needed to first establish an interest period for the English word presentation during the translation recognition task in order to apply the different measures. The interest period commenced from the onset of the English word and lasted for 3000 ms, until either the end of its display or the participant gave a response. The English word was either a translation equivalent, form-related to the translation equivalent, form-unrelated to the translation equivalent, meaning-related, or meaning-unrelated. The focus was on the pupillary responses during this interest period only, for all five conditions. To answer the third research question with regard to the main effect of multimedia representations on the cognitive demand of L2 word recognition, Table 3.11 shows pupillary responses during the presentation of the English words in the translation recognition task for the three multimedia representation types, for all ten participants of the second study.

Table 3.11. Pupillometry results during the translation recognition task for the three multimedia representation types across the five conditions.

Teaching method Word relatedness	Pupil size mean	Pupil size max	Pupil size min
MDI	539	571	505
Form related	542	570	513
Form Unrelated	534	561	498
Meaning related	540	574	506
Meaning Unrelated	534	569	504
Translation equivalent	543	587	505
GUI	538	569	504
Form related	535	563	500
Form Unrelated	547	576	518
Meaning related	530	565	497
Meaning Unrelated	530	558	497
Translation equivalent	546	585	508
ZUI	534	565	502
Form related	534	560	506
Form Unrelated	534	561	504
Meaning related	536	568	505
Meaning Unrelated	531	559	500
Translation equivalent	535	577	496
Grand Total	537	569	503

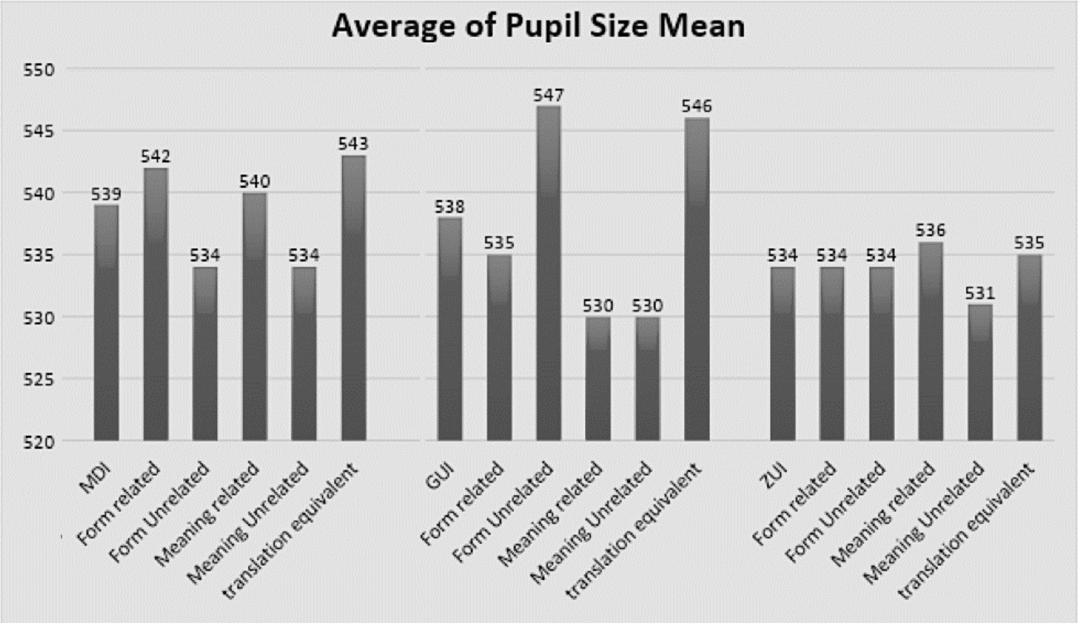


Figure 3.12. A line chart of the pupil size means in both control and critical conditions across the three multimedia representations.

Descriptive statistics

Table 3.11 and Figure 3.12. present the pupillometry results for all ten participants in the study. Responding to words learned through ZUI elicited the smallest pupil sizes (534), while the largest mean pupil size was recorded while processing words learned via MDI (539), suggesting this method was the most demanding one, followed closely by GUI (mean of 538).

For all three teaching method, the translation equivalent resulted in larger mean pupil size compared to the other forms of distracter. Another striking outcome is that the impact of form- and meaning-relatedness on pupil responses was different across all the teaching methods. Form-relatedness caused interference only for words learned through MDI, with a larger pupil size when the distracter was form-related than when it was unrelated. This effect did not appear with the other two visual-based teaching methods, GUI and ZUI. On the other hand, meaning-relatedness noticeably impacted the pupil responses when the words were learned via ZUI. Pupils were larger when the participants saw the English word in the translation recognition task.

Inferential statistics

In terms of the accuracy of judging the whether the word pair is a translation equivalent or not, a main effect of multimedia representation was tested in the formal hypothesis with an alpha level of .05. The effectiveness of the multimedia representation type and types of translation pair on the pupil size during the appearance of the English word in the translation recognition task was not significantly different. Post-hoc multiple pairwise comparisons showed that multimedia representations (ZUI) were more effective for vocabulary acquisition than multimedia representations (GUI and MDI) (see Table 3.12).

Table 3.12. Pairwise comparisons for multimedia representations on the translation recognition task.

Multimedia representation I	Multimedia representation J	Mean difference (I – J)	Standard error	Sig.
ZUI	MDI	4.88	3.5	0.17
	GUI	2	3.5	0.57
GUI	MDI	2.88	3.5	0.41

The pupillometry results for the translation recognition task revealed another dimension to how test-takers processed different translation pairs and the impact of the three multimedia representations on these. The results revealed that when the task was completed with words learned via ZUI pupil size was smaller (*M* = 533, *SD* = 281) than with GUI (*M* = 533, *SD* = 281) and MDI (*M* = 533, *SD* = 281), which induced the largest pupil sizes. The t-test results showed that there was no statistically significant difference between ZUI and GUI (*t* (18) = .03, *p* = .9 > .05), nor between ZUI and MDI (*t* (18) = .038, *p* = .97 > .05).

Participants' pupil size was larger during control conditions (translation equivalent) (*M* = 540, *SD* = 283) than in critical conditions (no translation pairs) (*M* = 534, *SD* = 284); this trend in pupil size mean was not statistically significant (*t* (18) = .047, *p* = .96 > .05).

With regard to the interference of form-related conditions, the only striking interference was caused by form-related pairs, and only with MDI. This is reflected in its trend of causing larger pupil size (*M* = 540, *SD* = 290), larger than other types of 'No' translation pairs (*M* = 535, *SD* = 283). However, the t-test results showed that there was no statistically significant difference (*t* (18) = .04, *p* = .96 > .05).

In ZUI, all the 'No' conditions had approximately the same effect on pupillary response (*M* = 532, *SD* = 280), except for the meaning-related translation pairs condition (*M* = 534, *SD* = 284), which showed a trend of slightly increasing the pupil size (*t* (18) = .011, *p* = .9 > .05).

In GUI, all the 'No' conditions had approximately the same effect on pupillary response (*M* = 531, *SD* = 284), except for the form-unrelated translation pairs condition (*M* = 546, *SD* = 292), which also increased pupil size but with no significant difference (*t* (18) = .118, *p* = .9 > .05).

3.5.2 AWMA results

Participants' scores in all three measures of the second study (accuracy, response time, and pupil size) were correlated with the participants' memory abilities in in terms of four components – verbal short-term memory, verbal working memory, visuospatial short-term memory, and visuospatial working memory – using SPSS. The correlation helped to explore whether the participants who scored higher in a specific teaching method actually possessed a specific cognitive ability. To determine this, the Pearson correlation coefficient (*r*) was used to assess the degree of the relationship between each of the study variables. Table 3.13 shows the second study participants' (N 10) AWMA mean scores.

Table 3.13: AWMA mean scores across the four components of memory for ten participants.

Participant	Verbal short-term memory	Visuospatial short-term memory	Verbal working memory	Visuospatial working memory
1	97.67	98.00	102.67	111.33
2	106.33	107.00	111.00	103.00
3	90.67	102.67	90.00	122.00
4	122.00	109.00	113.00	112.00
5	103.67	122.00	103.30	122.33
6	119.33	117.00	96.67	130.00
7	89.33	118.67	97.33	103.67
8	107.33	90.00	118.33	100.00
9	111.67	108.33	131.33	130.33
10	109.67	116.00	97.67	102.67
Total mean	105.767	108.867	106.13	116.499

It is clear from the table that each participant had differences in their abilities, reflected in their total score for each of the four memory components, calculated as the mean of the three subtests conducted for each memory component.

3.5.2.1 Accuracy

All four components of memory were found to correlate, to different degrees, with the accuracy results. The following tables will present the strongest correlations found.

Verbal short-term memory

The findings revealed that verbal short-term memory is the component most closely correlated with accuracy results, being correlated with six types of word pairs within the teaching methods.

The correlation between verbal short-term memory and the form-unrelated condition in GUI (r = .528) indicates a moderate relationship; however, this relationship is not statistically significant at p = .116 (see Table 3.14).

Table 3.14. Pearson correlations between verbal short-term memory and form-unrelated GUI.

Form-unrelated GUI	Pearson correlation	.528
	Sig. (2-tailed)	.116
	N	10

The correlation analysis showed a stronger relationship between meaning-related GUI and verbal short-term memory; yet this again was not statistically significant at r = .12, p = .052 (see Table 3.15).

Table 3.15. Pearson correlations between verbal short-term memory and y meaning-related GUI.

Meaning-related GUI	Pearson correlation	.627
	Sig. (2-tailed)	.052
	N	10

A more striking outcome for the verbal short-term memory is its significant correlation (r = .715, p = .02) with the meaning-unrelated CUI (see Table 3.16).

Table 3.16. Pearson correlations between verbal short-term memory and meaning-unrelated GUI.

Meaning-unrelated GUI	Pearson correlation	$.715^{*}$
	Sig. (2-tailed)	.020
	N	10

Another significant correlation was found between the verbal short-term memory at with the meaning unrelated MDI (r = .723, p= .018) (see table 3.17).

Table 3.17. Pearson correlations between verbal short-term memory and Meaning Unrelated MDI.

Meaning-unrelated MDI	Pearson correlation	$.723^{*}$
	Sig. (2-tailed)	.018
	N	10

The relationship between verbal short-term memory and form-related ZUI approaches showed a significant correlation (r= .620, p= .056) (see Table 3.18).

Table 3.18. Pearson correlations between verbal short-term memory and form-related ZUI.

Form-related ZUI	Pearson correlation	.620
	Sig. (2-tailed)	.056
	N	10

The last relationship for the verbal short-term memory in the accuracy results was found with the form-unrelated ZUI condition. The relationship between verbal short-term memory and form-unrelated ZUI showed a significant correlation (r= .720, p= .019) (see Table 3.19).

Table 3.19. Pearson correlations between verbal short-term memory and form-unrelated ZUI.

Form-unrelated ZUI	Pearson correlation	.720*
	Sig. (2-tailed)	.019
	N	10

Visuospatial short-term memory

Visuospatial short-term memory had only two relationships with accuracy results in the translation recognition task. The first correlation was found with the form-unrelated condition in GUI, which showed a moderate relationship; however, this relationship was not statistically significant (*r* = .538, *p* = .108) (see Table 3.20).

Table 3.20. Pearson correlations between visuospatial short-term memory and form-unrelated GUI.

Form-unrelated GUI	Pearson correlation	-.538
	Sig. (2-tailed)	.108
	N	10

The second relationship for visuospatial short-term memory in the accuracy results was found with the meaning-related ZUI condition, which again was not statistically significant (*r* = . 563, *p* = .09) (see Table 3.21).

Table 3.21. Pearson correlations between visuospatial short-term memory and meaning-related ZUI.

Meaning-related ZUI	Pearson correlation	.563
	Sig. (2-tailed)	.090
	N	10

Verbal working memory

Verbal working memory showed moderate relationships with the participants' recall scores, reflected in two types of word pairs within two of the teaching methods.

1- The correlation between verbal working memory and the meaning-unrelated condition in GUI showed a trend towards significance (*r* = . 509, *p* = .133) (see Table 3.22).

Table 3.22. Pearson correlations between verbal working memory and meaning-unrelated GUI

Meaning-unrelated GUI	Pearson correlation	.509
	Sig. (2-tailed)	.133
	N	10

The relationship between verbal working memory and meaning-unrelated GUI approaches showed a significant correlation (r= .576, p= .081) (see Table 3.23).

Table 3.23. Pearson correlations between verbal working memory and meaning-unrelated ZUI.

Meaning-unrelated ZUI	Pearson correlation	.576
	Sig. (2-tailed)	.081
	N	10

Visuospatial working memory

The findings also revealed that the fourth memory component, visuospatial working memory, was correlated – but not significantly – with two types of word pairs, both meaning-related. Table 3.24 shows that there was a moderate relationship between the visuospatial working memory and meaning-related conditions of CUI, though not at statistically significant level (*r* = .516, *p* = .09).

Table 3.24. Pearson correlations between visuospatial working memory and meaning-related GUI.

Meaning-related GUI	Pearson correlation	.516
	Sig. (2-tailed)	.127
	N	10

A further correlation was found between the visuospatial working memory and meaning-related MDI, which indicates a stronger relationship; however, this relationship was also not statistically significant (*r* = .617, *p* = .057) (see Table 3.25).

Table 3.25. Pearson correlations between visuospatial working memory and meaning-related MDI.

Meaning-related MDI	Pearson correlation	.617
	Sig. (2-tailed)	.057
	N	10

Accuracy results with the 12 AWMA subtests

A closer examination of the AWMA subtest results reveals more details regarding the relationships between the participants' accuracy in judging the 'Yes' and 'No' word pair conditions and their memory abilities. The next section will present only the statistically significant correlations with each subtest and provide some information about the specific cognitive ability measured by the subtest.

Digit recall and form-unrelated MDI

This task required the participants to access their verbal short-term memory. In this task, the participant heard a sequence of digits and was asked to recall each sequence in the correct order. The findings revealed that the digit recall task results were strongly positively correlated with the form-unrelated condition in MDI (r = . 878, p=0.001) (see Table 3.26).

Table 3.26. Pearson correlations between digit recall and form-unrelated MDI.

Form-unrelated MDI	Pearson correlation	$.878^{**}$
	Sig. (2-tailed)	.001
	N	10

Word recall and meaning-unrelated GUI

The word recall task also measured verbal short-term memory. In this task, the participant heard a sequence of words and was then asked to recall each sequence in the correct order. The correlation analysis results showed a strong relationship with meaning-unrelated GUI (r = . 632, p=0.05) (see Table 3.27).

Table 3.27. Pearson correlations between word recall and meaning-unrelated GUI.

Meaning-unrelated GUI	Pearson correlation	$.632^{*}$
	Sig. (2-tailed)	.050
	N	10

Non-word recall and meaning-unrelated MDI

The third and final subtest of the verbal short-term memory set was the non-word recall task. In this task, the participant heard a sequence of non-words and was asked to recall each sequence in the correct order. Table 3.28 shows that the non-word recall task results were significantly correlated with the meaning-unrelated MDI (r = .640, p = .046).

Table 3.28. Pearson correlations between non-word recall and meaning-unrelated MDI.

Meaning-unrelated MDI	Pearson correlation	.640*
	Sig. (2-tailed)	.046
	N	10

Non-word recall and form-related ZUI

Non-word recall was also found to correlate with the form-related ZUI (r = . 647, p=0.43) (see Table 3.29).

Table 3.29. Pearson correlations between non-word recall and form-related ZUI

Form-related ZUI	Pearson correlation	.647*
	Sig. (2-tailed)	.043
	N	10

Non-word recall and form-unrelated ZUI

Non-word recall was also found to correlate with form-unrelated ZUI (r = .645, p=0.44) (see Table 2.30).

Table 3.30. Pearson correlations between non-word recall and form-unrelated ZUI

Form-unrelated ZUI	Pearson correlation	.645*
	Sig. (2-tailed)	.044
	N	10

Listening recall and meaning-unrelated GUI

This subtest was used to measure verbal working memory. The first task was of listening recall. The participant heard a series of spoken sentences. Then, they were asked to judge whether the sentence was ‘true’ or ‘false’. Finally, they were asked to recall the last words of all the sentences in sequence. The trials of the listening recall task began with one sentence; the number of sentences then increased until the participant was incapable of recalling three correct trials in a row. As shown in Table 3.31, it was found that behaviour in this task correlated significantly with meaning-unrelated GUI (*r* = .635, p = .048).

Table 3.31. Pearson correlations between listening recall and meaning-unrelated GUI.

Meaning-unrelated GUI	Pearson correlation	$.635^{*}$
	Sig. (2-tailed)	.048
	N	10

Listening recall and form-unrelated MDI

The results also revealed that listening recall was significantly correlated with form-unrelated MDI (*r* = .686, p=.029) (see Table 3.32).

Table 3.32. Pearson correlations between listening recall and form unrelated MDI

Form-unrelated MDI	Pearson correlation	$.686^{*}$
	Sig. (2-tailed)	.029
	N	10

Backwards digit recall and meaning-unrelated ZUI

The third subtest was the backwards digit recall task. In this task, the participant heard a sequence of spoken digits and was required to recall them in backwards order. The trials of this test began with two numbers; this was then increased by one number in each block. The test ended when the participant became incapable of recalling four correct trials. Table 3.33 shows that a correlation was found between performance in the backwards digit task and the meaning-unrelated condition in ZUI, indicating a strong and statistically significant relationship (*r* = . 761, *p* = . 011).

Table 3.33. Pearson correlations between backwards digit recall and meaning-unrelated ZUI.

Meaning-unrelated ZUI	Pearson correlation	$.761^{*}$
	Sig. (2-tailed)	.011
	N	10

Dot matrix and meaning-related MDI

The dot matrix task required participants to access their visuospatial short-term memory. The participant was shown the position of a red dot in a series of 4 x 4 matrices for two seconds. They were then asked to remember the position and indicate it by pointing to the squares on the computer screen for the examiner. The number of red dots increased until performance broken down. The findings revealed that performance in the dot matrix task correlated significantly (*r* = .702, *p* = .024) with meaning-related conditions of MDI (see table 3.34).

Table 3.34. Pearson correlations between dot matrix and meaning-related MDI.

Meaning-related MDI	Pearson correlation	.702*
	Sig. (2-tailed)	.024
	N	10

Spatial recall and meaning-related MDI

The third subtest of tests to measure the visuospatial working memory was the spatial span. In this test, the participant was presented with a picture of two arbitrary shapes with a red dot on the shape on the right. The participant was required to identify whether the shape on the left was the same as or opposite to the shape on the right. The shape with the red dot was also rotated. After each trial, the participant was asked to recall the position of the red dots on the shapes, in the correct order. The final correlation was found between the accuracy results in the meaning-related MDI condition and the spatial recall task; the correlation indicated a strong and statistically significant relationship (*r* = .637, *p* = .047) (see Table 3.35).

Table 3.35. Pearson correlations between spatial recall and meaning-related MDI.

Meaning-related MDI	Pearson correlation	.637*
	Sig. (2-tailed)	.047
	N	10

3.5.2.2 Response time

The results for participants' response time (RT) to all five word pair conditions were found to correlate with two of the memory abilities, as follows.

Verbal short-term memory

The findings revealed that verbal short-term memory correlated with RT results in three word pair conditions.

The first correlation was between the verbal short-term memory and the meaning-related condition in GUI, which indicated a moderate relationship; however, this relationship did not reach the conventional significance level (*r* = .552, *p* = .098) (see Table 3.36).

Table 3.36. Pearson correlations between verbal short-term memory and meaning-related GUI.

Meaning related GUI	Pearson correlation	-.552
	Sig. (2-tailed)	.098
	N	10

Table 3.37 shows another relationship between verbal short-term memory and the form-related MDI. The correlation bordered on a statistically significant value (*r* = .615, *p* = .058).

Table 3.37. Pearson correlations between verbal short-term memory and form-related MDI.

Form-related MDI	Pearson correlation	-.615
	Sig. (2-tailed)	.058
	N	10

The results revealed a correlation between verbal short-term memory and form-unrelated MDI. The relationship approached accepted levels of statistical significance (*r* = .616, *p* = .058) (see Table 3.38).

Table 3.38. Pearson correlations between verbal short-term memory and form-unrelated MDI.

Form-unrelated MDI	Pearson correlation	-.616
	Sig. (2-tailed)	.058
	N	10

Visuospatial working memory

The second memory component that showed striking correlations with the RT results was the visuospatial working memory. This included the only significant correlation with the translation equivalent GUI (*r* = .687, *p* = .028) (see Table 3.39).

Table 3.39.. Pearson correlations between visuospatial working memory
and translation equivalent GUI.

Translation equivalent GUI	Pearson correlation	.687*
	Sig. (2-tailed)	.028
	N	10

The findings also revealed that the visuospatial working memory correlated with the translation equivalent MDI. The correlation was trending towards significance (r = .501, p = .14) (see Table 3.40).

Table 3.40. Pearson correlations between visuospatial working memory
and translation equivalent MDI.

Translation equivalent MDI	Pearson correlation	.501
	Sig. (2-tailed)	.140
	N	10

Table 3.41 shows that there was a correlation between the visuospatial working memory and the translation equivalent ZUI. The relationship was statistically significant (r = .792, p = .006).

Table 3.41. Pearson correlations between visuospatial working memory
and translation equivalent ZUI.

Translation equivalent ZUI	Pearson correlation	.792**
	Sig. (2-tailed)	.006
	N	10

Response time results for the 12 AWMA subtests

The findings revealed that response time was significantly correlated with just two AWMA subtests: counting recall and ‘Mr. X’. The counting recall task was the component that most closely correlated with RT results, as it correlated with all four types of word pairs within the teaching methods.

Counting recall and form-related GUI

The second subtest that measured the verbal working memory was the counting recall task. In this task, a visual array of blue triangles and red circles was presented to the participant. The participant was required to give two responses after seeing the presented array. First, they were asked to count the circles in the array, and second, to recall the tallies of circles. This test trial began with one visual array, and continued by adding more visual arrays in each of the subsequent blocks, until the participant was unable to remember four trials correctly.

The correlation between this task and the form-related condition in GUI (*r* = .733, p=0.016) indicated a strong relationship (see Table 3.42).

Table 3.42. Pearson correlations between counting recall and form-related GUI.

Form-related GUI	Pearson correlation	$.733^{*}$
	Sig. (2-tailed)	.016
	N	10

Counting recall and form-unrelated GUI

The correlation analysis results shown in Table 3.43 revealed another relationship between counting recall and form-unrelated GUI (*r* = .680, p=0.031), which was also statistically significant at *p* = .031.

Table 3.43. Pearson correlations between counting recall and form-unrelated GUI.

Form-unrelated GUI	Pearson correlation	$.680^{*}$
	Sig. (2-tailed)	.031
	N	10

Counting recall and meaning-related MDI

There was a strong correlation between counting recall and meaning-related MDI. This relationship was statistically significant (*r* = .800, *p* = . 005) (see Table 3.44).

Table 3.44. Pearson correlations between counting recall and Meaning related MDI.

Meaning-related MDI	Pearson correlation	$.800^{**}$
	Sig. (2-tailed)	.005
	N	10

Counting recall and meaning-unrelated ZUI

The final significant correlation for the counting recall task was with the meaning-unrelated ZUI (*r* = .667, p= .035) (see Table 3.45).

Table 3.45. Pearson correlations between counting recall and meaning-unrelated ZUI.

Meaning-unrelated ZUI	Pearson correlation	$.667^{*}$
	Sig. (2-tailed)	.035
	N	10

Mr. X and translation equivalent GUI

The Mr. X task required participants to access their visuospatial working memory. The participant was shown a picture of two Mr. X figures and asked to judge whether the blue Mr. X was holding the ball in the same hand as the yellow Mr. X. The blue Mr. X was rotated. In addition to this, the participant had to remember the location of the ball in Mr. X's hand in sequence. The correlation between this task and the translation equivalent conditions in GUI (*r* = .873, *p* = .001) indicated a very strong relationship (see Table 3.46).

Table 3.46. Pearson correlations between Mr X and translation equivalent GUI.

Translation equivalent GUI	Pearson correlation	$.873^{**}$
	Sig. (2-tailed)	.001
	N	10

Mr. X and translation equivalent ZUI

It was also found that Mr. X correlated significantly with translation equivalent conditions in ZUI (*r* = .871, *p* = .001) (see Table 3.47).

Table 3.47. Pearson correlations between Mr. X and translation equivalent ZUI.

Translation equivalent ZUI	Pearson correlation	$.871^{**}$
	Sig. (2-tailed)	.001
	N	10

3.5.2.3 Pupillometry

The pupillometry results revealed no significant relationships between the participants' individual differences in short-term and working memory and their pupillary response. That is to say, the pupillary responses of the participants while the English word was presented on the display were not affected by their memory.

However, looking at the correlation between pupil size and the AWMA results, the findings revealed negative correlations between some of the AWMA components and subtests and the size of the participants' pupils, which were measured during the period of presenting the English word in the translation recognition task. During this period, the only AWMA component that correlated with the mean pupil size was the verbal short-term memory, with an average of r = - .28; this was observed in all the conditions of the task. The correlation showed a trend towards significance at p = .43 > .05. This suggests a negative relationship, meaning that participants with a large verbal short-term memory span experienced less pupil dilation while retrieving L2 word meaning.

AWMA subtests

Other moderate negative but not significant correlations were found between pupil size and some of the AWMA subtests, such as non-word recall (r = - .45, p = .18), digit recall (r = - .24, p = .49), listening recall (r = - .55, p = .1), and Mr. X (r = - .22, p = .53). The only positive correlation was with the backwards digit recall task (r = .38, p = .27).

3.5.3 Summary

This second experiment of the main study investigated the participants' behaviours using the EyeLink Data Viewer software. It sought to identify patterns in participants' responses to the translation recognition task using three main measures: judgement accuracy, response time, and pupil size. The results of these measures were investigated separately using three types of teaching method (multimedia representations). In addition, the results of the translation recognition task were linked to participants' memory abilities. In regards to the third research question, the results revealed that participants were significantly faster and more accurate in judging the 'No' translation pair conditions than the 'Yes' conditions. Among the different multimedia representations, participants responded to MDI words faster and more accurately than GUI and ZUI words and, in both cases, the difference was statistically significant ($p < .05$). These rich findings will be discussed in detail in the next section.

To answer research question four, the EyeLink data was also correlated with the participants' AWMA scores. The analysis of the results using Pearson's correlation coefficient revealed statistically significant relationships between the participants' memory abilities and their scores in the translation recognition task. This result revealed that verbal short-term memory had a number of significant correlations with different word pair conditions across the teaching methods in terms of accuracy. The results for verbal short-term memory and visuospatial working memory were found to correlate with the response time results, where participants with high verbal short-term memory abilities were faster and more accurate in judging the translation pairs.

Lastly, the correlation results of the participants' performance during the translation recognition task revealed a relationship between the total amount of time that test-takers took to respond and the counting recall task. While no correlations were found with pupil size, the results also showed that there were significant relationships between participants' accuracy and performance in several AWMA subtests.

To sum up, the results related to the third research question, which drew on the EyeLink data, revealed that all participants were more accurate in and spent less time correctly judging the MDI words. Pupillometry results revealed no differences in pupil size when completing the translation recognition task between the three multimedia representations. Furthermore, the correlation results between the participants' performances during the translation recognition task and their AWMA scores showed a significant relationship, as research question four sought to determine. The results revealed significant correlations between both accuracy and response time and verbal short-term memory. There was also no significant relationship between pupil size while participants were shown the English words and their AWMA scores. There are several possible interpretations of this relationship, which will be discussed in the next section.

3.6 Discussion

The second study revealed several valuable insights into how second language learners access L2 words when they first begin learning novel L2 vocabulary. The study employed three measures during the translation recognition task (accuracy, response time, and pupillometry). The insights gained will be discussed in this section, referring to each research question in turn.

3.6.1 Discussion about research question 3

What is the impact of the three different interactive multimedia representations types (MDI, GUI, and ZUI) on response accuracy, response time, and pupil responses for L2 word retrieval?

The third question from the current research investigates how far the performance of L2 learners differs in terms of lexical access to learned L2 vocabulary with CAVA and different multimedia representations.

Paired sample T-tests were run with three main eye-tracker measures (accuracy, response time, and pupillometry), designed to analyse CAVA-2 data for the third research question. The findings indicated that L2 word learners' performances in the translation recognition task were affected by the teaching method used. In the following subsections, the results of the accuracy, response time, and pupillometry will be discussed before presenting a general examination of the third research question.

CAVA-2 accuracy

In accuracy analyses, the learner's performance in the translation recognition task indicated the effect of exposure to the different e-learning representation methods on ability to access and retrieve target lexis. The learners achieved different vocabulary learning scores depending on the teaching methods used. Given that the EyeLink software was used to reflect response accuracy, MDI learners can be inferred as the most accurate. The tendency that L2 beginner learners are presently shown to have, i.e. to link target L2 words to their translation equivalent, correspond with some of the bilingual mental lexicon models; namely, the Revised Hierarchical Model (Kroll and Stewart, 1994). The accuracy analyses also revealed that the MDI leads to better learning than ZUI. Although ZUI did not present the L1 word, the results suggested that the rich information about ZUI representation, which shows the visuospatial context of the target word, along with its sematic related words, enhanced the lexical access process, despite the fact that the stimuli in the translation recognition task were verbal (written L2 and L1 pairs). This approach to presenting words in the post-test task is most similar to how words were displayed in the MDI teaching phase.

CAVA-2 response time

The impact of the CAVA-2 representations was also apparent in the response time analyses. Time spent making a correct decision reflected the time the learner needed for lexical access processing. Again, the learners' performances varied across the e-learning representation methods. In line with the accuracy of the results, words learned through MDI led participants to be fastest to accept accurate *Yes* word pairs in the translation recognition task.

CAVA-2 pupillometry

In pupillometry analyses, learners' interaction with the translation recognition task confirmed the effect of learning on different e-learning representations. The learners' pupillary responses were different according to the teaching method. Given that the pupillometry measure reflects the task workload (Kahneman, 1973; Laeng, Sirois, & Gredebäck, 2012), accessing words encoded with GUI and MDI can be inferred as more demanding than ZUI. The learners' spatial processing capabilities seemed to take some of the information burden with ZUI. On the other hand, the MDI learners seemed to attempt to read the text and interpret the meaning simultaneously using one channel (the verbal channel), which also seemed to increase the workload. However, if we take advantage of both our verbal and visuospatial processing capabilities to accomplish learning, we might be able to reduce the load on an individual channel: allowing spatial processing capabilities to adopt some of the information burden. Pupil size was greatest for participants retrieving words learned through GUI, suggesting this task was the hardest: presenting word pairs in the translation recognition task verbally (in written forms) was not helpful when accessing a concept as learned through an illustration.

General discussion of research question 3

Through EyeLink measures used to observe participants' reactions to translation recognition tasks, the impact of multimedia representations as teaching methods has been widely understood. It has been shown that MDI learners are the most accurate and fastest to judge correct word pairs. Although the stimuli in the translation recognition task was verbal, presenting written word pairs, MDI was followed by ZUI, which is characterized by a rich visuospatial interface. The spatial processing capabilities of ZUI appeared to reduce the mental load of retrieving words. On the other hand, processing MDI words appeared to overload the verbal memory system, as learners needed to read the text and interpret the meaning using a verbal channel. The contextual differences in the translation recognition task compared to what GUI learners had been taught caused cognitive overload.

In addition to the main effect of multimedia representations, it is of interest to this researcher how early second language learners access L2 words, and how this might differ with changes in teaching methods. As previously mentioned in the literature review, theoretical models distinguish between a shared conceptual store for the two languages, and separate lexical (word form) representations of the bilingual mental lexicon. One of the major aims of this study is to test the predictions of Kroll and Stewart's (1994) Revised Hierarchical Model (RHM) in order to address lexical processing in L2 in CALL environments. The performance of ten native English speakers, who had been exposed to different vocabulary e-learning methods in CAVA-2, was compared in the context of translation recognition. With GUI and MDI teaching methods, there were reliable interference effects from semantic distracter conditions, the one exception being that of ZUI learners who appeared to experience less semantic interference. In other respects, all the different teaching methods' learners used, aside from noticeable differences in the accuracy of their performance, were similar with respect to the effects of L1 lexical form interference.

The next section considers how these findings might constrain future lexical processing models. Figure 3.13 displays a model which proposes an activation-based L2 word retrieval processing framework that guides interpretation of the research results. The model provides an operationalization of word recognition that responds to the translation recognition task. By word recognition, I refer to cognitive processes that involve retrieving different aspects of word knowledge, such as its form and meaning (learned through a translation equivalent (MDI), an illustrated picture (GUI), or a visuospatial overview (ZUI)). If the information about a specific aspect of word knowledge in a memory component reaches a threshold, that knowledge is retrieved, incrementing and facilitating the retrieval of other aspects.

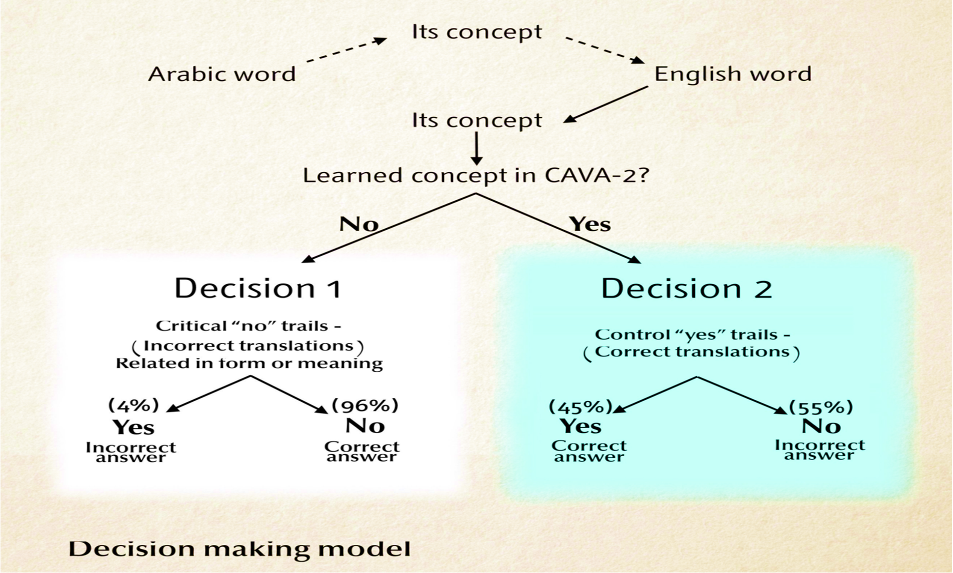


Figure 3.13. An activation-based L2 word retrieval processing framework.

Accuracy- Decision 1

The Arabic word form is most likely absent from the learner's mind, and did not affect this decision. As the participants could have easily rejected form related distracters, it seems like the English form is highly activated. The concept, on the other hand, seems to be active during the decision making process, as it is strongly linked to the English word form, because the meaning related distracters caused a high degree of interference. Zooming in on the meaning related distracters, ZUI prompted less interference, indicating that it was more effective when teaching the meaning of the word than GUI and MDI. GUI and MDI resulted in more semantic interference when their students faced additional difficulties rejecting meaning related answers. This indicates that GUI and MDI did communicate the meanings of Arabic words, but with less efficiency than ZUI. In the case of ZUI, 93% of meaning related distracters were rejected, showing more accuracy in this condition than in GUI or MDI. ZUI learners, however, were less accurate in the other *No* conditions, such as form related, form unrelated, or meaning unrelated. This indicates that meaning relatedness caused less interference for ZUI learners than it did for GUI and MDI learners.

Accuracy- Decision 2

Taking this decision proved more difficult for the participants, as they needed to access the Arabic form. Thus, it was unsurprising that 55% of responses to this condition proved incorrect. Only 45% of the responses provided correct answers. Words that had been learned using MDI, which is a verbal based teaching method showing a list of written Arabic and English words, were classified most accurately at 56%.

As seen in the model, the participant decides in the central stage on whether or not the concept that was extracted from the English word, which appeared in the translation recognition task, was learned and paired with one of the target Arabic words in the CAVA-2 learning phase. If yes, the English word form still proved active and connected to the concept. Therefore, it will be easier to determine if the Arabic word form was learned via MDI. Therefore, it can be argued that MDI could have strengthened the Arabic-English form link more than GUI or ZUI.

Even more striking is that ZUI was the most effective learning environment, as measured by accuracy, among the visual representations in decision 2. It successfully communicated the meaning of the word to the extent that it was close to the impact of MDI (ZUI 48%, MDI 56%). However, the participants took longer to make this decision when the words had been learned through ZUI.

Accuracy- general discussion

The accuracy and response time analyses revealed several important findings with potential implications for bilingual lexicon models and L2 lexical processing.

First, the experimental findings confirmed the predictions of the RHM model. Stronger links of concept to L1 emerged in decision 1. Moreover, the connection between the concept and the L2 word was found to be weaker, as indicated in decision 2. Accessing L2 was found to be better through L1, as proved by the MDI results (in decision 2).

A second finding presenting new constraints to bilingual lexicon models is that meaning-related interference was modulated by the teaching method. Models of word recognition have not, for the most part, provided accounts explaining the rules for how L2 words were learned. Therefore, it is recommended to take into account the engagement of teaching method impact at L2 recognition processing.

As mentioned in the methodology section, there were four critical conditions for each word learned. In the conditions in which distracters were meaning related, the activation of shared semantic codes arguably induced competition among a set of semantic candidates. As related lexical form distracters were unlikely to produce converging meaning activation, the different semantic information provided by learning methods might become available for use as a cue to retrieve and select target L2 words and suppress other competitors. The teaching method delivered different cues concerning word meaning, which were used by the participants to judge the pair as not being translation equivalents. By this explanation, the initial activation of Arabic word form is blind to the teaching method; only later in processing is information about the Arabic word form used as a criterion for making a response in decision 2 (for yes conditions).

It appears that the question should be whether there are different coding elements that characterize the design of each e-learning representation, which might lead to different sensitivity towards meaning during L2 recognition. In the literature regarding vocabulary acquisition, Jiang (2000) suggested a widely accepted explanation, which is that second language learners copy L1 semantic and syntactic information into L2 lexical items (Jiang, 2000). If the L2 word is fossilized onto its L1 equivalent to this extent, and never fully integrated with its specific semantic, syntactic, and morphological specifications, this means that sensitivity to the semantic information will underlie sensitivity to the residual L1 system, which is acting as the L2 system host. The findings of the present study, however, determined that the extent of these semantic information processes are dependent on the teaching method, and the impact of the semantic features used to design each method.

Although the effects from semantic interference differed across teaching methods, meaning related conditions produced more interference effects for all teaching methods, rather than other critical conditions. As the participants in the present study were less proficient learners, it might have been confusing to explain to them the sensitivity towards meaning for them, as previous evidence suggests that conceptual mediation improves only with increasing proficiency in the L2 (e.g. Kroll, Michael, Tokowicz, & Dufour, 2002; Kroll & Stewart, 1994). Some other studies, however, argued against complete conceptual processing at the earliest stages of L2 expertise (e.g. Dufour & Kroll, 1995; Talamas et al., 1999), suggesting a narrower range of semantic interference. However, as the analyses in the current experiment suggest, the strength of the semantic association did affect learners from the outset of their L2 acquisition. This is consistent with what other previous researchers have claimed; that is that L2 learners are sensitive to conceptual information from the early stages of acquisition (Altarriba & Mathis, 1997; De Groot & Poot, 1997). These findings differ across the literature, and can be reconciled in a number of ways. First, the type of measures used in some of the past studies result in different findings. For example, using comprehension rather than production measures showed that accessing meaning can be done directly for L2 words. Another aspect is using a small set of materials repeatedly (e.g. Altarriba & Mathis, 1997). In addition, the type of stimuli used when presenting the L2 word, whether it is a word or a picture (e.g. La Heij, Hooglander, Kerling, & Van Der Velden, 1996). In such studies, how the words are presented in the translation recognition task could affect the results. Displaying an L1 word to the learner might boost their ability to access its shared concepts with the paired L2 word, particularly, and less demandingly, when they were not required to speak in their L2. Moreover, the data showed that all less proficient participants in La Heij et al.'s (1996) experiment, regardless of their teaching methods, revealed inhibitory effects for form-related distracters.

In all, the model shows that the accuracy in taking decision 1 is considered as a concept decision. During decision 1, the word was not as important as the concept of the word. This was indicated when the form interference found to have no effect. ZUI representations was found to be the best to teach the concept. This might be a result of its coding elements of its visuospatial design. ZUI, however, tended to cause larger pupil size during meaning related and translation equivalent more than form related conditions. Moreover, ZUI had the least interference from semantic related distracters comparing with GUI and MDI. The accuracy of decision 2 was affected by the need of accessing and activating the Arabic form. MDI, followed closely by ZUI, made this easier.

Response Time- Decision 1

As mentioned in the methodology section, only correct answers were included as part of the response time. The participants were fastest to take decision 1 in relation to words learned through ZUI (critical *No* conditions) at 819 ms. The participants found ZUI words easier, and rejected incorrect translations more rapidly. The model shows that the learners took decision 1 after comparing the presented concept with previously learned ones. Thus, the concept is active when taking the decision. The dominance of ZUI indicated that the link between the concept and the word form was strengthened more by ZUI than MDI and GUI. On the other hand, MDI (838 ms) was slightly faster, suggesting that the concept-form link is perhaps weakest in GUI (849 ms).

Among the conditions mentioned, there were also differences in the participants' response time (see figure 3.14.). The participants were faster at meeting unrelated meaning conditions than meaning related conditions. As the model suggests, with the concept presented in mind, it might be affected by interference from other similar concepts, which are semantically close. This might be evidence that the concept was similarly activated, regardless of the teaching method. ZUI was the least affected condition by meaning relatedness (896 ms). The interference was only 89 ms. This finding supports the assumption that it better strengthened the form-meaning link than the other teaching methods. MDI had the least form interference during processing decision 1.

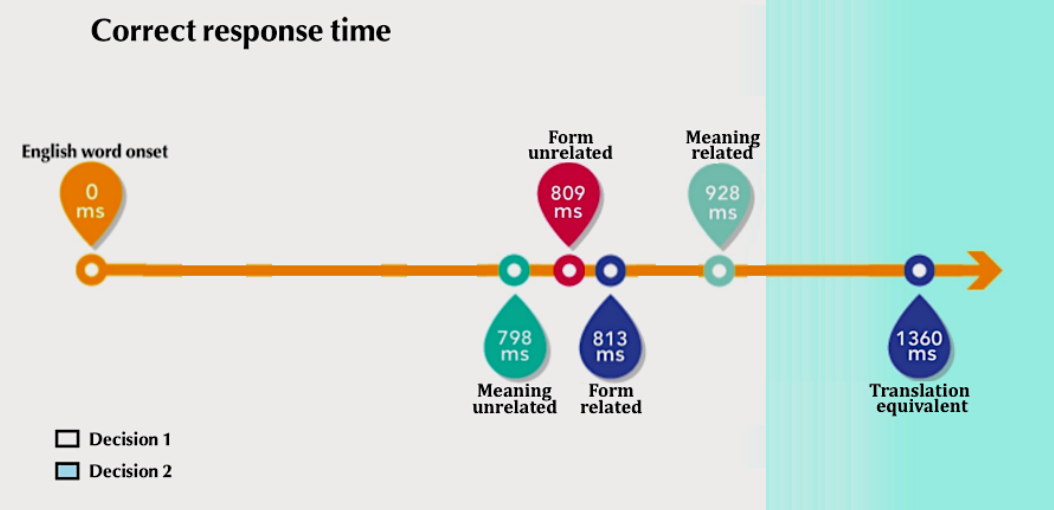


Figure 3.14. Response time across conditions.

Response Time- Decision 2

Compared to ZUI and GUI, MDI was the most influential, as the participants who learned via MDI took less time (1317ms) to say “Yes” that something is the correct translation (Decision 2, figure 3.13.). And this is further evidence that form is more active and present during decision 2, as indicated by MDI results. MDI communicated the meaning of the word through a translation equivalent, and this enhanced the process of learning the word form. ZUI was the most demanding, since it took the participants longer to think.

Response time- general discussion

The pattern of response time results for the control (Yes), as well as, critical (No) translation trials mirrored the accuracy of the findings for all the teaching methods. The learners' performance was impaired more by semantic similarity when taking decision 1 (no conditions), but was later less affected by form similarity. Decision 2, on the other hand, in which participants had to access the learned L2 Arabic word form, was found to be faster via L1. This is reflected in the MDI results. These findings suggest that learners rely heavily on the aspects of knowledge learned about the L2 word, which in turn varied according to the teaching method, even from their first contact with a new L2.

Indeed, the ZUI learners not only exhibited the least effect from semantic interference, but were also faster than the GUI and MDI learners to respond to semantically related pairs, suggesting that the ability to conceptually mediate L2 decreases the processing demand under these circumstances. The general pattern of interference and responses was consistent with a shift in reliance on meaning for D1, to reliance on form during D2, in terms of both accuracy and response time results. MDI learners produced observable semantic effect in response to latency and accuracy with critical *No* conditions, but they were the most accurate and fastest when responding *Yes* to true translation pairs.

The findings of the present study also differ in some respects from the pattern of results reported by Talamas et al. (1999), which included a number of methodological aspects that differ from the current work and so may account for the different results. In Talamas et al.'s (1999) study there was no exposure to any kind of immediate learning materials before the translation recognition task for participants, who were all bilingual to some degree. On the other hand, the findings reported in the present study provide support for the role of conceptual processing during the earliest stages of SLA. However, the current study results revealed that the participants did not spend significantly more time in semantic related conditions, although there was a significant effect from accuracy, indicating that the participants were then more likely to respond with a false positive to semantically related pairs.

Previous research on semantic priming across languages for bilinguals (e.g. Williams, 1994) suggests that the degree of semantic feature overlap between primes and targets provided an important determinant of priming. If second language learners who are not fluent were able to access some semantic information from L2 words (e.g. Dufour & Kroll, 1995), and if the process of concept manipulation might have developed in the early stages of language learning, then the ZUI learners in the present work might have proven less sensitive to semantically related distractors, as they were highly exposed to more information regarding the meaning of the target words.

The discussion of the current study analyses above shows the learners' performance on the translation recognition task was indeed affected by the type of teaching method they were exposed to. The most striking result was for the MDI learners. After the participants had to take decision 2, there was a teaching method effect for the MDI words, which seemed to strengthen the Arabic-English form link, resulting in the fastest correct decisions.

The analysis described provides a number of interesting implications for understanding the changes that occur in L2 recognition processing, which provokes changes in the teaching method. First, it suggests that the overall switch from reliance on meaning to reliance on form while changing the target memory (decision 1 to decision 2) in the second language is influenced by the nature of learned information, not only the change in the proficiency, which has been claimed to affect the learner's accessibility to that information. Second, it is in line with the latest connectionist approaches, which state that featural similarity between words is an important determinant for accessing semantic memory (e.g. McRae & Boisvert, 1998). The fact that the ZUI learners produced less semantic interference, and were quicker to take correct responses in decision 1 and, at the same time, were far more accurate for posing meaning related pairs than other teaching methods, suggests that ZUI learners might be able to access a wider range of semantic information for L2 than the GUI and MDI learners. However, the results revealed that even ZUI learners spent more time making correct answers in decision 2 (yes conditions), achieving a high percentage of accuracy. This establishes that the fundamental mechanism of L2 processing is not the same for all the teaching methods, as it differs in the encoding and retrieval process.

Unlike GUI and ZUI learners, MDI learners had greater access to Arabic word form information, as they could have accessed information that was available from activated Arabic-

English form links, and used it to override form-based competition among lexical candidates. Thus, MDI learners were less likely to produce form interference in the context of translation recognition.

The implication of findings for the interpretation of the current data is that it is likely that meaning-related pairs did produce activation that was potentially distracting for all GUI, MDI, and ZUI learners. However, as some researchers (e.g. De Groot & Comijs, 1995) have claimed, the translation recognition task involves lexical as well as conceptual-level processes. The learners appeared to be able to use higher-level feedback to respond to D1, and then take advantage of the information generated at the level of lexical form when responding to D2. If this interpretation were appropriate, then the switch when retrieving information between lexical and conceptual levels will be an essential feature of bilingual mental lexicon models.

If the GUI and ZUI learners did have direct and strong conceptual representations for Arabic words, then they would not be likely to use semantic information to resolve competition at the lexical level. The selection of a specific lexical candidate for GUI and ZUI learners is thus more demanding, and takes more time, as it goes through the concept to the target Arabic word form, while MDI learners created a direct Arabic-English lexical association, and thus, were faster.

Indeed, this is precisely the aspect of the observations of D2 that are described in the model (figure 3.13), in which contextual clues do not appear to eliminate the form-related errors made by GUI and ZUI learners, compared with MDI learners. Indeed, MDI learners apparently learned the lexical aspect of word knowledge best, from among the teaching methods.

Although the revised hierarchical model can provide explanations for the general pattern of differences among the teaching methods in decision 2, that model could not, in in its present form, accommodate the various effects of different semantic features, as described in the current analysis of decision 1.

The response time to physically respond to the translation recognition task suggests that conceptual activation was present in all trials, lasting for a mean of 835 ms after the onset of the translation, in which the semantically affected decision was taken (decision 1). For the majority of the participants, who are likely to express great certainty about their L1 knowledge, any significant activation of unlearned conceptual features might be expected to present enough evidence to respond correctly regarding whether the pair of words are not translation equivalents. If not, the more similar semantically related pairs will have a high degree of overlap, and produce longer RTs, such as true translation pairs, and participants will be more likely to perform deeper cognitive processing to manipulate the information, taking them about 425 ms longer to reach a correct decision (decision 2) around 1360 ms after the onset of the translation equivalent (English word). By this account, the difference in the response time for the participants when processing L2 words, regardless of the teaching method, is due to the ability to activate the necessary information to make an appropriate decision.

In summary, response time allowed to take decision 1 is a concept based decision, with ZUI proving the easiest (shortest time to decide, less affected by meaning relatedness). Meaning relatedness caused interference for all methods. The response time to take decision 2 seemed to be affected by level of knowledge of the Arabic word form. The most striking finding was that MDI learners were the fastest.

Pupillometry- general discussion

Although pupillometry contributed important gains to psychology, offering a window through which to examine the intensity of mental processing (Laeng et al., 2012), and the processing of foreign languages (Hyönä et al., 1995; Duñabeitia & Costa, 2015), the use of pupillometry to investigate the impact of multimedia has been scarce. Pupillometry is a method that can be employed to provide sensitive indices of cognitive effort (Beatty, 1982) as a means to track linguistic processing, not only when responding to external stimuli but also to reflect internal stimuli such as emotions and thoughts (Goldwater, 1972; Loewenfeld, 1993). Mental effort while retrieving L2 words as a linguistic process was assessed in the present study. The processing requirement for this kind of internal thoughts, in order to judge the word pair in the translation recognition task, was found to be impacted by the learning method employed in the present study. Since many previous studies (e.g. Kahneman 1973; Laeng et al., 2012) revealed a high correlation between pupillary dilation and cognitive load, pupillometry in the current study provided findings that might suggest a greater processing demand for MDI and GUI than ZUI words. In the absence of the use of pupillometry in multimedia research, it is clear that the influence of using multiple representations on L2 learners' performance in the current study might only be moderately linked to the available body of research. However, the pupil dilation evoked by the different teaching methods under certain conditions during L2 word accessing can be explained in relation to other broader pupillometry studies.

The larger pupil size caused by meaning related conditions to ZUI learners, as revealed in the present study, could arise as a result of processing demand. The increase in cognitive demand in this case might result from the competition between activated semantic related words triggered by meaning related conditions. This interference here is likely the reason for the work demand, which is reflected in pupillary responses. These results are in accordance with previous studies that observed the connection between processing demand and multiple probable interpretations in lexical (Ben-Nun, 1986) or syntactic ambiguity (Schluroff, Zimmermann, Freeman, Hofmeister, Lorscheid, & Weber, 1986).

The current experiment's findings pertaining to ZUI in meaning related conditions, as well as the findings of MDI in the form related conditions, echo Schmidtke's (2014) study, which revealed an effect from neighbourhood density; highlighting the number of competing words when retrieving spoken L2 words. The results of the present experiment with regard to the translation equivalent conditions (decision 2) could be linked to studies that found correlations between the recognition of old words and pupil dilation. In this case, the explanation for pupil size increase, when exposed to correct answers does not result from cognitive demand, but arises due to the strength of memory. Fernández et al. (2016) for instance, found increased pupil size when reading proverbs over regular sentences. This result, along with others (Goldinger, 1998; Papesh & Goldinger, 2012) that revealed greater pupil dilation when processing low-frequency words, were explained by word recognition. These researchers suggested that the cause for pupillary dilation during word processing is not limited to cognitive load.

Another explanation of pupil dilation during translation pairs could be the result of a participant accessing the Arabic word form in the target foreign language. As explained above in figure 3.13., the participants in the present study appeared to be taking sequential decisions. In decision 2, they judge the correct translation recognition pairs. To do this, they need to access L2 words. The result of the current study showed that this condition was more demanding than the other critical conditions in all teaching representation methods. This can provide evidence that supports the idea that second language processing evokes larger pupil sizes than initial language processing (Duñabeitia & Costa, 2015; Hyönä et al., 1995).

Unrelated GUI form conditions can be explained as the most difficult for learners, and the difficulties of this task are in line with some previous studies, which found larger pupil dilation with anomalous and negative sentences (Ahern, 1978; Beatty, 1982; Schluroff, 1982) reading complex sentences (Just & Carpenter, 1993), and listening to them (Stanners, Headley, & Clark, 1972).

3.6.2 Discussion about research question 4

What kind of performances (accuracy, response time, pupil size) can be observed across the three multimedia representations (MDI, GUI, ZUI), as indicated by the EyeLink Data Viewer? Is there a correlation between these performances and participants' individual differences in memory?

The fourth question posed in the current research investigates to what extent individual differences in cognitive abilities, specifically working memory and short-term memory spans, influence foreign language vocabulary acquisition. Pearson correlation coefficients were conducted to identify any correlations between the learners' scores in CAVA-2's translation recognition tasks across all three e-learning representations, and the participants' cognitive abilities assessed by the AWMA. In support of the division of the multicomponent model of working memory (Baddeley, 2012), the results show significant correlations between the AWMA components and the subtests with the scores for the three EyeLink measures (accuracy, response time and pupillometry) across the CAVA-2 representations. In the following subsections, the correlation results for accuracy, response time and pupillometry will be discussed respectively before a general discussion of the fourth research question is presented.

AWMA correlations with CAVA-2 accuracy

According to the analyses, the use of different e-learning representations to teach target Arabic L2 words, and use related distractors in meaning and form, highlighted the role of individual differences in memory. Learners with different memory abilities performed differently when completing the CAVA-2 accuracy task. Since the Pearson correlation is used to reflect the nature of such a relationship, highly verbal learners can be shown to be more accurate than other learners. Specifically, the ability to make accurate judgements when rejecting words unrelated in meaning and form (decision 1) was significantly correlated with short-term verbal ability. The outstanding performance of learners with a highly verbal short-term memory is presently shown to acquire L2 vocabulary which corresponds with the effectiveness of foreign language learning in the literature.

AWMA correlations with CAVA-2 response time

The AWMA creates significant correlations with the CAVA-2 representations, and these have been found in response time analyses. In addition to being found to be more accurate when taking decision 1, according to accuracy analyses, highly verbal short-term memory MDI learners also required less time to correctly judge *Yes* word pairs (decision 1), compared to ZUI and GUI learners. In addition, high visuospatial working memory learners spent more time processing ZUI and GUI words before offering correct answers in decision 2. However, learners with a higher working memory found that ZUI required more time to take decision 2.

AWMA correlations with CAVA-2 pupillometry

In pupillometry analyses, the learners' AWMA scores were found to correlate with the CAVA-2 representations during the translation recognition task. These findings confirmed the effect of individual differences in memory on L2 vocabulary learning. The learners' pupillary responses differed according to their abilities: regardless of teaching method. Highly verbal short-term memory learners showed relatively less pupil dilation than less verbal learners. They had a higher mental load, and found the task harder. However, the difference in AWMA correlations with CAVA in its two versions could have resulted from the verbal nature of the translation recognition task used in CAVA-2. The task might well have activated the verbal memory more than other memory components, probably generating the current results.

General discussion of research question 4 (accuracy and response time)

The results of the present study revealed that the three methods of learning resulted in modulating the cognitive processes during L2 word retrieving. This can be seen in the amount of demand that the translation recognition task placed on each type of memory components. Linking this demand to the memory component's capacity clarified that each individual was essential to enhance our understanding of L2 word processing.

Participants' individual differences in the area of memory were shown in every Eyelink measure (accuracy, response time, and pupillometry). The results of these measures correlated significantly with some of the AWMA components. The majority of these findings cohere: highly verbal short-term memory learners were more accurate, faster, and experienced less cognitive load.

The current findings are also in agreement with Service (1992), who concluded that high verbal working memory individuals with a higher ability to represent unfamiliar phonological material in the phonological loop probably acquire new foreign language vocabulary more readily. Although the phonological loop is modular in nature (Baddeley, 2012), with only a limited capacity (thought to be adversely affected by the displacement of information or overwriting), it is distinguished by its capacity to facilitate the learning process under the menu driven user interface (MDI) condition. To be effective, it requires that written materials, such as letters need to be subvocalized, with the reader verbally rehearsing the items, so that they can be fed into the phonological store.

The findings of the present study agree with many previous research studies, which found that verbal short-term memory is involved in the text comprehension process (e.g. Baddeley, Eldridge, & Lewis, 1981; Just & Carpenter, 1992). It has also been found that verbal short-term memory plays an essential role in foreign language acquisition in formal settings (e.g. Dufva & Voeten, 1999; Palladino & Cornoldi, 2004; Service, 1992; Speciale, Ellis & Bywater, 2004).

The current results further support the conclusions of previous vocabulary acquisition experiments in studies in laboratory settings (e.g. Atkins & Baddeley, 1998; Baddeley, Papagno, & Vallar, 1988; Gathercole & Baddeley, 1990; Papagno & Vallar, 1992), clarifying that the phonological loop is a language acquisition device. Whereas, verbal short-term memory has been specifically associated with vocabulary acquisition, working memory has been more generally linked to learning difficulties (Alloway, Gathercole, Willis, & Adams, 2004).

General discussion of research question 4 (pupillometry)

An important reason to investigate pupillometry as a physiological measure is that brain activity and cognitive processing do not map traditional behavioural measures of performance such as accuracy or response time directly. The claim made is that for the translation recognition task demand, there is a positive relationship between cognitive workload in a given task and memory capacity. Thus, the magnitude of observed pupillary responses should be related to L2 word processing, as influenced by the teaching method. The aim of pupillometry as a physiological measure is to propose an explicit mapping between the functions of memory as a cognitive system and multimedia representations during L2 word recognition. Word recognition is the cognitive processes that results when recalling the acquired aspects of word knowledge, such as the word meaning. In the current study, word meaning has been communicated through a translation equivalent (MDI), an illustrated picture (GUI), a visuospatial overview (ZUI), and the word form. These processes will be compared with others that might vary according to the teachings afforded by multimedia representation. Thus, this experiment was conducted to reveal similarities and differences in L2 word processing across the three designs of instructional multimedia representations, and to relate this to learners' individual differences in memory. To analyse the effect of each AWMA component demanded, a translation recognition task was conducted and manipulated as a window on task demand. To analyse the effect of the AWMA component's capacity, variations in individual differences in memory abilities among participants were considered. Then, the study examined the AWMA component demand and capacity, as it is related to pupil responses, which is known as an indicia of neural processing.

Systematic individual differences, in addition to task demand, have been shown by the pupil responses in the current study. This is in line with some previous studies. For instance, Ahern and Beatty (1979) had distributed their participants into two college student groups according to their verbal and SAT scores. The group with the average scores of 877 showed greater pupil dilation than the group with average scores of 1407 during mental multiplication. These results indicated that pupil size is sensitive to individual differences among the groups themselves, as well as to the task demand itself (Ahern & Beatty, 1979). Many other previous studies have clearly revealed a correlation between executive load, or working memory load and pupil size (e.g. Ahern & Beatty, 1979; Chatham, Frank, & Munakata, 2009; Hyönä, Tommola, & Alaja, 1995; Karatekin, Couperous & Marcus, 2004; Nuthmann & Van der Meer, 2005; Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004; Võ et al., 2008). It became apparent that pupillary responses reflected a response to other cognitive processing demands. For example, Kahneman and Beatty (1966) found a positive correlation between pupil size and the cognitive load of memory when increasing the number of digits in a string to be recalled.

One possible explanation as to why ZUI representations reduced pupil size is that the amount of demand that the translation recognition task imposes on the learner's memory can vary according to the number of activated memory subcomponents required to do the task, over a given time interval. Retrieving ZUI words in the present study seems to employ verbal as well as visuospatial components, and that might have helped to distribute task demand over these resource pools, resulting in a lesser cognitive workload overall than that required for GUI or MDI, as shown by pupillary responses. Furthermore, the size of the pupil during the task is found to vary among the study participants (due to memory abilities), and to show individual differences in cognitive performance. Participants with a high verbal short term memory experienced smaller pupil size. On the other hand, MDI word processing is postulated to draw on fewer components for encoding processing and storage than GUI or ZUI words do. Depending only on the verbal component, as in MDI, predicts the necessity for more brain activity in relevant areas of language processing, such as Werneck's area, during the task. Whereas, if the effort is distributed into verbal and spatial pools, which have been shown to be separated according to studies of cognitive abilities (Carroll, 1993; Shah & Miyake, 1996), I assume that the overall cognitive workload will be less. However, I further assume that different tasks can draw on each of the resource pools to varying degrees.

It has also been assumed that retrieving L2 words to complete the translation recognition task will draw on each of the memory components to differing degrees according to the teaching method. For example, MDI word retrieval might primarily be supported by verbal components, whereas GUI and ZUI would be expected to draw primarily on visuospatial components. Nevertheless, studies of cognitive abilities suggest some visuospatial tasks might involve verbal strategies (Reid & Nygren, 1988) and verbal comprehension might sometimes invoke spatial mental model construction (Friedman & Miyake, 2000). Thus, the three multimedia representations of the study are likely to draw on multiple memory components to a different extent.

The results supported the prediction that the amount of workload should be greater in the verbal memory components in the MDI condition than in the GUI and ZUI conditions, because the demand on the verbal channel is greater when both word form and its meaning are required. Pupillary responses are correlated to the presence of one channel demand that is imposed in the context of the MDI word recognition task, which is intended to associate the L1 and L2 word forms. On the other hand, the increase in pupil size during ZUI semantic related conditions could be due to an increase in the number of activated competitors and increased activity in the semantic set of the target word. Task demand can cause changes in pupil size, not only during language processing, but also while completing spatial tasks, and this suggests that the pupil size measure is an index of capacity utilization (Just & Carpenter, 1995). In their study, participants with high and low ability in spatial memory were exposed to a spatial task, where they judged whether a pair of irregular two-dimensional hexagons showed the same figure or mirror-image figures. The task was designed to increase capacity utilization with the degree of variation of the angular disparity of the hexagons. This increase in capacity utilization led to greater pupil dilation for all participants, with significant changes for individuals with lower spatial memory. This information processing, which takes place while reading sentences, is facilitated when contextual hints are provided (Esterman & Yantis, 2009). Contextual hints were also found to enhance working memory (Gazzaley, 2012).

3.6.3 Summary

Pupil size changes for low verbal participants show a trend towards being greater, but this did not achieve statistical significance for high verbal participants, indicating that the cognitive load for the same task was greater for individuals with less verbal ability. In all teaching methods, taking accurate judgements in unrelated conditions in meaning and form (decision 1) were significantly correlated with verbal short-term ability. Among them, high verbal short-term memory MDI learners gave correct answers more rapidly in decision 1. High verbal working memory ZUI learners took longer to make correct decisions in decision 2. High Mister X GUI and ZUI learners required more time to give correct decision 2. This means that the MDI learners did not need to spend a lot of time on decision 2 whether or not they had a high or low ability in verbal working memory. However, high verbal short-term learners showed a trend towards having smaller pupil sizes, although this did not reach statistical significance.

CHAPTER FOUR: CONCLUSION

Multimedia vocabulary CALL has always been used as an environment that allows software developers, educators, and teachers to communicate aspects of word knowledge via e-learning representations, employing various design elements. Each of these aspects of word knowledge represents a specific part of the word. In theory, a multimedia vocabulary CALL environment should be representative of the real life situations of each specific target concrete word, so the knowledge communicated through that e-learning module would then be rich and effective. When reviewing cutting-edge technology, one can see that it is almost possible to represent the majority of the situations in target language use domains by which the language learner will have the opportunity to interact virtually with a simulated real life situation. Presenting such situations, in which the target language takes place, might render the learner more able to see and learn aspects of the target word knowledge. Furthermore, the use and design of such advanced educational technologies should always be compatible with the learner's cognitive abilities, and render them flexible to individual differences among users. CALL makers rarely base their technology on models of memory that limit the effectiveness of multimedia vocabulary CALL applications. Relating multimedia CALL design in teaching vocabulary to SAL and memory research has been encouraged by some researchers (e.g. Handedly, 2014; Pederson, 1987), but few convincing empirical attempts have been made to support that call.

In essence, using simple design elements such as text and uncontextualized pictures to communicate word meaning does not facilitate L2 learners to take advantage of the different elements that can present other features to characterize target objects in real life situations. These visuospatial design elements can also stimulate and employ different components of the learners' cognitive abilities, such as spatial memory, which would normally allow the learner to deeply process spatial information along with other information associated with word meaning. They could also be used to communicate the word associations related to the semantic set as an extra dimension of word knowledge.

Vocabulary learning for L2 learners should not normally be limited to their ability to know the word form and part of its meaning. Including simple textual and visual aspects of the word would most likely be suitable. However, this could also perhaps be explained as a limited drawing of CALL research to SLA and memory theory; one that would threaten the success of CALL interventions, and hinder its potential contributions to furthering our understanding of the complex process of language acquisition. Furthermore, because word knowledge is not limited to form-meaning mapping, the ZUI visuospatial word associations involved different types of spatial design elements. It would be threatening to the learning effectiveness of the word if an attempt were made to exclude all these aspects from the learner. This would be a case of preventing their spatial abilities during the learning process. Consequently, the main purpose of the present study was to investigate the effects when implementing such visuospatial components in a CALL environment: showing the semantic and positional relationship between the target words, on the performance of the L2 learners in multimedia vocabulary CALL environments. The motivation behind this study was a form of response to support the notion of relating the CALL design to SAL and memory theory, to better inform the design of future CALL software.

In order to achieve the purpose of this research, two CAVA programs were designed and used to conduct the two main studies that comprise the present thesis. This research design required the collection of four sets of quantitative data to answer the four main research questions: a) CAVA-1 immediate and delayed post-tests data, b) AWMA correlation data with CAVA-1, c) CAVA-2 EyeLink measures of the translation recognition task- accuracy, response time, and pupillometry, and d) AWMA correlation data with CAVA-2.

The data collected was analysed to answer the four main research questions. To address the first question (What is the effect of three different interactive representation types; namely verbal-based (L2-L1: MDI), visual-based (L2-Picture: GUI) and spatial-based (L2-Context: ZUI) on second language vocabulary acquisition?), a within-subjects design was used to investigate the difference in participants' performance (*n* = 50). Each participant was exposed to a total of six lessons, composed of a total of 24 words. A one way within-subjects -repeated measures- ANOVA was conducted to explore if there was any difference in participants' performance in the three different multimedia representations. The results of this analysis revealed the ZUI learners showed a trend towards higher scores in the CAVA-1 immediate post-test than GUI and MDI, this did not attain statistical significance. However, in the delayed post-test, ZUI learners experienced more difficulties retrieving L2 words than the GUI learners and MDI learners did. In this condition GUI was shown to be the most effective approach. This result implies that the use of spatial design elements in ZUI might positively affect the short-term performance of learners compared to traditional approaches to using L2-L1 lists or L2-pictures in vocabulary multimedia learning. Although, the use of spatial elements in vocabulary educational software is scarce, which makes it difficult to compare these results with the previous body of CALL literature, these results concur with the earlier finding that spatial memory decays rapidly (Hole, 1996; Fagan et al., 2013). Moreover, the long-term impact of GUI compared with MDI is in line with Dual Coding research by Paivio (1991), who found that encoding information through verbal and visual channels leads to better retention than employing just one modality of information.

The second research question asked (What is the effect of individual differences in verbal, visuospatial working memory and short-term memory capacity on the initial uptake of a brand new L2 vocabulary?). Pearson correlations were used to explore the relationship between the participant's memory abilities, as assessed by AWMA, and their vocabulary learning based on CAVA-1 representations. Each participant was examined with a total of twelve subtests, designed to measure their memory capacity according to four main components (verbal short-term memory, visuospatial sort-term memory, verbal working memory, and visuospatial working memory). The results confirmed ZUI short-term impact as significantly correlated with most of the subtests. GUI and MDI, on the other hand, were significantly correlated with working memory abilities, and not with short-term memory components.

In response to the third research question (What kind of performance differences (in accuracy, response time, pupil size) can be observed during lexical access across five conditions (form and meaning relatedness) as indicated by the EyeLink Data Viewer? a relationship was sought between these performances and CAVA-2 multimedia representations (MDI, GUI, ZUI)?). In addition, paired sample *t*-tests were performed with three EyeLink measures (i.e. accuracy, response time, and pupillometry) for three of the CAVA-2 multimedia representations. The results of this analysis indicated that the MDI learners were significantly more accurate and faster. In addition, all the learners proved to be faster and more accurate at judging the *No* translation pairs conditions than the *Yes* conditions, and the difference that arose was statistically significant ($p < .05$).

To answer the fourth question in this research (What kind of performances (accuracy, response time, pupil size) can be observed during lexical access across five conditions (form and meaning relatedness) as indicated by the EyeLink Data Viewer? Is there a correlation between these performances and the participants' individual differences in memory?), Pearson correlation coefficients were conducted, revealing several statistically significant relationships between the participants' memory abilities and their responses as set out in the three EyeLink measures. Compared with the other AWMA components, verbal short-term memory was found to be correlated significantly with the learners' accuracy in word-pair conditions across teaching methods. The results imply that high verbal short-term memory learners were more accurate and faster at judging the translation pairs. There was also a relationship between response time and the counting recall task. In addition, the results showed a very slight trend towards significance relative to negative correlations for verbal short-term memory capacity with mean pupil size (r = - .28, p = .43) in all the conditions of the task.

These data sets were comprised of four main types of results: First, learners' performance in the three representational types of CAVA-1 (i.e. MDI, GUI, and ZUI), which revealed the participants' vocabulary recall scores in the ZUI trend toward significance being higher than GUI and MDI. Other differences in learners' performance were detected in the delayed post-test (i.e. GUI effects lasted longer than MDI and ZUI which experienced faster decay). Second, correlations between the learners' cognitive abilities measured by AWMA with CAVA-1 immediate and delayed post-tests scores, where they showed the ZUI representation to be more highly correlated than MDI and GUI, with all of the AWMA components in the short-term. Third, learners' performance in the translation recognition task after exposure to the three representational types of CAVA-2 (i.e. MDI, GUI, and ZUI), which were assessed by the three Eyelink measures (accuracy, response time and pupillometry), proved that MDI had strengthened the form-meaning link most, followed closely by ZUI and then GUI. Lastly, the learners' AWMA correlations with CAVA-2, which generally revealed that verbal high short-term memory learners' were more accurate, faster and experienced less pupil dilation.

Comparing the results of the two main studies undertaken for the current thesis, it is apparent that individual differences in working memory played an essential role in learners' different outcomes. However, with a small sample size in the second study and the variable nature of the posttests used in the two studies, caution must be applied when comparing the impact of the three multimedia representations used in CAVA (MDI, GUI, and ZUI). Meanwhile the posttests in the first study were presented with the same interfaces, and the posttest in the second study was a translation recognition task, which was verbal in nature. Taking the type of test into account when comparing the results of the two studies, it emerged that applying different tests affected both the correlations with AWMA and also CAVA outcomes. For example, in the first study, MDI came last in both the immediate and delayed posttests. Conversely, in the second study, MDI learners were found to be the fastest and most accurate in completion of the translation recognition task. However, in both the first and second study, ZUI learners achieved relatively high scores.

Taken together, these results have implications for the design of multimedia vocabulary CALL interventions, as will be discussed in the following sections of this chapter. Most importantly, the inclusion of visuospatial components should be considered in a vocabulary multimedia CALL design in order to reflect the different aspects of word knowledge. The implications that resulted from the two main studies of the thesis, as well as its limitations, are presented in the next sections, followed by a discussion of suggested guidelines for future vocabulary CALL research.

4.1 Implications of the study

4.1.1 Implication 1: Spatial representations

The findings of the first and second study have some implications for the design of multimedia vocabulary CALL software. The repeated ANOVA measures outcomes in research question one, when comparing the immediate post-test scores for the three types of representations, as the learners' performance was higher with ZUI than GUI and MDI. This implies that the ZUI representation with the spatial elements might have improved learners' acquisition of the target L2 concrete words. Therefore, this research suggests that designers should take spatial elements into consideration when they attempt to develop multimedia vocabulary CALL software. The results of this research indicate the importance of including visuospatial components to teach different aspects of the word knowledge beyond mapping form-meaning. The construct of L2 multimedia learning programs should take advantage of the learner's spatial memory abilities, to better improve the learning and processing of L2 words.

4.1.2 Implication 2: Theoretical grounded design

The use of Baddeley's (2012) most recent multicomponent model of working memory as a memory theory, as well as using the more specific vocabulary learning model, that is Kroll and Stewart's Revised Hierarchical Model (1994), as underpinning the theoretical ground for designing CAVA, is an attempt to support the call to link vocabulary CALL design to SLA and Memory theory. The complementary nature of this research was shown in the two studies, offering a bigger and more comprehensive picture of the L2 vocabulary learning process in CALL environments. This was achieved by shedding light on both the outcome from the CAVA post-tests, and the L2 word recognition process that underlies that outcome. Relating the design of CAVA and the explanation of the findings to memory theory and research will hopefully assist with the goal of improving vocabulary CALL design.

4.1.3 Implication 3: Individual differences

Conducting the long version of the AWMA, with its twelve subtests provided a reliable and valid advanced comprehensive test to systematically investigate learners' individual differences in working memory and short-term memory. Considering the learners' individual differences as shown in AWMA, in addition to the applying of CAVA and pupillometry, served to shape the current research and the comprehensive picture of vocabulary learning in multimedia CALL environments. This is another essential implication of the current study and its complementary nature, which assisted in furthering our understanding of SLA.

4.2. Limitations of the study

4.2.1 Limitation 1: Verbal task

In the second study undertaken in this research, the use of verbal stimuli in the translation recognition task represented a change from the context in which the ZUI and GUI words were being learned. In contrast, MDI learners might well have been aware of the same verbal cues that were encoded along with the target memory (L2 word), and then those cues were helped to retrieve it. Generally speaking, context-dependent memory is encoded with the spatiotemporal, physiological, mood, or cognitive context cues, and it is slower and more difficult to retrieve the target memory when those cues are absent. Future research would ideally ask: if the purpose of the translation recognition task is to assess learners' ability to access the target L2 word, as communicated via visuospatial representations, then would excluding these visuospatial elements from the post-test present a serious threat to the validity of the task construct?

4.2.2 Limitation 2: Correct responses

Looking at the pupillary responses during trials where participants made correct judgements was important for comparison with the response time results. Comparisons between time, accuracy and pupil size were made in some previous studies (e.g. MacDonald, Just, & Carpenter, 1992; Miyake, Just & Carpenter, 1994), as it was found that comprehension accuracy and reading time for ambiguous structures served as an indication of the greater demand for lexical ambiguity. The relationship between pupillary responses and reading accuracy, and time devoted to ambiguous language structures, suggests that solving this ambiguity resulted from a resource-dependent process, and that the extra demand was reflected in pupil response. Moreover, it was observed that complex sentences took longer to process than simple sentences. The pupillary responses were also detected, creating greater changes in pupil diameter (0.249 mm) as elicited by complex sentences when compared with smaller changes (0.203 mm) in the pupil diameter upon reading simpler sentences (Just & Carpenter, 1993).

4.2.3 Limitation 3: Generalizability

Generalizability was a potential limitation of the second study in the present research. The sample was small (i.e. 10 participants) so that the findings could be generalized to a wider population of L2 word learners. Although the number of learned words was large (108), and delivered divided into twelve topics, a higher number of participants was expected to increase the validity of the scores.

4.3 Recommendations for future research

This research with its outcomes, limitations, and implications posed some suggestions for future vocabulary CALL research. These suggestions were derived from all of the studies in this research. Certain suggestions were presented regarding the use of spatial design elements in CAVA, and the effect of this on long-term performance. Further suggestions focused on employing other working memory components, such as the episodic buffer, and also on the use of translation recognition tasks and picture naming.

4.3.1 Recommendation 1: Better employing spatial design elements

It was shown in the first study posed in this research that ZUI spatial features, which present target words in positions that allow the learner to infer extra information about the word, were highly correlated with memory components leading to greater achievement in terms of learners' vocabulary recall than the other representations developed on a short-term basis. The advantage of short term recall from ZUI in the current research related to the process of learning with no rehearsal. A richer set of spatial information can be practiced and learned, and could also be transformed into improved longer lasting word knowledge. Investigating the role of the spatial design elements in consolidation of the L2 word memory trace could also prove to be a successful idea linking vocabulary CALL design to memory research.

4.3.2 Recommendation 2: Testing the central executive and the episodic buffer

It is worthwhile exploring the role of the central executive and the episodic buffer in empirical studies using reliable assessment tasks. Testing the relationship between each of these components with multimedia representations might help to further our understanding of SAL. Although some studies have measured the phonological loop, visuospatial sketchpad, and central executive (Schüler et. al, 2011), to date, no study has yet investigated the contribution of the episodic buffer in multimedia learning.

4.3.3 Recommendation 3: Using picture naming task in EyeLink

To eliminate the effect of decontextualization, which happened as a result of using the translation recognition task in the second study in this research, the post-tests were expected to use the same representations that the participants had learned with. The translation recognition task constrained the retrieval process by presenting verbal stimuli only. This limits the value of the second study as a means to test the RHM model, as there was limited opportunities to compare between the CAVA-1 post-test findings in the first study, and the CAVA-2 findings in the second study.

4.4 Summary

The present thesis developed and assessed an advanced technological implementation software, to further clarify the vocabulary learning processes in multimedia CALL environments. Two experimental studies were conducted using different methodological measurements; i.e. CAVA, AWMA, and pupillometry, allowing for a better and deeper understanding of how L2 learners interact with multimedia CALL representations. The recruitment of learners with no knowledge of the target language minimized the effects of previous knowledge, and allowed tracking of the learning process in its earliest stages. In addition to the use of the long version of AWMA as a comprehensive and reliable memory test, it is thought to be the first research to employ pupillometry to investigate the efficacy of L2 multimedia learning. This was achieved by comparing and including spatial design elements with more traditional visual types of design methods, such as pictures and texts; highlighting second language learners' acquisition of novel L2 concrete words. The results demonstrated how the L2 learners interacted with, and perceived the three types of CAVA multimedia representations differently. Differences were evident in the learners' vocabulary recall scores in the post-tests, as well as in the correlations of those with their AWMA results. It is assumed that the findings of the present research, despite its limitations, had contributed to furthering our understanding of multimedia CALL and SLA.

REFERENCES

Absatova, K. (2015). Applying Automated Working Memory Assessment and Working Memory Rating Scale in Russian Population. *The Russian Journal of Cognitive Science, 3*(1-2), 21-33.

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology*, *40*(1), 23-31.

Ahern, S. K. (1978). *Activation and intelligence: Pupillometric correlates of individual differences in cognitive abilities.* Unpublished doctoral dissertation, University of California, Los Angeles.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*(4412), 1289-1292.

Aitchison, J. (1990) *Words In the Mind*. Oxford: Basil Blackwell.

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), 499-517.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology, 5*(1), 202-232.

Alloway, T. P. (2007). *Automated Working Memory Assessment*. London: Pearson Assessment.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child development*, *77*(6), 1698-1716.

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of experimental child psychology*, *87*(2), 85-106.

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of memory and language*, *36*(4), 550-568.

Andrade, J. (Ed.). (2001). *Working memory in perspective*. Hove, UK: Psychology Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes1. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Atkins, P. W., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19*(4), 537-552.

Austin, K. A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers & Education, 53*(4), 1339–1354.

Baddeley, A. (1986). *Working Memory*. New York: Oxford University Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*(11), 417-423.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology, 63*, 1-29.

Baddeley, A. & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *The Quarterly Journal of Experimental Psychology, 33*(4), 439-454.

Baddeley A. D., Eysenck, M. W., & Anderson, M. C. (2015). *Memory*, 2nd Edition. Psychology Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, *105*(1), 158-173.

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of memory and language, 27*(5), 586-595.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883-2901.

Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge: Cambridge University.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin, 91*(2), 276-292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of psychophysiology, 2*, 142-162.

Bederson, B. B. (2011). The promise of zoomable user interfaces. *Behaviour & Information Technology*, *30*(6), 853-866.

Ben-Nun, Y. (1986). The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing. *Brain and Language, 28*(1), 1-11.

Bergman, R. E., & Moore, T. V. (1990). *Managing interactive video/multimedia projects*. Englewood Cliffs, NJ: Educational Technology Publications.

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.

Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning, 30*(7), 709–725.

Bolt, R. A. (1979). *Spatial Data Management.* Architecture Machine Group, MIT, Cambridge, Massachusetts.

Branch, R. M. (2009). *Instructional design: The ADDIE approach* (Vol. 722). Springer Science & Business Media.

Briesemeister, B. B., Hofmann, M. J., Tamm, S., Kuchinke, L., Braun, M., & Jacobs, A. M. (2009). The pseudohomophone effect: evidence for an orthography–phonology-conflict. *Neuroscience Letters, 455*(2), 124-128.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, *22*(5), 349-368.

Brouwer, H., Fitz, H., & Hoeks, J. C. (2010). Modeling the noun phrase versus sentence coordination ambiguity in Dutch: evidence from surprisal theory. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics* (pp. 72-80). Association for Computational Linguistics.

Çakmak, F., & Erçetin, G. (2018). Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL, 30*(1), 24-47.

Carlesimo, G. A., Perri, R., Turriziani, P., Tomaiuolo, F., & Caltagirone, C. (2001). Remembering what but not where: independence of spatial and visual working memory in the human brain. *Cortex, 37*(4), 519-534.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and individual differences*, *19*(2), 246-251.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a development shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences, USA, 106*, 5529–5533.

Chen, H. (2002). Investigating the effects of L1 and L2 glosses on foreign language reading comprehension and vocabulary retention. Paper presented at the annual meeting of the Computer-Assisted Language Instruction Consortium, Davis, CA.

Chen, H. C., & Leung, Y. S. (1989). Patterns of lexical processing in a nonnative language. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 15*(2), 316 – 325.

Chen, I.-J. (2016). Hypertext glosses for foreign language reading comprehension and vocabulary acquisition: Effects of assessment methods. *Computer Assisted Language Learning*, *29*(2), 413-426.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The modern language journal*, *80*(2), 183-198.

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning.* San Francisco: Pfeiffer.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, *25*(3), 301-315.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*(3), 38-63.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of memory and language, 45*(4), 721-736.

Comesaña, M., Perea, M., Piñeiro, A., & Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Child Psychology*, *104*(1), 22-33.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin*, *104*(2), 163– 191.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory* (pp. 62–101). Cambridge: Cambridge University Press.

Craik, F. I. (2002). Levels of processing: Past, present ... and future? *Memory, 10*(5-6), 305-318.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684.

Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of verbal learning and verbal behavior, 12*(6), 599-607.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450-466.

De Groot, A., & Comijs, H. (1995). Translation recognition and translation production: Comparing a new and an old tool in the study of bilingualism. *Language Learning*, *45*(3), 467-509.

De Groot, A., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language learning*, *47*(2), 215-264.

Dehn, M. J. (2011). *Working memory and academic learning: Assessment and intervention.* John Wiley & Sons.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo–spatial memory. *Neuropsychologia*, *37*(10), 1189-1199.

Della Sala, S., & Logie, R. H. (2002). Neuropsychological impairments of visual and spatial working memory. *Handbook of memory disorders*, *2*, 271-292.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in cognitive sciences, 11*(9), 379-386.

Dilenschneider, R. F. (2018). Examining the conditions of using an on-line dictionary to learn words and comprehend texts. *ReCALL*, *30*(1), 4-23.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175-197.

Doran, G. T. (1981). There's a SMART way to write management's goals and objectives. *Management review, 70*(11), 35-36.

Doolittle, P. E., & Mariano, G. J. (2008). Working memory capacity and mobile multimedia learning environments: Individual differences in learning while mobile. *Journal of Educational Multimedia and Hypermedia, 17*(4), 511–530.

Doolittle, P. E., Terry, K. P., & Mariano, G. J. (2009). Multimedia learning and working memory capacity. In R. Zheng (Ed.), *Cognitive effects of multimedia learning* (pp. 17–33). London: Premier Reference Source.

Dubois, M., & Vial, I. (2000). Multimedia design: the effects of relating multimodal information. *Journal of Computer Assisted Learning*, *16*(2), 157-165.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.*, *55*, 51-86.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, *23*(2), 166-180.

Dufva, M., & Voeten, M. J. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied psycholinguistics, 20*(3), 329-348.

Duñabeitia, J. A., & Costa, A. (2015). Lying in a native and foreign language. *Psychonomic bulletin & review, 22*(4), 1124-1129.

Duquette, L., Renié, D., & Laurier, M. (1998). The evaluation of vocabulary acquisition when learning French as a second language in a multimedia environment. *Computer Assisted Language Learning*, *11*(1), 3-34.

Dutke, S., & Rinck, M. (2006). Multimedia learning: Working memory and the learning of word and picture diagrams. *Learning and Instruction, 16*(6), 526–537.

Dziemianko, A. (2019). The role of online dictionary advertisements in language reception, production, and retention. *ReCALL*, *31*(1), 5-22.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (No. 3). University Microfilms.

Ekstrom, R. B., Dermen, D., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests* (Vol. 102). Princeton, NJ: Educational testing service.

Ellis, R. (2000). Task-based research and language pedagogy. *Language teaching research*, *4*(3), 193-220.

Esit, Ö. (2011). Your verbal zone: an intelligent computer-assisted language learning program in support of Turkish learners' vocabulary learning. *Computer Assisted Language Learning*, *24*(3), 211-232.

Esterman, M., & Yantis, S. (2009). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex, 20*(5), 1245-1253.

Eysenck, M. W., & Keane, M. T. (2010). *Cognitive psychology: A student's handbook*. Taylor & Francis.

Fagan, W. F., Lewis, M. A., Auger-Méthé, M., Avgar, T., Benhamou, S., Breed, G., LaDage, L., Schlägel, U.E., Tang, W.W., Papastamatiou, Y.P. and Forester, J., Mueller, T. (2013). Spatial memory and animal movement. *Ecology letters, 16*(10), 1316-1329.

Fagehi, A. Y. (2013). Strategies for Defining and Understanding Critical Technology Integration Terms (Doctoral dissertation, University of Kansas).

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological review, 95*(3), 307.

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive psychology, 20*(4), 439-462.

Farmer, E. W., Berman, J. V., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 675-688.

Fernández, G., Biondi, J., Castro, S., & Agamenonni, O. (2016). Pupil size behavior during online processing of sentences. *Journal of integrative neuroscience, 15*(04), 485-496.

Field, J. (2012). The cognitive validity of the lecture-based question in the IELTS Listening paper. *IELTS Collected Papers 2: Research in Reading and Listening Assessment,* 2, 391-453.

Finke, R. A. (1989). *Principles of mental imagery*. The MIT Press.

Frank, S., & Thompson, R. (2012, January). Early effects of word surprisal on pupil size during reading. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, *129*(1), 61-83.

Galotti, K. M. (2017). *Cognitive psychology in and out of the laboratory*. Sage Publications.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of memory and language, 28*(2), 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*(4), 439-454.

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. New York: Psychology Press.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1-16.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental psychology, 28*(5), 887.

Gay, G. (1986). Interaction of learner control and prior understanding in computer-assisted video instruction. *Journal of Educational Psychology, 78*(3), 225-227.

Gazzaley, A. (2012). Top-down modulation de cit in the aging brain: An emerging theory of cognitive aging. In D. T. Stuss and R. T. Knight (eds.), *Principles of Frontal Lobe Function*, 2nd edn. (pp. 593–608). New York: Oxford University Press.

George, D. & Mallery, M. (2010). *SPSS for Windows step by step: a simple guide and reference*. Boston, MA: Pearson Education, Inc.

Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: approaches to second language vocabulary learning. *Language, Learning & Technology, 14*(2), 4-11.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review, 105*(2), 251-79.

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: Insights from eye movements and pupillometry. *Journal of Experiment Psychology; Learning, Memory, & Cognition, 35*(5), 1105-1122.

Goldstein, A. G., & Chance, J. E. (1971). Visual recognition memory for complex configurations. *Attention, Perception, & Psychophysics, 9*(2), 237-241.

Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological bulletin, 77*(5), 340-355.

Granholm, E., & Verney, S. P. (2004). Pupillary responses and attentional allocation problems on the backward masking task in schizophrenia. *International Journal of Psychophysiology, 52*(1), 37-51.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*(4), 457-461.

Grundy, J., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, *33*(3), 325–340.

Guasch, M., Ferre, P., & Haro, J. (2017). Pupil dilation is sensitive to the cognate status of words: further evidence for non-selectivity in bilingual lexical access. *Bilingualism: Language and Cognition, 20*(1), 49-54.

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics (IRAL)* 50, 239-76.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning*, *60*(2), 5-24.

Gustafson, K. L., & Branch, R. M. (1997). *Survey of Instructional Development Models* (3rd ed.). Syracuse University, NY: ERIC Clearinghouse on Information Resources.

Gustafson, K.L., & Branch, R.M. (2002). What is instructional design? In R.A. Reiser & J.V. Dempsey (Eds.), *Trends and issues in instructional design and Technology*. (pp. 16-25). New York: Merrill.

Gyselinck, V., Cornoldi, C., Dubois, V., De Beni, R., & Ehrlich, M. F. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology*, *16*(6), 665-685.

Gyselinck, V., Ehrlich, M. F., Cornoldi, C., De Beni, R., & Dubois, V. (2000). Visuospatial working memory in learning from multimedia systems. *Journal of Computer Assisted Learning*, *16*(2), 166-176.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics,* volume 2, pages 159–166.

Handley, Z. (2014). Vocabulary CALL for Young ESL/EFL Learners: A systematic review of the research evidence. In *Engaging Language Learners through Technology Integration: Theory, Applications, and Outcomes* (pp. 299–325). New Jersey, US: IGI Global.

Harman, H. H., Ekstrom, R. B., & French, J. W. (1976). *Kit of factor reference cognitive teats.* Educational Testing Service. Princeton NJ.

Hayet, M. (1994). A French learner's workbench. *Computers & Education*, *23*(1), 97-106.

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: language and cognition, 1*(3), 213-229.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349-350.

Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System*, *41*(3), 639-653.

Hitch, G. J., Halliday, S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition, 16*(2), 120-132.

Hole, G. J. (1996). Decay and interference effects in visuospatial short-term memory. *Perception*, *25*(1), 53-64.

Hulstijn, J. (2000) Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady and T. Huckin (eds) *Second Language Vocabulary Acquisition* (pp. 203-224). Cambridge: Cambridge University Press.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology, 48A*, 598–612.

James, M. O. (1996). *Improving second language reading comprehension: A computer-assisted vocabulary development approach.* Doctoral Dissertation: Manoa: University of Hawaii.

Janisse, M. P. (1977). *Pupillometry. The psychology of the pupillary response.* New York: Wiley.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied linguistics*, *21*(1), 47-77.

Jolicoeur, P., & Kosslyn, S. M. (1985). Is time to scan visual images due to demand characteristics? *Memory & Cognition, 13*(4), 320-332.

Jonassen, D. H. (1992). What are cognitive tools? In *Cognitive tools for learning* (pp. 1-6). Springer, Berlin, Heidelberg.

Jones, L. C. (2009). Supporting student differences in listening comprehension and vocabulary learning with multimedia annotations. *Calico Journal*, *26*(2), 267-289.

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The modern language journal*, *86*(4), 546-561.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review, 99*(1), 122-149.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*(2), 310-339.

Just, M. A. and Carpenter, P. A. (1995). Spatial working memory: cognitive capacity in mental rotation, Unpublished Manuscript. Department of Psychology, Carnegie Mellon University, Pittsburgh, PA.

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology, 64*(10), 1971-1989.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

Kahneman, D., & Peavler, W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of Experimental Psychology, 79*, 312–318.

Kaplan-Rakowski, R., & Loranc-Paszylk, B. (2019). The impact of verbal and nonverbal auditory resources on explicit foreign language vocabulary learning. *System*, *85*, 1-10.

Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology, 41*(2), 175-185.

Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, *28*(2), 166-186.

Khezrlou, S., Ellis, R., & Sadeghi, K. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System*, *65*, 104-116.

Kim, D., & Gilman, D. A. (2008). Effects of text, audio, and graphic aids in multimedia instruction for vocabulary learning. *Educational Technology & Society, 11*(3), 114-126.

Kim, S. H. (2019). The effects of using images and video clips for vocabulary learning: Single words vs. phrasal verbs. *STEM Journal, 20*(2), 19-42.

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133*(3), 355-381.

Klein, W., & Dimroth, C. (2009). Untutored second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 503–522). New York: Academic Press.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*, 12-35.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P. Robinson (Ed.),

Second language task complexity: Researching the Cognition Hypothesis of language learning and performance (pp.39-60). Amsterdam: Benjamins.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology, 7*(3), 341–370.

Kosslyn, S. M. (1976). Using imagery to retrieve semantic information: A developmental study. *Child Development, 47*(2), 434–444.

Kosslyn, S. M. (2006). *Graph design for the eye and the mind.* Oxford, England: Oxford University Press.

Kosslyn, S. M., & Rabin, C. S. (1999). Imagery. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 387–389). Cambridge, MA: MIT Press.

Kosslyn, S. M., Seger, C., Pani, J. R., & Hillger, L. A. (1990). When is imagery used in everyday life? A diary study. *Journal of Mental Imagery, 14*, 131–152.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, *32*(1), 89-97.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212-218.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory, 2*(389-395), London: John Wiley & Sons.

Kroll, J. F., & Steward, E. (1990). *Concept mediation in bilingual translation.* Paper presented at the 31st Annual Meeting of the Psychonomie Society, New Orleans, LA.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language, 33*(2), 149-174.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second language research*, *18*(2), 137-171.

Kuchinke, L., Võ, M. L. H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology, 65*(2), 132-140.

La Heij, W., Hooglander, A., Kerling, R., & Van Der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language*, *35*(5), 648-665.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspectives on psychological science, 7*(1), 18-27.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian modern language review, 59*(4), 567-587.

Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review, 63*(1), 149-166.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology*, *21*(2), 32-51.

Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System*, *81*, 14-25.

Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences, 3*(6), 223-232.

Levelt, W. J., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological review, 98*(1), 122.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Massachusetts: MIT Press.

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford, UK: Oxford University Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.

Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward*. Hove, UK: Language Teaching Publications.

Li, Z. N., Drew, M. S., & Liu, J. (2014). Introduction to multimedia. In *Fundamentals of Multimedia* (pp. 3-24). Springer International Publishing.

Lim, L. (2014). Engaging student interpreters in vocabulary building: Web search with computer workbench. *ReCALL*, *26*(3), 355-373.

Lin, C. C., & Tseng, Y. F. (2012). Videos and Animations for Vocabulary Learning: A Study on Difficult Words. *Turkish Online Journal of Educational Technology-TOJET*, *11*(4), 346-355.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*, 21(4), 861–83.

Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology,* 44(1), 87-112.

Loewenfeld, I. (1993). *The pupil: Anatomy, physiology, and clinical applications*. Detroit, MI: Wayne State University Press.

Logie, R. H. (1995). *Essays in cognitive psychology. Visuo-spatial working memory*. NJ, US: Hillsdale.

Logie, R., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial, or central executive? In R. H. Logie, & M. Denis (Eds.), *Mental Images in Human Cognition (Advances in Psychology)* (pp. 105-115). Amsterdam: Elsevier B.V.

Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1272-1299.

Ma, Q., & Kelly, P. (2006). Computer assisted vocabulary learning: Design and evaluation. *Computer Assisted Language Learning*, *19*(1), 15-45.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive psychology, 24*(1), 56-98.

MacLachlan, C., & Howland, H. C. (2002). Normal values and standard deviations for pupil diameter and interpupillary distance in subjects aged 1 month to 19 years. *Ophthalmic and Physiological Optics, 22*(3), 175-182.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*(3), 379-413.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, *43*(1), 31-41.

McCarthy, M. (1994) *Vocabulary*. Oxford: Oxford University Press.

McGeoch, J. A., & Irion, A. L. (1952). *The psychology of human learning* (2nd Ed.). Oxford, UK: Longmans, Green & Co.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 558-572.

Miller, G. A., & Fellbaum, C. (2007). WordNet then and now. *Language Resources and Evaluation*, *41*(2), 209-214.

Min, H. T. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning, 58*(1), 73-115.

Mirzaei, A., Domakani, M. R., & Rahimi, S. (2016). Computerized lexis-based instruction in EFL classrooms: Using multi-purpose LexisBOARD to teach L2 vocabulary. *ReCALL*, *28*(1), 22-43.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of memory and language, 33*(2), 175-202.

Mohsen, M. A. (2016). Effects of help options in a multimedia listening environment on L2 vocabulary acquisition. *Computer Assisted Language Learning*, *29*(7), 1220-1237.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Moreno, R., & Mayer, R. E. (2004). Personalized Messages That Promote Science Learning in Virtual Environments. *Journal of Educational Psychology, 96*(1), 165-173.

Moscovitch, M., & Craik, F. I. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15*(4), 447-458.

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society: B, 364*, 1273–1280.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17-38.

Nation, I. S. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nation, I. S. (2013). *Learning vocabulary in another language* (2nd edition). Cambridge, UK: Cambridge University Press.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum.

Nuthmann, A., & Van Der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology, 42*(3), 306-317.

Paivio, A. (1975). Coding distinctions and repetition effects in memory. *Psychology of learning and motivation*, *9*, 179-214.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian journal of psychology*, *45*(3), 255-287.

Paivio, A. (2006). Dual coding theory and education. In S. Neuman (ed.), *Pathways to Literacy Achievement for High Poverty Children*. The University of Michigan School of Education.

Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of verbal learning and verbal behavior*, *20*(5), 532-539.

Palladino, P., & Cornoldi, C. (2004). Working memory performance of Italian students with foreign language learning difficulties. *Learning and individual differences, 14*(3), 137-151.

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *The Quarterly Journal of Experimental Psychology Section A, 44*(1), 47-67.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30*(3), 331-347.

Papesh, M. H., & Goldinger, S. D. (2008). Pupil-blah-metry: Word frequency reflected in cognitive effort. In *Annual Meeting of the Psychonomic Society*, Chicago, IL.

Papesh, M. H., & Goldinger, S. D. (2012). Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics, 74*(4), 754-765.

Pavičić Takač, V. (2008). *Vocabulary learning strategies and foreign language acquisition.* Clevedon, UK: Multilingual Matters.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-225). London: Longman.

Pazzaglia, F., Toso, C., & Cacciamani, S. (2008). The specific involvement of verbal and visuospatial working memory in hypermedia learning. *British Journal of Educational Technology*, *39*(1), 110-124.

Pederson, K. (1987) Research on CALL. In W. F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation* (pp. 99-131). Lincolnwood, IL: National Textbook.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283-290.

Pinker, S. (1985). Visual cognition: An introduction. In S. Pinker (Ed.), *Visual cognition* (pp. 1–63). Cambridge, MA: MIT Press.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*, 1–10.

Plass, J., & Jones, L. (2005). Multimedia learning in second language acquisition. In Mayer, R.E. (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467-488). New York: Cambridge University Press.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25-36.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*(2), 221-243.

Poarch, G. J., Van Hell, J. G., & Kroll, J. F. (2015). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism: Language and Cognition*, *18*(3), 357-371.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly journal of experimental psychology, 60*(2), 211-229.

Potter, M. C., So, K. F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of verbal learning and verbal behavior, 23*(1), 23-38.

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production.* Amsterdam, Philadelphia: John Benjamins.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied linguistics, 15*(1), 36-57.

Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. *Language Learning & Technology, 23*(2), 105-124.

Rast, R. (2010). The use of prior linguistic knowledge in the early stages of L3 acquisition. *IRAL-International Review of Applied Linguistics in Language Teaching, 48*(2-3), 159-183.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 124* (3), 372–422.

Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge university press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146–161.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: Elsevier.

Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139*(1), 5-21.

Rimrott, A. (2010). Computer-assisted vocabulary learning: Multimedia annotations, word concreteness and individualized instruction. Doctoral dissertation, Simon Fraser University, Burnaby, BC, Canada.

Rouhi, A., & Mohebbi, H. (2013). Glosses, spatial intelligence, and L2 vocabulary learning in multimedia context. *3L: The Southeast Asian Journal of Language Studies, 19*(2), 75–87.

Rubin, D.C., & Wallace, W.T. (1989). Rhyme and reason: Analyses of dual cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 698–709.

Rumelhart, D. E., & Norman, D. A. (1988). Representation in memory. In R. C. Atkinson, R. J. Hernsein, G. Lindzey, and R. D. Luce (Eds.), *Stevens Handbook of Experimental Psychology* (pp. 545-571). New York: Wiley.

Rusanganwa, J. A. (2015). Developing a multimedia instrument for technical vocabulary learning: A case of EFL undergraduate physics education. *Computer Assisted Language Learning*, *28*(2), 97-111.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*, 344–355.

Sato, T., Matsunuma, M., & Suzuki, A. (2013). Enhancement of automatization through vocabulary learning using CALL: Can prompt language processing lead to better comprehension in L2 reading? *ReCALL*, *25*(1), 143-158.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and language, 17*(1), 133-145.

Schluroff, M., Zimmermann, T. E., Freeman Jr, R. B., Hofmeister, K., Lorscheid, T., & Weber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and Language, 27*(2), 322-344.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in psychology, 5*, 1-16.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 199–227). Cambridge: Cambridge University Press.

Schmitt, N. (2007). Current perspectives on vocabulary teaching and learning. In J. Cummins and C. Davison (Eds.), *International Handbook of English language teaching,* (pp. 827-841). New York: Springer.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* Basingstoke, UK: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913-951.

Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review*, *23*(3), 389-411.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology, 45*(1), 21-50.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of experimental psychology: General*, *125*(1), 4-27.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly journal of experimental psychology*, *22*(2), 261-273.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied psycholinguistics, 25*(2), 293-321.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, *99*(2), 195-231.

SR research Ltd. (2005–2008). EyeLink User Manual (version 1.4.0).

Stanners, R. F., Coulter, M., Sweet, A. W., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion, 3*(4), 319-340.

Stanners, R. F., Headley, D. B., & Clark, W. R. (1972). The pupillary response to sentences: Influences of listening set and deep structure. *Journal of Verbal Learning and Verbal Behavior, 11*(2), 257-263.

Sternberg, R. J., & Sternberg, K. (2012). *Cognitive psychology* (6th Ed.). Belmont, CA: Wadsworth, Cenage Learning.

Stockwell, G. (2007). A review of technology choice for teaching language skills in areas in the CALL literature. *ReCALL Journal, 19*(2), 105-120.

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362*, 901-915.

Sun, Y., & Dong, Q. (2004). An experiment on supporting children's English vocabulary learning in multimedia context. *Computer assisted language learning*, *17*(2), 131-147.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in second language acquisition*, *28*(3), 387-422.

Sutherland, I. E. (1963). Sketchpad, A Man-Machine Graphical Communication System, (Doctoral dissertation), Massachusetts Institute of Technology.

Sweller, J. (2002). Visualisation and instructional design. In *Proceedings of the International Workshop on Dynamic Visualizations and Learning* (pp. 1501-1510).

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251-296.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*(1), 45-58.

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, *28*(3), 341-348.

Tozcu, A., & Coady, J. (2004). Successful learning of frequent vocabulary through CALL also benefits reading comprehension and speed. *Computer assisted language learning, 17*(5), 473-495.

Tsai, K. J. (2019). Corpora and dictionaries as learning aids: inductive versus deductive approaches to constructing vocabulary knowledge. *Computer Assisted Language Learning*, DOI: 10.1080/09588221.2018.1527366.

Tulving, E. (1972). Episodic and semantic memory. *Organization of memory, 1*, 381-403.

Türk, E., & Erçetin, G. (2014). Effects of interactive versus simultaneous display of multimedia glosses on L2 reading comprehension and incidental vocabulary learning. *Computer Assisted Language Learning*, *27*(1), 1-25.

Uriel, W. (1953). *Languages in contact, Findings and problems*. New York: Linguistic Circle of New York.

Van Hell, J. G., & De Groot, A. M. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and cognition*, *1*(3), 193-211.

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of memory and language*, *39*(3), 458-483.

Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*(2), 167-174.

Varol, B., & Erçetin, G. (2019). Effects of gloss type, gloss position, and working memory capacity on second language comprehension in electronic reading. *Computer Assisted Language Learning*, DOI: 10.1080/09588221.2019.1643738.

Võ, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology, 45*(1), 130-140.

Ware, C. (2004). *Information Visualization: Perception for Design* (2nd Edition). San Francisco, CA: Morgan Kaufman.

Whipple, B., Ogden, G., & Komisaruk, B. R. (1992). Physiological correlates of imagery-induced orgasm in women. *Archives of sexual behavior, 21*(2), 121-133.

Williams, J. N. (1994). The relationship between word meanings in the first and second language: Evidence for a common, but restricted, semantic code. *European Journal of Cognitive Psychology*, *6*(2), 195-220.

Williamson, V., Baddeley, A., & Hitch, G. (2010). Musicians' and non-musicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory and Cognition 38*(2), 163–75.

Wyatt, H. J. (1995). The form of the human pupil. *Vision Research, 35*(14), 2021-2036.

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology, 13*, 48–67.

Yeh, Y., & Wang, C. W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *Calico Journal*, *21,* 131-144.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, *10*(3), 85-101.

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal, 20*(1), 33-58.

Young, J. D. (1996). The effect of self-regulated learning strategies on performance in learner controlled computer-based instruction. *Educational technology research and development*, *44*(2), 17-27.

APPENDICES

Appendix 1

Participants' consent form (first study)

Dear participant, my name is Saad Alzahrani, PhD student in the in the Department of Education at the University of York. I am writing to invite you to take part in a research study.

Title of Study

Effective vocabulary learning in multimedia CALL environments: psychological evidence

Brief Description of Study

This study aims to explore the effectiveness on foreign language learning via different types of multimedia representation of new words. You will learn new Arabic words using a computer-based instructional program. After that, we will investigate how many words you have learned via a recall test. The study will take no more than two hours. On the following day, you will be asked to take part in a memory test, which will last approximately 45 to 60 minutes. You will receive £5 per hour for your participation. You are free to stop your participation at any point, without giving any specific reason, without your rights affected.

All the data collected during the experiment will be fully anonymized, and they will be securely stored in a password-protected computer / locked office, and only the researcher will have access to these data. If your individual data are used in future presentations or publications, you will not be identified. If you would like to ask any further questions about this study or to get your tests scores, please email Saad Alzahrani on (saad.alzahrani@york.ac.uk) or the Chair of the Ethics Committee Dr Paul Wakeling (paul.wakeling@york.ac.uk).

INFORMED CONSENT

I have read the statement concerning the research that I am being asked to take part in, and I have had the opportunity to ask questions. I understand that I may withdraw at any time during data collection time, and within 15 days after the completion of data collection completion, and if I decided to do so my data will be safely disposed of. I understand that my data will be kept confidential and I am happy to take part in the research.

(Participant)

Printed Name: Signature: Date:

(Researcher)

Printed Name: Saad Alzahrani Signature: Date:

Appendix 2

Participants’ consent form (second study)

Dear participant,

My name is Saad Alzahrani, PhD student in the in the department of education at the University of York. I am writing to invite you to take part in a research study.

Title of Study

Effective vocabulary learning in multimedia CALL environments: psychological evidence

Brief Description of Study

This study aims to explore the effectiveness on foreign language learning via different types of multimedia representation of new words. You will learn new Arabic words using a computer-based instructional program. After that, you will be asked to take a translation recognition task in the Eye tracker lab at the department of Education where your eye track will be recorded. The whole procedure will take approximately two hours of your time. You will receive £5 per hour for your participation. You are free to stop your participation at any point, without giving any specific reason, without your rights affected.

All the data collected during the experiment will be fully anonymized, and they will be securely stored in a password-protected computer / locked office, and only the researcher will have access to these data. If your individual data are used in future presentations or publications, you will not be identified. If you would like to ask any further questions about this study or to get your tests scores, please email Saad Alzahrani on (saad.alzahrani@york.ac.uk) or the Chair of the Ethics Committee Dr Paul Wakeling (paul.wakeling@york.ac.uk).

INFORMED CONSENT

I have read the statement concerning the research that I am being asked to take part in, and I have had the opportunity to ask questions. I understand that I may withdraw at any time during data collection time, and within 15 days after the completion of data collection completion, and if I decided to do so my data will be safely disposed of. I understand that my data will be kept confidential and I am happy to take part in the research.

(Participant)

Printed Name: Signature: Date:

(Researcher)

Printed Name: Saad Alzahrani Signature: Date:

Appendix 3

CAVA

The two versions of CAVA, which were designed and used in the present research, are provided in soft copies in the attached CDs. Each file includes the software, its post-tests and general instructions for installation and use.