University of Southampton

FACULTY OF HUMANITIES

Modern Languages

L2 Incidental Vocabulary Learning and Retention through Different Modalities of Audio-visual Input

by

Hassan Alshumrani

Thesis for the degree of Doctor of Philosophy

April, 2019

University of Southampton

Abstract

Faculty of Arts and Humanities

Modern Languages

Thesis for the degree of Doctor of Philosophy

L2 Incidental Vocabulary Learning and Retention through Different Modalities of Audio-visual Input

Hassan Alshumrani

While the bulk of previous studies on incidental vocabulary learning through audio-visual materials have looked at the differential effects of some input modalities (e.g., L1 subtitles vs L2 captions), little research has examined the effects of other important modalities of audio-visual input. The present research study investigates L2 incidental vocabulary short-term learning and long-term retention in four different audio-visual input conditions. More precisely, adopting a quasi-experimental research design, this study compares the effects of four modalities of audio-visual input: video, audio, and caption (VAC), video and audio (VA), caption and audio, (CA), and audio only (A only) on incidental learning and retention of knowledge of 36 target words' spoken form recognition, meaning recall, and meaning recognition. Additionally, the study examines the predictive roles of an item-related variable (frequency of occurrence) and a learner-related variable (working memory) in incidental vocabulary learning through the four different input conditions.

The study used a range of data collection methods. Vocabulary knowledge was assessed through three vocabulary tests: spoken form recognition, spoken meaning recall, and spoken meaning recognition. These were administered at three different time points as, pre-tests, immediate post-tests, and delayed post-tests. Working memory capacity was measured using two verbal tests, (forward digit recall and backward digit recall) and two visuospatial tests, (dot matrix and odd one out).

The study demonstrated that the four audio-visual input conditions resulted in L2 incidental vocabulary learning of the three vocabulary knowledge dimensions. The findings showed that the four modalities of audio-visual input had differential effects on incidental short-term learning of the three vocabulary knowledge types. The captioning conditions (CA and VAC) were more effective than the non-captioning conditions (VA and A only) for fostering form learning. The visual condition (VA) was the most effective condition for promoting meaning knowledge. Additionally, large attrition rates of the three vocabulary knowledge dimensions were found across the four experimental groups. The results also demonstrated that the effect of frequency of occurrence varied based on the modalities of audio-visual input and the target word knowledge aspects. In relation to the role of working memory, the findings indicated that individual differences in working memory capacity did not account for the variations in the vocabulary scores on the immediate and delayed post-tests. A number of pedagogical implications regarding the effects of the different modalities of audio-visual input on vocabulary learning and retention are presented.

Table of Contents

List of Tables

Table of Figures

Research Thesis: Declaration of Authorship

I, Hassan Alshumrani declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

L2 Incidental Vocabulary Learning and Retention through Different Modalities of Audio-visual Input

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: Date:

Acknowledgements

First and foremost, I would like to thank Allah (God) the Almighty for granting me the strength and ability to complete this research project. His endless grace, bounties, and assistance made this thesis possible.

My special thanks go to my first supervisor, Dr Sarah Rule, for all that she offered me while working on this study. I am deeply indebted to Dr Sarah for her unremitting patience, continuous support, insightful comments, constant guidance, continuous reassurance, and encouragement. I would not have been able to complete this PhD research without her help.

I also owe a great deal to my second supervisor, Dr Alison Porter, for her support, invaluable insights, and thought-provoking feedback. She was involved in the different phases of this study. Her invaluable guidance and suggestions contributed significantly to improving this thesis.

I wish to extend my sincere gratitude to my beloved parents, Ahmed and Saleha for their unconditional love, ongoing prayers, endless support, and continuous encouragement. To them, I am eternally grateful. I am also particularly grateful to my elder brother, Mohammed, for his support, advice, encouragement and words of concern.

My acknowledgements would not be complete without thanking my beloved wife, Manal, for her love, support, and ongoing encouragement. My heartfelt thanks are also due to my lovely daughters, Lujane and Reema, for the joy they brought in my life. My family made my stay in Southampton enjoyable and unforgettable. I would not have been able to undertake this endeavour without their presence.

I would like to acknowledge the support of the individuals who helped me with my field work. A very special thank you goes out to Mr. Adel Almutairi, Dr Abdullah Alshehari, and Mr. Sami Almaliki for allowing me to conduct my study in their classes. I should also thank all my participants for their cooperation and agreeing to participate in the study. Without their help, this study would have been impossible.

Lastly but not least, I would like to thank my best friend, Saleh Alzahrani, for his prayers, remembrance, and words of concern. He is a source of optimism and wisdom for me.

Chapter 1 Introduction

The present study is concerned with L2 incidental vocabulary short-term learning and long-term retention through different modalities of audio-visual input. It also examines the role of both an item-related variable (i.e. frequency of occurrence) and a learner-related variable (i.e. working memory) in L2 incidental vocabulary learning and retention through different modalities of audio-visual input. This introductory chapter aims to present a complete picture of the current research study by first providing some background information about this study. It then highlights the motivation for carrying out the study. The chapter further proceeds by presenting the aims and the research questions of the study. Finally, it outlines the structure of the study. Overall, this chapter intends to provide the reader with a holistic overview of the study, prior to moving forward to the theoretical and conceptual frameworks and the research methodology in the following chapters.

1.1 Background to the study

The importance of vocabulary is widely acknowledged and is perceived by language acquisition researchers as one of the most essential components in any language. One of the most oft-cited statements to advocate this importance is put forward by Wilkins (1972, p. 111) who states, "without grammar very little can be conveyed, without vocabulary nothing can be conveyed". As such, Nation and Waring (1997) assert that vocabulary knowledge plays a key role in attaining proficiency in a second/foreign language. Schmitt, Cobb, Horst, and Schmitt (2015, p. 213) add "knowledge of vocabulary is fundamental to all language use, and so must be learned in some manner in order for learners to become communicative in a new language". These statements collectively mark the perceived importance of vocabulary learning as a key aspect of first language learning (L1) as well as second/foreign language learning (L2/FL) that language learners need to focus on. However, vocabulary learning is not an easy task for L2/FL learners, as they need to learn/acquire voluminous amounts of vocabulary items in order to use the language properly (Laufer, 1992).

The number of words that L2/FL learners have to know in order to be able to comprehend spoken and written discourse is well established. It is suggested that they need to know around 3000 to 5000 word families to comprehend spoken discourse (Laufer, 1992; Peters & Webb, 2018; Van Zeeland & Schmitt, 2013). However, the amounts of words needed for understanding written texts are far greater than the amount needed for spoken discourse. According to Nation (2006), L2/FL learners need to know no less than 8000-9000 word families in order to understand written discourse. Unfortunately, many L2/FL learners fall well short of attaining such thresholds, as Schmitt (2010) noted. In effect, it has been said that these figures are daunting and represent one of the largest challenges posed for L2/FL learners (Schmitt, 2008). Therefore, it is a pressing concern for L2/FL teachers and learners to find out efficient ways to reach the proposed vocabulary threshold.

One of the central debates surrounding vocabulary learning is the role of incidental versus intentional learning approaches in learning new words (Hulstijn, 2003; Nation, 2001; Sonbul & Schmitt, 2010). Although the significance of the intentional approach in vocabulary learning is well-documented (Hulstijn, 2003; Nagy & Herman, 1987; Schmitt, 2008), it is argued that intentional learning alone cannot account for the thousands of words deemed necessary for a proficient use of English because of the limited amount of time devoted to vocabulary instruction and learning in most L2/FL classrooms (Peters & Webb, 2018). After reviewing eight different studies conducted in eight different countries, Laufer (2000) found that the vocabulary size of L2 learners ranged between 1.000 and 4.000 word families, which according to Schmitt, Jiang, and Grabe (2011) lags far behind the required number of words. Therefore, researchers have advocated the potential of the incidental learning approach to be used alongside the intentional learning approach to foster students' vocabulary knowledge, as it has been found that learners were able to expand their vocabulary knowledge incidentally through different input modes (Peters & Webb, 2018). Research has established that incidental vocabulary learning can take place through written input (reading) (Horst, 2005; Laufer & Rozovski-Roitblat, 2011; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003) as well as aural input (listening) (Brown, Waring, & Donkaewbua, 2008; Van Zeeland & Schmitt, 2013; Vidal, 2003, 2011). Although reading and listening are highly valuable resources for word learning, the process of incidental vocabulary acquisition through reading and listening is usually slow and fallible (Horst, Cobb, Cobb, & Meara, 1998). This in turn has been found to lead to small vocabulary gains (Pellicer-Sánchez & Schmitt, 2010; Van Zeeland & Schmitt, 2013).

Nevertheless, researchers have recently started to examine the potential of audio-visual inputs to boost incidental vocabulary learning. Several studies (for example, Montero Perez, Peters, Clarebout, & Desmet, 2014; Montero Perez, Peters, & Desmet, 2018; Peters, Heynen, & Puimège, 2016; Peters & Webb, 2018; Rodgers, 2013; Sydorenko, 2010) have shown that words can be acquired incidentally through audio-visual input. Rodgers (2013) indicated that audio-visual input is a promising way to fuel learners' vocabulary knowledge. His participants were able to learn approximately a quarter of the target words from watching multiple episodes of a TV programme. The basic premise of audio-visual materials is that they involve presenting input/information simultaneously in two or more input modalities/modes. The implementation of multiple modalities of input is found to improve language learning (Mayer, 2014; Sydorenko, 2010) and enhances retention (Chun & Plass, 1996). This view is supported by Paivio's (1986) dual coding theory and Mayer's (2005) cognitive theory of multimedia learning. These theories assume that incoming information is processed by two cognitive systems, verbal and nonverbal. The activation of these two systems in processing information augments the recollection of the presented information and leads to better learning results. However, since not all input modalities are equally effective (Mayer, 2014), it has become crucial to examine which modalities of audio-visual input are more effective for facilitating incidental vocabulary learning and retention.

Moreover, given the recognised importance of working memory (WM) in learning through multiple input modalities (Mayer, 2014), a crucial question remains, as to which modalities of audio-visual input can be more beneficial to incidental vocabulary learning and retention for learners with different WM capacities. WM is believed to play an essential role in different learning domains, more importantly many scholars have acknowledged its part in learning through the audio-visual input (Alloway, 2007; Baddeley, Papagno, & Vallar, 1988; Gupta, 2003; Schüler, Scheiter, & van Genuchten, 2011; Schweppe & Rummer, 2014). Given one of the main characteristics of WM is its limited capacity (Baddeley, 1986), this issue of limited memory resources can have adverse effects on learning through multiple modalities (text and pictures) (Mayer & Moreno, 2003). That is, the effectiveness of different input modalities can depend on the individual's WM capacity (Plass, Chun, Mayer, & Leutner, 2003). Since WM underlies our ability to learn from multiple input modalities, it is thus essential to shed some light on its role in incidental vocabulary learning and retention through different modalities of audio-visual input.

Although incidental vocabulary learning through audio-visual input has attracted an increasing amount of academic interest, the effects of different modalities of audio-visual input have received less attention. In addition, the role played by both the word-related variable (frequency of occurrence of the target words) and the learner-related variable (WM) has hardly been addressed in audio-visual studies. This research study, therefore, is an attempt to investigate L2 incidental vocabulary short-term learning and long-term retention of three distinct word knowledge dimensions: spoken form recognition, meaning recall, and meaning recognition through different modalities of audio-visual input. It also aims to explore the predictive roles of both frequency of occurrence and WM in L2 incidental vocabulary learning and retention through multiple modalities of audio-visual input.

1.2 Motivation of the study

The central motivation for carrying out this research project is based on my experiences as an English language instructor. The dearth of information in the literature is also a significant factor in my motivation. As a teacher, I have noticed that most Saudi learners of English are unable to express themselves clearly either when dealing with academic topics or with common everyday topics. They struggle to read and comprehend course materials. This suggests that they have a restricted vocabulary. A number of studies bear testimony to Saudi English learners being plagued by their poor vocabulary knowledge (Al Qahtani, 2005; AlSaif, 2011). These studies have revealed that Saudi students' vocabulary size ranged between 1000 and 2000 words, which is very limited compared to the vast amount of vocabulary needed for proficient use of English. This is very disheartening, given the huge budgets spent on English language teaching and learning in Saudi Arabia.

It is my contention that learners' difficulties in vocabulary coupled with their poor knowledge may arise from a number of factors. Firstly, the fundamental formats applied for vocabulary extension are word-lists and rote memorization. However, reliance on such intentional techniques only has been subject to some criticisms. For example, Nagy and Herman (1987) argued that the intentional learning approach cannot account for the many thousands of words learners need to learn in order to reach the threshold proposed by the literature for proficient language use. Similarly, Cobb (1999, p. 345) commented "students typically need to know words measured in thousands, not hundreds, but receive language instruction measured in months, not years". Further, vocabulary learning is not given the same importance as other aspects of language (Al Qahtani, 2005). Great emphasis is placed on the planning and teaching of grammar in most language classes. So much so that most language classes end up becoming grammar teaching sessions. The grammar-translation method is still common and widely adopted (Al-Homoud & Schmitt, 2009; Assalahi, 2013). The dominance of grammar instruction has led some language teachers to devalue the importance of vocabulary learning (Al Seghayer, 2001).

In the light of the above, it becomes necessary to explore other efficient ways to improve L2 vocabulary learning alongside the intentional learning/instructional techniques. The growing interest globally in researching vocabulary learning is motivated by the different approaches available with the aim of developing and enhancing the learning and retention of L2 words. One such approach that has received considerable attention is incidental vocabulary learning. In general, incidental word learning can occur through different modalities, including reading, listening, and viewing audio-visual materials. Most researchers have focused on reading and generally reported small vocabulary gains (Horst, 2005; Pellicer-Sánchez & Schmitt, 2010; Pigada

& Schmitt, 2006). In these studies, subjects were required to read large volumes of texts in order for incidental vocabulary learning to occur. However, many Saudi students' level in reading abilities is unsatisfactory (Al-Qahtani, 2016; Al-Roomy, 2013), which may make reading a difficult task. In fact, in my experience as an English language teacher, I have noticed that most Saudi learners have low levels of pleasure reading in English and the amount of time they spend on reading is rather low compared to the amount of time they spend on watching TV or films. This apparent lack of interest of reading may render reading an ineffective medium for learning.

Apart from the fact that most Saudi students prefer to watch videos and TV programmes, the motivational potential of the use of audio-visual materials in English learning has been evidenced by a great deal of past research (Montero Perez et al., 2014; Peters et al., 2016; Rodgers, 2013). This observation thus prompted me to explore the potential of audio-visual materials, as they may appeal to Saudi students' interests. However, one of the concerns raised is what modalities of audio-visual input are more conducive for fostering vocabulary uptake? An investigation of the effects of different modalities of audio-visual input bears important implications for the field of vocabulary learning and teaching. Of equal importance is the role of WM in learning through different input modalities. Since individual differences in WM capacity is one of the main factors that may play a predictive role in mediating the effects of different input modalities, it is thus essential to explore this issue (Kozan, Erçetin, & Richardson, 2015). However, it is astonishing that there is little attention paid to the influential role of WM in vocabulary learning through different modalities of audio-visual materials.

Although the L2 literature is rich with studies carried out on the effects of different modalities on incidental vocabulary learning, the need to undertake this doctoral research stems from the following observations. Firstly, the extant audio-visual studies have been found wanting because of several serious methodological flaws, as will be discussed in section 2.3.2. Secondly, none of the previous studies has examined the role of WM in incidental vocabulary learning and retention through multiple modalities of audio-visual input. To this end, this study will attempt to fill the gaps identified in the literature. It therefore sets out to examine the effectiveness of different modalities of audio-visual materials on L2 incidental vocabulary learning and retention. Additionally, it aims to explore the predictive role played by the item-related factor (frequency of occurrence) and the learner-related factor (WM) in incidental vocabulary learning through different input conditions.

1.3 The purpose and research questions of the study

Given the aim of filling the void observed in the literature, this experimental research intends to achieve the following aims. Firstly, it aims to examine the effectiveness of four different modalities of audio-visual input, video, audio and caption (VAC), video and audio (VA), caption and audio (CA), and audio only (A only), on L2 incidental vocabulary short-term learning and long-term retention of three aspects of vocabulary knowledge: spoken form recognition, meaning recall, and meaning recognition. The research further aims to compare the effects of these input conditions - VAC; VA; CA; and A only - on L2 incidental vocabulary learning and retention. It also attempts to examine the effect of the frequency of occurrence on incidental vocabulary learning of the three vocabulary knowledge aspects. Furthermore, the study aims to look at the relationship between WM and incidental vocabulary learning and retention. Finally, it aims to examine the role of WM in mediating the effects of the different input modalities on L2 incidental vocabulary learning and retention. These aims are set out in the following research questions:

1. To what extent do audio-visual conditions, VAC, VA, CA, and A only, enhance/promote L2 incidental vocabulary short-term learning of spoken form recognition, meaning recall, and meaning recognition?
2. Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary short-term learning of the spoken form recognition, meaning recall, and meaning recognition?
3. Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary long-term retention of the three vocabulary knowledge aspects, spoken form recognition, meaning recall, and meaning recognition?
4. To what extent does the frequency of occurrence affect L2 incidental vocabulary learning from the four experimental conditions?
5. Is there a relationship between WM and incidental vocabulary short-term learning and long-term retention? If so, which subset(s) of WM, phonological loop or visuospatial sketchpad, contribute(s) to this relationship?
6. Which of the WM task types, simple or complex, best predicts incidental vocabulary short-term learning and long-term retention?
7. Does WM mediate the effect of the four modalities of audio-visual input on incidental vocabulary short-term learning and long-term retention?

To answer the above research questions, a range of data collection instruments were deployed. The instruments include three vocabulary tests: spoken form recognition, spoken meaning recall, and spoken meaning recognition. These were administered at three different times as: pre-tests, immediate post-tests, and delayed post-tests. The study also used four WM measures: forward digit recall, backward digit recall, dot matrix, and the odd one out to gauge the capacity of two components of WM, the phonological loop and the visuospatial sketchpad.

1.4 Structure of the thesis

This thesis consists of seven chapters:

Chapter one is the introductory chapter. It briefly introduces the study to the reader. The chapter opens with an overview of the study. It then demonstrates the motivation for undertaking this research. It also presents the aims of the study and the research questions. Finally, it concludes with an outline of the structure of the study.

Chapter two is divided into three main sections. The chapter aims to provide a thorough overview of the research literature pertaining to the three main themes of the study: vocabulary learning, theoretical underpinnings of the study, and the various modalities of audio-visual input. The first section discusses different issues related to vocabulary learning. The second section presents the theoretical framework that underpins the current study. The last section reviews relevant studies conducted in the field of incidental vocabulary learning through multiple modalities of input and concludes with a description of the perceived gaps in the literature.

Chapter three outlines the methodology of the study. It is dedicated to provide a layout of the study design. It begins with a discussion of the philosophical stances and the research paradigms. It then moves on to provide a detailed account of the design, participants, materials, data collection methods, and the procedures of the study. Next, issues of validity and reliability are discussed. This is followed by a description of the two pilot studies carried out before the main study. The last part of this chapter presents the analytical framework that will be followed for the analysis of the data.

Chapter four presents the analysis of the vocabulary data. The first section reports the findings of the first research question pertaining to incidental vocabulary learning through the four different audio-visual input conditions. The second and third sections present the results related to the differential effects of the four input conditions on incidental vocabulary short-term learning and long-term retention. The last section of this chapter reports the findings of the fourth research question that looked at the role of frequency of occurrence on incidental vocabulary learning.

Chapter five is devoted to the analyses of the WM data. The first part of this chapter begins with the results of the relationship between WM and incidental vocabulary short-term learning and long-term retention. It then reports the results of the correlational and multiple regression analyses carried out. Finally, it presents the finding of the role of WM in incidental vocabulary learning and retention through the multimodal input conditions.

Chapter six provides a detailed discussion of the findings analysed in chapter four and five in relation to the relevant literature and the theoretical accounts. The principle findings of each research question will be discussed in a separate section.

Chapter seven is the concluding chapter. This chapter opens with an overview of the study. It then gives a brief summary of the key findings of this study. This is followed by a discussion of the contribution and implications of the study. Finally, the chapter describes the limitations of the study and proposes some future directions.

Chapter 2 Literature Review

This chapter presents a theoretical and conceptual overview of the study. It is divided into three major sections; each section corresponds to one main theme. The first section discusses a number of issues related to L2 vocabulary learning. The second part of this chapter provides the theoretical framework that underpins the study. The last section focuses on the concept of input modalities and provides a review of a wide range of existing literature in the field of L2 incidental vocabulary learning, through multiple modalities of input. The chapter concludes with a summary of the focus of the study and the rationale behind it.

2.1 Vocabulary learning

In the field of L2 education, grammar has traditionally been given precedence and due attention, while vocabulary occupied a marginal position due to, as Beltrán, Abello-Contesse, and del Mar Torreblanca-López (2010) commented, the influence of the structuralist view, which considers language as a set of grammatical rules, rather than unlimited subsystems, like words. Oxford and Scarcella (1994, p. 241) illustrated the issue of the neglect of vocabulary remarking; "vocabulary instruction is ignored in many L2 classes, on the assumption that students themselves will simply find ways to memorize words without any help". However, after decades of neglect, particularly in 1980s, the field of vocabulary learning has garnered a certain salience (Schmitt, 2000). Nowadays, a wealth of studies have investigated vocabulary from various perspectives (Schmitt, 2010). In particular, researchers have explored different issues related to the field of vocabulary learning, including vocabulary breadth and depth (Qian, 1999; Vermeer, 2001; Webb, 2005; Wesche & Paribakht, 1996); vocabulary learning strategies (Hamzah, Kafipour, & Abdullah, 2009; Kalajahi & Pourshahian, 2012; Lawson & Hogben, 1996); and incidental and intentional vocabulary learning approaches (Al-Homoud & Schmitt, 2009; Alamri & Rogers, 2018; Hennebry, Rogers, Macaro, & Murphy, 2017; Laufer & Rozovski-Roitblat, 2011, 2015; Sonbul & Schmitt, 2013; Webb, 2008; Webb & Chang, 2015).

Vocabulary is a fundamental component of overall language competence. In essence, the development of other language skills such as writing, reading, listening, and speaking is largely dependent on vocabulary knowledge (Laufer, 1992; Stæhr, 2009). Words are the key part of a language and they, as Milton (2009, p. 3) argued "are the building blocks of language and without them there is no language". The importance of words is unquestionable and lies within all stages of language acquisition. In effect, vocabulary knowledge has been directly linked to other language skills, including reading comprehension and listening comprehension, suggesting that vocabulary knowledge plays a predictive part in reading and listening comprehension (Laufer & Ravenhorst-Kalovski, 2010; Qian, 2002; Schmitt et al., 2011; Stæhr, 2009). Laufer and Ravenhorst-Kalovski (2010) noted that participants with a large vocabulary size were able to able to read the frequent lexical items in the text more fluently than their counterparts with a smaller vocabulary size.

The complex nature of word knowledge makes vocabulary learning a difficult undertaking. Vocabulary learning involves learning several characteristics about words and is inherently 'incremental' for a number of reasons (Schmitt, 2010). Firstly, vocabulary knowledge entails several aspects that are not acquired simultaneously, for example knowledge of spoken and written form, meaning, collocation, grammatical functions, associations etc. Second, the development of each dimension of word knowledge lies on a continuum. This means that receptive knowledge of vocabulary is acquired first and this knowledge then gradually turns into a productive knowledge (Schmitt, 2010). In addition, the level of receptive and productive mastery of each vocabulary knowledge aspect differs. Schmitt (2000, p. 4) noted, "one thing we can be sure of is that words are not instantaneously acquired. Rather they are gradually learned over a period of time". Moreover, Schmitt (2000, p. 6) contended that vocabulary learning is an "not all-or-nothing" phenomenon, that is, knowledge of words develops gradually.

Nevertheless, despite the complex processes involved in vocabulary acquisition, a wide range of studies have attempted to describe the process in L2 word acquisition (for example, Al-Homoud & Schmitt, 2009; Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010; Peters et al., 2016; Rodgers, 2013; Van Zeeland & Schmitt, 2013). These studies unanimously agreed that vocabulary can be acquired in an incidental way. The incidental vocabulary learning approach has been the subject of extensive research over the past four decades. A discussion of the distinction between incidental and intentional vocabulary learning is provided below. However, prior to that, some conceptual overviews regarding the nature of word knowledge as well as receptive and productive vocabulary knowledge are considered.

2.1.1 To know a word

Vocabulary knowledge is an essential aspect in language learning, however, knowing what word knowledge entails is problematic. One of the frequently asked questions is what does it mean to know a word? Different answers to this question can be expected. The most common one would be that vocabulary knowledge includes two facets: form and meaning. Although these two facets are of paramount importance, vocabulary knowledge also entails other important aspects (Nation, 2001; Schmitt, 2010). Laufer and Goldstein (2004) argued that vocabulary knowledge is more complicated than a form-meaning connection. In effect, vocabulary knowledge is a complex construct that can be divided into separate aspects. Nation (2001, p. 23) stated "words are not isolated units of language, but fit into many interlocking systems and levels". Similarly, Nation (2004) posited that vocabulary learning is a cumulative process that involves enhancing and developing receptive and productive knowledge of words. The importance of mastering the different aspects of word knowledge lies in the fact that it features a proper and effective use of the target language (Schmitt, 2010).

Over the last forty years, there have been many attempts to define what knowing a word means. One of the well-established ways of conceptualizing vocabulary knowledge is known as the 'dimension approach' (Schmitt, 2010) or the 'multi-componential approach' (Pellicer-Sánchez, 2016). Richards (1976) introduced one of the first influential taxonomies that adopted the 'dimension approach' to explicitly define what constitutes word knowledge. In his taxonomy, Richards divided vocabulary knowledge into a variety of related aspects as follows: (1) knowledge of the degree of the probability of meeting a word in speech or print, (2) knowledge of the limitations imposed on the word use, (3) knowledge of the associative syntactic behaviour of the word, (4) knowledge of different forms that can be derived from the base form, (5) knowledge of the word relationship with other words, (6) knowledge of the semantic value of the word, and (7) knowledge of the distinct or related meanings of a word. Richards' framework involves knowledge of both receptive and productive knowledge of various dimensions of knowledge of a word. Although many lexical studies have used this taxonomy in assessing different types of vocabulary knowledge, it is said it is not comprehensive (Qian, 2002), as it did not include all possible knowledge types of words, such as spelling and pronunciation.

Several years later, Nation (2001) elaborated on Richards' framework and proposed a more exhaustive taxonomy. In his taxonomy, Nation included all the possible constructs of vocabulary knowledge in which he differentiated between receptive and productive knowledge of words. According to Nation's framework, knowing a word involves knowledge of three principle dimensions: form, meaning and function/use. Each of the three main aspects entails several sub-features of lexical knowledge that need to be acquired in order to develop full knowledge of a word (see Table 2.1).

Table 2.1. Different types of word knowledge

Form	Spoken form	R	What does the word sound like?
		P	How is the word pronounced?
	Written form	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use (register, frequency)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?
P= productive knowledge		R= receptive knowledge	

(Adapted from Nation, 2001, p. 27)

As shown in Table 2.1, Nation (2001) classified word knowledge into three main types: form, meaning and use. Each main type entails several sub-categories of lexical knowledge. The first category of Nation’s classification is form. 'Form’ entails knowledge of the spoken and written form of a given word. It also involves knowledge of different word parts. The second main category of Nation's taxonomy is ‘meaning’. Knowledge of meaning refers to knowing how to connect the form and the meaning of a word, knowing the relationship between the concept and its referent, and knowing the associations of a word. In addition, the third main aspect in Nation’s framework is ‘use’. Knowledge of the proper use of a word involves, firstly, knowledge of the grammatical functions of the word. The grammatical functions include knowledge of different parts of speech (e.g., noun, verb, adjective, adverb, etc.). Further, knowledge of use entails knowledge of collocations and knowledge of constraints of use of such words. Nation (2001) also applied receptive and productive knowledge to all the nine aspects of word knowledge.

The use of vocabulary knowledge framework of Nation (2001) has some implications. Schmitt (1995) contended that the use of this framework would produce more accurate measurements of vocabulary knowledge. Schmitt (2010) also stated that this framework thoroughly demonstrates the ‘incremental nature’ of word learning. Further, according to Schmitt (1995), one of the merits of Nation’s (2001) framework is that unlike other approaches that view vocabulary knowledge in a binary manner (i.e., as know/don’t know), Nation’s framework covers the different types of vocabulary knowledge. It can be claimed that this framework shows how well a learner knows a word or the quality/depth of knowing a word. However, despite the wide use of this taxonomy and its significance in illustrating the different types of word knowledge, it is thought that this framework is merely descriptive and provides no explanation for the acquisition processes of such aspects (Schmitt & Meara, 1997). Nation's (2001) classification of the different types of vocabulary knowledge is also said to include interrelated aspects despite being classified separately/differently (Schmitt & Meara, 1997).

A growing number of vocabulary studies have implemented the ‘dimension approach’ or ‘multi-componential approach’ in their designs. For instance, utilising a multiple-choice test and translation test, Waring and Takaki (2003) assessed participants’ incidental vocabulary uptake of meaning recognition and meaning recall knowledge. Their subjects were able to recognise the meanings of considerably more target words than to produce a translation for the target words. Similarly, Pellicer-Sánchez and Schmitt (2010) examined the learning of four types of word knowledge: recognition of word spelling, recall of grammatical class, and receptive and productive meaning knowledge from reading an authentic text. They found that subjects acquired more receptive word knowledge components (i.e., recognition of form and meaning) than productive knowledge aspects (i.e., recall of word class and recall of word meaning).

Drawing on Nation's (2001) framework, the present study examines three different word knowledge components: spoken form recognition, meaning recall, and meaning recognition using three different vocabulary tests (see section 3.11, for a detailed discussion on the tests developed to assess these word knowledge components). These knowledge types are seen as a primary step in vocabulary learning (Schmitt, 2010; Webb, 2005), they have been chosen because they are considered the main components of vocabulary knowledge that deserve due attention. Although it is desirable to assess all word knowledge facets in a study, it is practically impossible to do so due to the lack of a unified methodology or measures for testing some word aspects like register; the limit of time; and the difficulty of controlling for the potential of cross-test effects (Schmitt, 2010). In this case, Schmitt (2010) advised researchers to focus on certain vocabulary knowledge dimensions.

2.1.2 Receptive and productive vocabulary knowledge

In the field of vocabulary research, the concepts receptive and productive vocabulary knowledge have attracted much interest, as they are seen as one of the ways to distinguish word knowledge types (Nation, 2001). In effect, they are one of the fundamental aspects of the folklore surrounding vocabulary learning (Waring, 1997). Receptive/recognition and productive/recall are the two levels of word mastery. Receptive knowledge is linked to listening and reading and entails recognizing the spoken and the written form of a word while reading, listening and retrieving its meaning. Productive knowledge, on the other hand, is associated with speaking and writing; it involves producing the meaning of a word in spoken and/or written contexts and producing the written or spoken form (Laufer & Goldstein, 2004; Nation, 2001; Schmitt, 2010).

Though the distinction between receptive and productive knowledge seems simple, Read (2000) noted that researchers defined the two terms differently. This in turn has made the comparison amongst studies examining receptive and productive vocabulary knowledge truly troublesome. In this regard, Waring (1999) warned that without a clear definition of these terms, studies cannot meaningfully say much about the development of receptive and productive vocabulary knowledge of L2 learners. Similarly, Schmitt (2010) emphasised that future researchers should clearly delineate what type/level of word knowledge mastery (recognition/receptive or recall/productive) is being addressed in their studies. Accordingly, in the present research, receptive/recognition vocabulary knowledge is defined and measured as; the ability of the learners to recognize the spoken form (i.e. spoken form recognition) and the meaning (i.e. meaning recognition) of the target words, while productive/recall knowledge is used to refer to the ability to recall and produce the meaning of spoken target words (i.e. meaning recall).

It is generally assumed that receptive knowledge is greater and easier to acquire than productive knowledge (Nation, 2001; Pellicer-Sánchez & Schmitt, 2010; Schmitt, 2000; 2010; Webb, 2005) and that receptive knowledge precedes productive knowledge (Melka, 1997). This means that receptive vocabulary knowledge is acquired first and productive knowledge is developed later at a more advanced level (Laufer & Goldstein, 2004). Schmitt (2000) exemplified the assumption that receptive vocabulary knowledge is greater than productive vocabulary knowledge, by referring to a situation that most of L2 speakers may have experienced. The situation is when we come across and recognize a word in a speech or a text, but we are not able to use that lexical item while speaking or writing. To him, this situation demonstrates the existence of different degrees of word knowledge.

Another key issue that vocabulary researchers and scholars have long debated, is whether receptive and productive lie on a 'continuum' or whether they are 'dichotomous' (Schmitt, 2010), to quote Laufer and Goldstein (2004, p. 405) "[there is] no consensus [as] to whether this distinction is dichotomous or whether it constitutes a continuum". On the one hand, Melka (1997) assumed that the distance between receptive and productive knowledge places on a 'continuum' and that receptive knowledge of a word gradually progresses to productive knowledge as one learns more about that word. However, Schmitt (2010) criticised Melka's assumption, arguing that there is still no clear explanation of the threshold between receptive and productive mastery. In other words, the minimum required amount of word knowledge that enables receptive knowledge to transfer to productive level, has not yet been explored. Conversely, for others, like Meara (1997), receptive and productive are two opposite poles, as they represent distinctive types of associational knowledge. Active words are connected to other lexical items, while passive words are only initiated by an external stimulus. According to his view, words do not naturally transfer from a receptive state to a productive state. Nation (2001) criticised this view on the basis that language use is driven by meaning, not association.

Despite this lack of a clear theoretical understanding about the distinction between receptive and productive vocabulary knowledge, the concepts receptive and productive knowledge have been incorporated into different vocabulary knowledge frameworks. For example, Nation (2001) indicated that the concepts receptive and productive knowledge cover all the nine aspects necessary to know a word. So that each of the nine subdivisions of his taxonomy of word knowledge entails receptive and productive knowledge of words. For example, receptive knowledge of a word form like *"unaltered"* involves the recognition of its spoken and written form, whereas productive knowledge of this word entails producing the form in writing or in speech. To put it simply, a learner can master the nine dimensions of vocabulary knowledge receptively and productively. A number of previous incidental vocabulary studies, which have compared the acquisition of receptive and productive vocabulary knowledge dimensions (Brown et al., 2008; Montero Perez et al., 2014; Peters & Webb, 2018; Van Zeeland & Schmitt, 2013; Waring & Takaki, 2003; Webb, 2008; Zhong, 2009), have concluded that receptive word knowledge develops faster than productive knowledge. The present study examines incidental learning and retention of knowledge of meaning receptively (meaning recognition) and productively (meaning recall), along with receptive knowledge of form (spoken form recognition).

2.1.3 Incidental and intentional learning

Fortunately, L2 vocabulary researchers no longer argue about the importance of vocabulary instruction and learning, but instead discuss the most effective and efficient ways for vocabulary teaching and learning. Thus, the best ways of vocabulary learning in the field of L2 learning have been subject of many discussions. Researchers have identified two prominent approaches of vocabulary learning, intentional and incidental (Hulstijn, 2003; Schmitt, 2000). Despite the widely acknowledged importance of the incidental vocabulary approach in vocabulary learning, the term 'incidental learning' is considered problematic, as it lacks clear operational definitions. Hulstijn (2003, p. 357) highlighted this situation stating that incidental learning "has often been rather loosely interpreted in common terms not firmly rooted in a particular theory". Likewise, Wesche and Paribakht (1999) commented that scholars have not agreed on a precise definition of incidental learning.

The use of the concepts of incidental and intentional learning originated in the field of learning psychology, wherein both terms were used as methodological concepts in experiments referring to "the presence or absence of an explicit instruction to learn" (Hulstijn, 2003, p. 354). In the heyday of cognitive psychology in 1970s, interest declined in the concepts of incidental learning and intentional learning as essential accounts in the explanation of human learning. However, these concepts were retained by theorists of information processing and memory as they could be used as suitable means to examine the impact of different kinds of information processing tasks on long-term information retention (Hulstijn, 2003). Craik and Lockhart's (1972) study is considered one of the most influential seminal works, that has encouraged researchers to examine incidental learning with the proposal of "a levels of processing framework". In fact, Craik and Lockhart's study enabled the terms incidental and intentional learning to survive from the 1970s to the current time (Hulstijn, 2003).

Because of the lack of theoretical ground of the concepts incidental and intentional learning, defining them appears to cause confusion. Nonetheless, several authors have strived to describe the processes involved in incidental and intentional learning (Ellis, 1994; Gass, 1999; Huckin &

Coady, 1999; Hulstijn, 2003; Schmidt, 1994). Schmidt (1994, p. 137) described the term incidental learning as “learning without the intent to learn or the learning of one thing (e.g., grammar) when the learner’s primary objective is to do something else (e.g., communicate)”. Later, Hulstijn (2003) derived three more definitions from Schmidt’s statement. The first definition refers to learning without the intention to do so. This definition puts intentional learning in opposition to incidental learning. The second definition is about the learning of one item, while the focus of attention is paid to another aspect. Based on the assumption of this definition, the acquisition of the language aspects is a ‘by-product’ of other tasks. The last more precise definition is about the acquisition of formal features while concentrating on semantic ones. Wesche and Paribakht (1999, p. 176) presented a similar definition of incidental learning. To them, incidental vocabulary learning takes place when learners "are focused on comprehending meaning rather than on explicit goal of learning new words".

The common definition that different scholars agreed upon is that in incidental learning conditions, learners acquire new knowledge or competences as a ‘by-product’ of reading (Huckin & Coady, 1999), oral input (Ellis, 1994), or the focus of the classroom in general (Gass, 1999). In other words, incidental learning takes place as a ‘by-product’ of engagement in different activities such as reading, listening, or writing (Hulstijn, 2003). However, some researchers criticised the conceptualization of incidental learning as a ‘by-product’ of other tasks. For example, Gass (1999) argued that such an interpretation disregards the active role played by the learners in the learning process, as they can place their attention to any part of stimuli input. Gass concluded that the claim that learning was incidental was inaccurate. To put it another way, defining incidental learning as a ‘by-product’ of other tasks implies that conscious processes have no part to play in the learning process. Huckin and Coady (1999) lent support to the idea that incidental learning involves some kind of attention and that the learners have an active role to play in locating their attention to the stimuli. Ellis (1994) took also a related position in advocating the active role played by the learner in the incidental vocabulary learning condition. He argued that although incidental word learning means learning without intending to learn, this does not mean that learning words takes place without the learner noticing the acquired words.

From a pedagogical perspective, Wesche and Paribakht (1999) distinguished between the concepts incidental and intentional in relation to the presence or absence of forewarning of the implementation of a subsequent retention test on given skills or materials. In the incidental learning condition, learners are not informed about the post-test on the learning materials. While in the intentional learning condition, they are warned in advance about the test. On the other hand, intentional learning is seen as kind of learning in which the learner's attention is directed towards the information presented. This type of learning encompasses some explicit tasks geared for practice and learning (Ellis, 1994; Gass, 1999). Hulstijn (2003, p. 349) defined intentional word learning as “the deliberate committing to memory of thousands of words (their meaning, sound, and spelling) and dozens of grammar rules”.

The terms incidental and intentional learning are also distinguished from each other based on the type of attention paid during the learning process. According to Ellis (1994), learners pay two types of attention to vocabulary learning, focal and peripheral. He posits that in the incidental learning mode, the learner's focal attention is focused on the meaning of the message, yet s/he pays peripheral attention to the form. In contrast, in the intentional learning condition, the focal attention of the learner is focused on the language and particularly on attempting to associate between the form and meaning. Schmidt (1994) echoed Ellis’s argument regarding the importance of attention in both different learning conditions, arguing that there needs to be some sort of attention for any learning to take place, regardless of its type.

Although the majority of the issues related to incidental vocabulary learning remain unresolved, one issue that many researchers agreed upon is that incidental vocabulary learning takes place from context. A number of studies in the field of L1 vocabulary acquisition (Huckin & Coady, 1999; Nagy & Herman, 1987; Nagy, Herman, & Anderson, 1985) have affirmed that the bulk of L1 words are acquired incidentally from context. Nagy et al. (1985), for example, found striking evidence for this assumption, as schoolchildren have incidentally acquired a large number of vocabulary items from free reading. In the same way, Nagy and Anderson (1984) reported that L1 speakers incidentally learn about 1000 new vocabulary items every year. The findings reported in L1 studies have encouraged L2 researchers to explore the contribution of the incidental learning approach in vocabulary development (Al-Homoud & Schmitt, 2009; Horst, 2005; Hu, 2013; Laufer & Rozovski-Roitblat, 2011, 2015; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Schmitt, 1998; Waring & Takaki, 2003; Webb, 2008). These studies collectively attested to the advantage of the incidental approach in fostering vocabulary growth; however, most of the vocabulary learning gains reported were relatively small (this will be discussed in detail in the following section).

Despite the fact that incidental vocabulary learning has attracted considerable interest in the field of L2 vocabulary learning, it has also been the subject of a considerable share of criticism. It has been argued that the results that have emerged from this approach are usually ‘negligible’ (Peters et al., 2016). In fact, the learning gains reported in many of the previous incidental studies were relatively small compared to the gains reported in intentional studies (Al-Homoud, 2007). Webb (2008) also attacked the incidental learning approach, arguing that the process of incidental learning is usually slow, which requires multiple numbers of encounters with the target words to develop. In addition, learners with low proficiency levels are likely to incorrectly guess the meanings of the target words (Bensoussan & Laufer, 1984). On the other hand, researchers have identified a number of benefits for incidental vocabulary learning over intentional vocabulary instruction/learning. Huckin and Coady (1999) pointed out that the contextualised nature of the incidental approach helps learners make a richer sense of the target words. A further key benefit of the incidental approach is related to the notion that it is a pedagogically efficient approach, because learners can develop other aspects of the language alongside vocabulary learning (Huckin & Coady, 1999). Nation and Webb (2011) also mentioned that incidental vocabulary learning from context fosters the depth of word knowledge and leads to different degrees of knowing a word.

It is nevertheless argued that both incidental and intentional learning approaches are of equal importance to vocabulary learning. Schmitt (2000) posited that although the significance of incidental learning is widely acknowledged, intentional learning is essential for L2 learners to develop the vocabulary 'threshold' necessary to enable them to benefit from the incidental approach. Similarly, Hulstijn (2003) maintained that both approaches ought to be treated as complementary to each other and both should be implemented in L2 language programmes. The long debate regarding which approach is more efficient and effective remains unresolved. Therefore, it is suggested that a good vocabulary plan should incorporate both approaches to ensure promoting as much vocabulary knowledge facets as possible and teachers should balance opportunities for students to learn from the two approaches (Al-Homoud, 2007).

A variety of factors affect incidental vocabulary learning from context, including the frequency that the target words occur in the text (Laufer & Rozovski-Roitblat, 2011; Nation & Webb, 2011; Schmitt, 2010; Zahar, Cobb, & Spada, 2001). Webb (2007, p. 64) stressed the importance of repetition for vocabulary learning, remarking "learners who encounter an unknown word more times in informative contexts are able to demonstrate significantly larger gains in [different] vocabulary knowledge types than learners who have fewer encounters with an unknown word". Similarly, Nation (2001, p. 76) mentioned that repetition "adds to the quality of knowledge and also to the quantity or strength of this knowledge". A great deal of vocabulary research has provided empirical evidence for this assumption indicating that repeated exposures increase the chances of word acquisition (Brown et al., 2008; Pellicer-Sánchez, 2016; Rodgers, 2013; Vidal, 2011; Waring & Takaki, 2003; Webb, 2007). However, despite the widely held recognition of the importance of the frequency of occurrence, there is no agreement yet on the necessary numbers of repetitions for incidental vocabulary learning to develop, which makes the available results inconsistent and inconclusive. As a result, different studies proposed different numbers of necessary encounters. For instance, Webb (2007) suggested 10 repetitions, Brown et al. (2008) found 7 to 9 repetitions necessary, Pellicer-Sánchez and Schmitt (2010) indicated that the minimum number of repetitions was 10 +, and Waring and Takaki (2003) said that 20 repetitions are necessary. Webb (2007) explained that one of the key reasons for these inconclusive results is the lack of control over the type of context in which the target items were encountered. Because of this inconsistency in the literature, Laufer and Rozovski-Roitblat (2011) concluded that there is no definite number of repetitions needed to ensure word learning, and different aspects of word knowledge may require different numbers of encounters.

In short, it is obvious that there is still no clear understanding of the construct of incidental vocabulary learning (Hennebry et al., 2017). However, in operational terms, what distinguishes the incidental learning approach from the intentional learning approach is whether learners are made aware of the purpose of the activity and the implementation of the subsequent tests after the activity. In this regard, in the current study, incidental vocabulary learning is viewed as exposing subjects to audio-visual inputs (video clips) in different input conditions (VAC, VA, CA, and A only) and asking them to watch or listen for pleasure, which will hopefully lead to vocabulary acquisition and subsequently testing their knowledge of the target words. The learners were not forewarned about the implementation of the subsequent vocabulary tests nor informed about the purpose of the study.

2.1.4 Previous research on L2 incidental vocabulary learning

Incidental vocabulary learning has been examined through different input modalities, including reading (Al-Homoud & Schmitt, 2009; Horst, 2005; Hu, 2013; Laufer & Rozovski-Roitblat, 2011, 2015; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Schmitt, 1998; Waring & Takaki, 2003; Webb, 2008); listening (Brown et al., 2008; Van Zeeland & Schmitt, 2013; Vidal, 2003, 2011); reading while listening (Brown et al., 2008; Webb & Chang, 2015); and viewing audio-visual materials (Montero Perez et al., 2018; Peters & Webb, 2018; Rodgers, 2013; Sydorenko, 2010; Winke, Gass, & Syodorenko, 2010). Most of these previous incidental vocabulary studies have focused on reading, while listening and viewing have been given less attention. For example, Waring and Takaki (2003) asked their subjects to read a graded reader and complete three vocabulary tests: word form recognition, prompted meaning recognition, and unprompted meaning recognition. The researchers found that incidental vocabulary learning is possible through reading but with relatively few gains. Receptive word knowledge (i.e., form recognition and meaning recognition) was developed better than the productive knowledge, meaning recall. Of the 25 target words, only one item was retained three months later. It was also observed that repetition positively affected the learning of the different word-knowledge types.

Pellicer-Sánchez and Schmitt (2010) studied incidental vocabulary learning of form recognition, word class recall, meaning recall and meaning recognition from reading a novel. The findings revealed that the participants made sizable gains from reading the novel. Similar to the findings of Waring and Takaki (2003), Pellicer-Sánchez and Schmitt (2010) reported that the receptive vocabulary knowledge aspects, meaning recognition and spelling recognition, were better improved than the productive knowledge aspects of word class recall and meaning recall. Frequency of occurrence also affected learning, as there was a considerable increase in learning uptakes after repetitions for the four knowledge components. These two studies lend support to the claim that incidental vocabulary growth occurs through reading.

A study that compared the effects of different input conditions was that of Brown et al. (2008). Brown et al. compared incidental vocabulary learning and retention from listening-only mode; reading-only mode; and reading-while-listening-mode, using two vocabulary tests, multiple-choice and meaning translation tests. The results showed that participants were able to increase their knowledge from the three input modes, however the lowest gains were reported in the listening-only mode. Participants also performed better on the multiple-choice test than on the translation test, indicating that receptive knowledge develops faster than productive knowledge. Unlike the results of Waring and Takaki (2003), Brown et al. (2008) found relatively little loss of the target words have taken place over three months. They also found that short-term learning and long-term retention were both affected by the frequency of word encounters, meaning that the more a word was met, the better it was learned and retained.

A similar study by Vidal (2011) looked at incidental vocabulary learning and retention from listening and reading of three academic texts using a modified version of Wesche and Paribakht's (1996) vocabulary knowledge scale (VKS). The results showed that both input modes, reading and listening led to small vocabulary gains, however the reading group significantly outperformed the listening group. In contrast, the listening mode was more effective for the long-term retention than the reading mode. Vidal (2011) also found that frequency of occurrence predicted performance in reading and listening, however its effect on reading was more evident. Small vocabulary gains from listening were also reported in a study by Van Zeeland and Schmitt (2013) who examined incidental vocabulary learning of form recognition, grammar recognition, and meaning recall. The results also demonstrated that much of the acquired knowledge of the three knowledge aspects was not retrievable over the course of two weeks. The effects of frequency of occurrence varied across the three knowledge types with meaning being the least affected. Thus, Van Zeeland and Schmitt (2013) concluded that more than 15 encounters are needed to foster vocabulary learning through listening.

Recently, the potential of audio-visual input as a source for incidental word acquisition started to attract considerable attention of a number of researchers. For example, Rodgers (2013) asked his participants to view multiple episodes of a TV programme and tested their knowledge of form and meaning relationship through two multiple-choice meaning tests at two levels of difficulty (tough and sensitive tests). The tough test had distractors that were very similar to the correct answer with regards to the form and meaning, while the sensitive test had distractors that were different in form and meaning from the correct answer. The findings indicated that the experimental group made significant, large gains from viewing the TV programme. On average, the participants made a gain of 6.4 words on the tough test and 6.8 items on the sensitive test. A medium-sized correlation between word learning on the tough test and repetition was also found.

A recent study by Peters and Webb (2018) looked at incidental vocabulary learning through watching a lengthy TV programme. In two experiments, Peters and Webb measured three types of word knowledge: form recognition, meaning recall, and meaning recognition. The findings indicated that incidental vocabulary learning occurred through audio-visual input and that receptive and productive knowledge of meaning was significantly affected. Knowledge of form was not significantly improved because of the effect of the pre-test, according to Peters and Webb (2018). The results also revealed a positive relationship between repetition and vocabulary learning.

Overall, as the aforementioned studies prove, incidental vocabulary learning is possible through different input modes and frequency of occurrence is one of the key determinants of learning. What can be inferred from these studies is that incidental vocabulary gains are usually low and are the cause of steady attrition. In addition, the amounts of gains differed across the different word-knowledge types and by the different input modes. It can be summarised that the odds of learning a word incidentally increase when the word is met more often.

2.1.5 Section summary

To sum up, vocabulary learning appears to be more complicated than simply learning its form and meaning as vocabulary knowledge entails several aspects which are not acquired at the same time. Research into vocabulary learning has shown that vocabulary items can be acquired through basically two main approaches, intentional and incidental. The present study focuses on incidental learning and retention of three different aspects of word knowledge and two levels of mastery, receptive and productive through different conditions of audio-visual materials. The following section presents the theoretical framework that will serve to guide the present study.

2.2 Theoretical background

It is important to investigate the differential effects of multiple modalities of input on incidental vocabulary learning and retention within an appropriate theoretical framework. This section provides the theoretical underpinnings that guide the present study. It opens with a discussion of the dual coding theory of Paivio (1986) then continues to introduce of the cognitive theory of multimedia learning by Mayer (2005). This is followed by an overview of working memory model of Baddeley and Hitch (1974). A detailed account of Wen's (2016b, 2019) model of working memory is also provided. Finally, a rationale for drawing on these theories to understand the effects and value of different input modalities in incidental vocabulary learning is provided.

2.2.1 The dual coding theory

Psychological approaches play an essential role in developing a sound understanding of science and the practice of education. Over the last five decades, psychologists have sought to identify the psychological mechanisms that underlie human behaviour and experience (Clark & Paivio, 1991). This effort has resulted in the birth of several theories and models that account for the cognitive processes involved in language learning. One of the well-established theories in this regard is the dual coding theory (DCT) of Paivio (1986, 1991). DCT is a general theory of cognition that was developed to provide equal significance to linguistic and non-linguistic representations. It is made up of some basic structural and functional assumptions (see Figure 2.1). One of the main assumptions of DCT is that memory and cognition are served by two separate symbolic systems; a verbal system that is specialized in dealing with verbal information and a nonverbal system – imaginal – which is responsible for processing nonverbal information (Paivio, 1986, 2013; Sadoski & Paivio, 2004). Together, these two cognitive systems process information of language and information of the world.

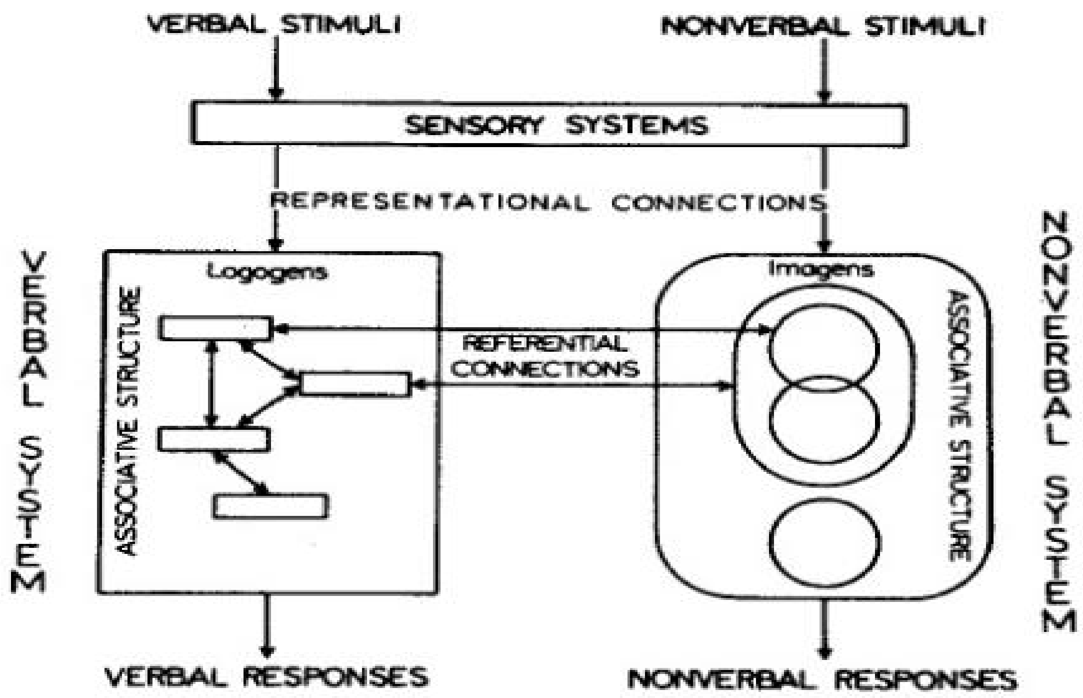


Figure 2.1. Dual coding theory. Adopted from Mental Representations: A Dual Coding Approach by Paivio (1986, p. 67)

The verbal and nonverbal systems have unique structures and functions. They are functionally independent of each other in the sense that "one system can be active without the other or both can be active in parallel" (Paivio, 1991, p. 259). However, Paivio (1986) assumed that the activation of both systems can have additive effects on recall. The two systems are interconnected in the sense that activity in one representational unit (logogen/imagen) can extend to the other unit (Paivio, 1991). These two systems have two levels of the interconnection. The first level is the referential interconnection that refers to the link between the two representational units, imagens and logogens. A good example for this interconnection is naming objects or pointing to named objects. The other level of interconnection is the associative interconnection which occurs within the representational units in the two systems (verbal/imaginal), so that, a logogen can activate other logogens such as in a word association task, and an imagen can also activate other imagens (Paivio, 1986).

The verbal and nonverbal systems have their own representational units and hierarchical organizations (Sadoski & Paivio, 2004). The representational units in the verbal system and nonverbal system are logogens and imagens respectively. Sadoski and Paivio (2004, p. 7) defined a logogen as "anything learned as a unit of language in some sense modality". While, the term imagen is used to refer to “representations from which mental images are generated” (Paivio, 1986, p. 59). These representational units are modality-specific in the sense that they correspond to relevant modalities of the presented materials, and are activated in different ways (Paivio, 1991). The direct activation of logogens can occur via sensory input like seeing a text. Imagens, on the other hand, can be directly activated by sensory input such as seeing objects or pictures. There can also be indirect activation of the two units either through forming images for words or naming objects (Sadoski & Paivio, 2004). The sizes of the representational units are also distinct. This means that logogens could represent letters, words, sentences, and imagens could respond to an object, a segment of that object, or a series of objects (Paivio, 1986). Furthermore, the two systems differ in their organization. Imagens are organised into nested sets. That is, smaller objects, pictures are embedded in larger ones. (Paivio, 1991; Sadoski & Paivio, 2004). For instance, a human face is composed of several components, such as nose, eyes, mouth, among others and these components consist of even smaller parts like nostril and iris. Logogens, on the other hand, are structured in a sequential fashion. That is, smaller units (e.g., letters) are organised into larger units (e.g., words/phrases) (Paivio, 1986).

Another assumption of the DCT is that there are three levels of processing that occur between or within the verbal and nonverbal systems. (Paivio, 1986, 1991, 2013; Sadoski, 2005; Sadoski & Paivio, 2004). The first level is representational processing, which concerns the activation of either of the representational units, logogens or imagens by a stimulus. An example of this level of processing is the simple recognition of verbal (e.g., book, car) or/and nonverbal/imaginal stimuli. The second type of processing is the referential processing that refers to the activation that occurs across-systems, meaning words activate images, and images activate words. The final level of processing is associative processing, which relates to the activation within each representational unit by other units within the verbal and nonverbal systems. Paivio (1991) gave an example of a task that requires the activation of these three levels of processing. The example is when the word '*dog'* is used in a task, the simple recognition of this word requires the representational processing that activates the corresponding logogen. At the associative level, the word '*dog'* is associated with the verbal associate '*cat'*, and at the referential processing level, images of '*cat'* and '*dog'* are activated.

In the present study, DCT is chosen as a solid framework to provide a firm explanation of the findings of this research for the following reasons. Firstly, this theory provides a detailed account of the effects of verbal and nonverbal information on memory and recall. Its emphasis on the structures and functions of its modality-specific subsystems makes it a unique theory. Since the purpose of this study is to investigate L2 incidental vocabulary learning and retention by exposing the learners to different combinations of verbal and visual information in different audio-visual input conditions, DCT is an appropriate theory to explain the mental processes involved in dealing with the given information and how the use of verbal and nonverbal information can affect memory and recall. DCT is also appropriate in explaining the additive effect of pictures on recall of the target lexical items. Research has indicated that learners easily acquired and retained words that were associated with pictures (Clark & Paivio, 1991), because having the verbal and nonverbal memory codes to represent a word increases the odds of recalling it (Reed, 2006). Therefore, it is expected that learners involved in the visual input conditions (VAC and VA) would learn the target words better and retain them longer than those in the non-visual conditions (CA and A only).

A further rationale underlying my interest in using the DCT relates to the assumptions of WM. Seeing as WM is an aspect to be investigated in this study, it is essential to use a theory that maps the assumptions of WM. DCT and WM are closely linked, DCT shares a similar conception of the WM model of Baddeley (1986) in the sense that their basic premise is the existence of two modality-specific systems, which are responsible for different types of information. DCT postulates that information is processed by verbal and non-verbal systems, which are analogous to Baddeley's WM modality-specific systems, the phonological loop and the visuospatial sketchpad. In addition, both DCT and WM theories assume that information is processed separately by the respective system before being fused. The phonological loop deals with auditory information, while the visuospatial sketchpad is responsible for pictorial information. According to DCT, dual coding of information allows verbal and pictorial information, which are temporarily represented in the distinct WM systems (the phonological loop and visuospatial sketchpad), to be referentially connected. Moreover, WM plays a core role in information processing as its limited capacity impedes the additive effect of dual processing of information on recall and learning.

An additional reason for drawing upon DCT is that it distinguishes between abstract language and concrete language. Sadoski (2005) described abstract language as a language that has no direct access to the pictorial cognitive system (for example, the word *fact*), while concrete language (e.g., *book*) is seen as the language that can have direct access to the nonverbal systems because it has sensory referents. Abstract language is processed by the verbal system, whereas concrete language is processed by both channels (verbal/imaginal) (Paivio, 1991; Sadoski, 2005). Hence, concrete language is better remembered, as it is processed and represented in the two systems. The advantage of concrete concepts is that one can form a mental image of their referents (Sadoski, McTigue, & Paivio, 2012). Since the present study uses concrete target words, DCT is thus useful in explaining the effects of dual coding of the concrete target words on retrieval. Finally, DCT features predominantly in studies that have examined the effects of different modalities of input on language learning (Al Seghayer, 2001; Chun & Plass, 1996, a full report on these studies is provided in section 2.4.1). The findings of those studies lend support to the DCT.

Those studies have confirmed that DCT provided useful explanations for a variety of memory effects, namely the superiority effect of images and the concreteness effect.

2.2.2 The cognitive theory of multimedia learning

The cognitive theory of multimedia learning (CTML) developed by Mayer (2005), seeks to provide a cognitive account of how individuals learn from multimedia environments. In his theory, Mayer (2005, p. 31) attempted to explain how people learn from words and pictures. One of his goals in developing this theory was to help teachers and material designers to develop multimedia instructional messages in light of how “the human mind works”. Mayer (2014) based the CTML on three cognitive theories: the generative theory of Wittrock (1974); the dual coding theory of Paivio (1986); and Sweller’s (1994) cognitive load theory. However, Mayer (2005) took his theory a step further to view the learner as an active knowledge constructor, because he/she integrates multiple pieces of information (Gyselinck, Jamet, & Dubois, 2008). In this regard, Mayer’s (2014) CTML is based on three assumptions derived from cognitive science: the dual channel, the limited capacity of WM, and active processing. The dual processing channel assumption proposes that humans process verbal and nonverbal information through two independent corresponding processing systems. The verbal channel deals with verbal information received via the ears, and the visual channel is responsible for visual information received through the eyes. The second assumption of CTML is limited capacity, which concerns the limit of WM capacity in which learners can only process and store a limited amount of verbal and visual information. Moreno and Mayer (2002) argue that the learner’s visual and verbal processing channels can be overloaded when too many elements to be processed are presented simultaneously. The third assumption (i.e., active processing) relates to the cognitive process of building mental connections between verbal and visual information and integrating them with existing knowledge in long-term memory.

Mayer (2005) argued that the core tenet of his theory is that meaningful learning takes place when learners build connections between visual and verbal information and process them actively in relevant processing channels, before integrating them within existing knowledge. As such, meaningful learning in a multimedia environment involves several cognitive processes (Mayer, 2014; Mayer & Moreno, 2002). Firstly, learners are given verbal and visual information (i.e. multimedia presentation), received through their eyes and ears. They then need to hold the received information in the relevant WM modality-specific systems. After that, learners organise the verbal and nonverbal information into coherent verbal and visual representations in WM. Finally, they create referential connections between the verbal and visual information before integrating them with corresponding prior knowledge. The CTML has received a large amount of support from a wide range of studies that have examined the effects of combinations of words and pictures on different learning domains (for example, Acha, 2009; Akbulut, 2007; Al Seghayer, 2001; Austin, 2009; Chun & Plass, 1996; Doolittle & Altstaedter, 2009; Gyselinck, Ehrlich, Cornoldi, De Beni, & Dubois, 2000).

Following from the basic cognitive assumptions of CTML, Mayer and his associates (Mayer, 2003; Mayer & Johnson, 2008; Mayer & Moreno, 2002) developed twelve instructional principles that ultimately yielded a better understanding of how to best develop multimedia learning materials and conditions. To provide insightful explanations for some of its findings, the present study draws on two of those influential principles, namely the multimedia principle and redundancy principle, both widely supported by empirical research (Sorden, 2012). The rationale behind using the multimedia principle in this study is related to the point that this principle is appropriate in explaining the superiority of the multimodal input conditions (VAC, VA, and CA) over the single mode input condition (A only) in learning more words. The redundancy principle is also suitable to explain the reasons for differential effects of the visual groups (VAC and VA) on incidental vocabulary learning. The next sections provide brief discussions on these two principles with regards to their implications on vocabulary learning through different audio-visual learning conditions.

2.2.2.1 The multimedia principle.

The multimedia principle assumes that learning through a combined presentation of words and pictures is more effective than learning through a single presentation mode (i.e., words only) (Mayer, 2005). This is because the presence of two sources of information, (i.e. pictures and words in the same input condition) would allow learners to establish a mental connection between the pictorial and verbal representations in WM, paving the way for a more effective recall and learning of the presented information (Mayer, 2014). In other words, learners may learn better if they are exposed to text coupled with pictures than when they are presented with words only. The multimedia principle has received a wide support from a number of past studies, which examined the effects of combinations of texts and visuals on various types of learning (for example, Akbulut, 2007; Mayer, Dow, & Mayer, 2003; Mayer & Moreno, 1998). Akbulut (2007), for example, compared the differential effects of three different gloss types: (a) text only; (b) text accompanied by static pictures; and (c) text coupled with a video clip on incidental vocabulary learning and reading comprehension of Turkish EFL learners. The researcher indicated that the gloss conditions that had text coupled with the static pictures or video clip were more effective for incidental vocabulary learning than the single gloss condition. Since the present study employs multimodal verbal and visual input conditions (VAC, VA, CA) and a single input condition (A only) to present the target words, it is thus expected that learners in the multimodal conditions would outperform those in the single input condition in the three vocabulary measures.

2.2.2.2 The redundancy principle

The redundancy principle refers to situations in which learners have to divide their attention between three types of stimuli, visual images, captions (written text) and audio. Mayer (2014, p. 247) noted that redundancy takes place in situations “when the same information is presented concurrently in multiple forms or is unnecessarily elaborated”. Rather than having positive impacts on learning, the redundancy principle concludes that redundant information has a detrimental effect on learning, as it may cause distraction (Mayer, 2005). For example, in the present study, adding printed text (on-screen captions) to video with audio (as is the case in the VAC condition) learners may direct their attention to the captions and thus pay less attention to the audio and the visual images in the video. Since the audio/narration (verbatim text) and the on-screen text are identical, eliminating captions may result in better performance (Mayer, Heiser, & Lonn, 2001). A range of previous studies have found that learning has been negatively affected by the presence of redundant information, as they caused extraneous processing (Mayer et al., 2001; Mayer & Moreno, 2002). For instance, Mayer and Moreno (2002) compared the performance of two groups on retention and transfer tests. The first group watched animation with audio, while the other group received animation, narration and on-screen text (caption). The first group who saw the animation accompanied by the narration outperformed the group who had the three types of stimuli on the subsequent retention and transfer questions. The authors concluded that the redundant information (captions) was distracting, which thus impeded learning. Mayer and Johnson (2008, p. 2) contended that redundancy hinders learning, because redundant information “[…] creates extraneous processing – such as inducing the learner to visually scan between the caption at the bottom of the screen and the graphic and to try to mentally reconcile the incoming spoken and verbal stream”. Similarly, Mayer and Chandler (2001) found that a deeper understanding of the content was enhanced by materials without redundancy. Because of limited WM capacity, extraneous processing would make the learner less capable of processing the necessary information for learning (Mayer & Johnson, 2008). In the present study, the multimodal condition (VAC) included visual images, audio, and on-screen text (captions). Thus, according to the redundancy principle, it is assumed that participants involved in conditions like this (VAC) would not perform as well as their counterparts who would not be subject to conditions of redundancy (i.e., the VA condition), presumably, because the redundant information (captions) may cause distraction and extraneous processing.

2.2.3 The multicomponent model of working memory

Human beings are biologically endowed with special mechanisms that enable them to acquire/learn languages. Working memory (WM) is one of the key cognitive processes on which learning and use of language draws (Juffs & Harrington, 2011), as it enables us to store, encode, and retrieve information (Dehn, 2008). Psychologists have developed several theoretical accounts of WM that are distinct from each other in their nature, structure, and function (Baddeley & Hitch, 1974; Cowan, 1999; Engle & Kane, 2004). However, in recent decades, Baddeley and Hitch (1974) WM model has emerged as a dominant and widely accepted model of memory. It remains the most explicitly articulated and influential model of WM that has been advanced to account for the temporary storing and processing of information whilst doing other cognitive tasks, such as reasoning, problem solving, and learning (Dehn, 2008). Baddeley and Hitch’s (1974) WM model provides a sound framework for guiding an empirical investigation of different complex cognitive tasks. In particular, it explains the different memory processes involved in daily tasks as well as demanding activities that require considerable effort and thinking.

Baddeley and Hitch’s (1974) tripartite WM model is used as a theoretical framework for the present research study for the following reasons; firstly, in recent years, this model has become the crucial explanation for the role of memory in learning and has attracted considerable attention, its different components are supported by an abundance of research. The second reason for using this model is its unified structure, which makes it easier to compare the relationship between its different systems (e.g., the phonological loop and the visuospatial sketchpad in this study) and vocabulary learning. Another important reason for adopting this WM model is related to the fact that this WM model was found very useful in explaining the involvement of the two modality-specific systems, the phonological loop and the visuospatial sketchpad in learning through words and pictures. The original WM model of Baddeley and Hitch (1974) has been proposed, according to Baddeley (2007), to highlight the functional role of WM, instead of only its storage capacity.

Various definitions of WM have been proposed in the field of memory research. For instance, Baddeley (2007, p. 1) defined WM as “a temporary storage system under attentional control underpins our capacity for complex thought”. Similarly, Alloway (2006, p. 134) viewed the construct WM as "a mental workspace that can be flexibly used to support everyday cognitive activities that require both processing and storage such as, mental arithmetic". Juffs and Harrington (2011, p. 138) did not conceive WM as “memory PER SE, if by that term we simply mean the capacity to store the products of our experience in the world. Rather, it is better understood as a system that controls and regulates behaviour”. Alloway, Gathercole, and

Pickering (2006, p. 1698) argued that WM "is not a single store, but a memory system comprised of separable interacting components". The present study reserves the term WM to refer to the learners' capacity to simultaneously store and manipulate information (i.e. vocabulary items) for a short period of time. The learners' WM capacity is assessed by four simple and complex tasks, forward digit recall, backward digit recall, dot matrix, and the odd-one-out.

One of the prominent features of WM is the limited capacity of its different systems, which means that only a limited amount of information can be concurrently manipulated in those systems. The notion of WM limitations is a contentious issue, as there is no unified definition of this limited capacity. However, several attempts have been made to identify the range of the capacity of WM. Baddeley (1986) proposed that the capacity of WM can process approximately seven items of information at once. Therefore, if any specific task imposes demands that exceed the available limited amount of cognitive resources, maintenance and processing of information will be impeded. However, the term *'item'* is not clearly defined and it is a controversial issue. Individual differences in WM capacity are one of the main determinants for variations in learning outcomes and therefore have important consequences on learners' abilities to develop new skills (Alloway , Gathercole, Kirkwood, & Elliott, 2008; Kozan et al., 2015; Lusk et al., 2009).

Alloway (2006) gave the example of mental arithmetic to illustrate how WM functions in everyday activities. Mental arithmetic involves attempting to multiply two figures (e.g., 28*38) without the help of any other resources, as a calculator. WM comes into play in this task in the following ways. Firstly, it tries to store the two numbers temporarily. Secondly, it retrieves the previously learned rules of multiplication, recalls the outcomes of intermediary calculations, and adds the intermediary calculations to work out the correct solution. Without WM, this simple mental arithmetic cannot be successfully achieved. Alloway (2006) contended that since WM has limited capacity, any internal or external distraction can cause the calculation process to fail.

The following provides a discussion of the distinction between short-term memory and WM. This is followed by an overview of the different components of WM. Next, studies that have examined the link between WM and vocabulary learning are reviewed. Finally, the study examines the role of individual differences in WM capacity in learning from different modalities of input.

2.2.3.1 The distinction between working memory and short-term memory

In modern theories of memory and cognition, WM and short-term memory are deemed key constructs. Though the terms WM and short-term memory are sometimes used interchangeably in the literature, they differ in several fundamental ways. Firstly, short-term memory refers to the ability to hold information for a short period and is conceptualised as a passive temporary store of information (Cowan, 2008). A perfect example of a task that relies on short-term memory is remembering a mobile number. On the other hand, WM is more complicated than short-term memory in the way that it involves both the storage and the processing of information (Alloway et al., 2006). The term WM originated from the concept of short-term memory (Baddeley, 2012). Therefore, the concept of short-term memory is replaced by the WM concept, which is more dynamic as it encompasses both storage and processing abilities (Unsworth & Engle, 2007). Gathercole and Alloway (2008) reserved the term short-term memory to refer to the slave subsystems of WM, namely the phonological loop and visuospatial sketchpad on the one hand. On the other hand, they used the WM concept to describe active processing subsets, such as articulatory rehearsal and inner scribe.

Besides the aforementioned theoretical conceptions, the constructs short-term memory and WM are distinguished from one another with relation to their measures as well as the relationships of their measures with other cognitive activities. On the one hand, some researchers argued that short-term memory is tapped by tasks that impose storage demands (Alloway & Alloway, 2010). Short-term memory tasks, which required subjects to remember a list of words, digits, nonwords, or shapes in the correct presentation order (Unsworth & Engle, 2007), were found to correlate highly with cognitive tasks, such as vocabulary learning/development (Gathercole & Baddeley, 1990; Gupta, 2003; Martin & Ellis, 2012). On the other hand, some researchers maintained that WM is measured by tasks that engage participants in storing and processing activities (Juffs & Harrington, 2011). One of the most widely known and oft-used WM tasks is the reading span task of Daneman and Carpenter (1980). The reading span task gauges a subject’s ability to read a series of sentences, judge the veracity of the sentences, and remember the last word of each sentence in the correct order of presentation. Several researchers have documented that WM tasks correlate strongly with higher-level cognitive abilities, such as reading, reasoning, and numeracy (Alloway & Alloway, 2010; Daneman & Carpenter, 1980). Daneman and Carpenter (1980), for instance, found that their reading span associated more strongly with different reading comprehension measures than did the word span task (which is considered a short-term memory task). Taken together, WM tasks and short-term memory tasks theoretically tap the storage abilities of the participants, however WM tasks impose extra processing demands (Unsworth & Engle, 2007). Nevertheless, it is worth mentioning that there is an opposing view regarding the theoretical classifications of tasks as either short-term memory tasks or WM tasks, as discrepant findings have emerged regarding the relationship between different memory tasks and higher order cognitive tasks. For example, a study by Bayliss, Jarrold, Baddeley, and Gunn (2005) reported that simple tasks (i.e., digit span and Corsi span) that involve storage activity were reliable predictors of complex learning tasks, such as reading and mathematics. Hence, Kormos and Sáfár (2008) argued such evidence indicates that complex and simple memory tasks tap the same underlying constructs.

To avoid the apparent paradox in classifying memory tasks as either short-term memory tasks or WM tasks, I used theoretically neutral labels to classify memory tasks as either simple or complex based on the demands imposed by the task. In other words, this study categorised the four memory tasks (forward digit recall, backward digit recall, dot matrix, and the odd one out) as either simple memory tasks or complex memory tasks based on the activities required by the task. That is, the term simple memory task is used to describe the tasks that involve storage activity, while the term complex task is used to refer to the tasks that impose both storage and processing demands. This is because the simple (i.e., forward digit and dot matrix) and complex (i.e., backward digit and odd-one-out) tasks are assumed to tap the same cognitive systems of Baddeley's model of WM (the phonological loop or visuospatial sketchpad, respectively), but differ with regards to the extent of the processing demands they impose. Unsworth and Engle (2006, p. 77) concluded "simple and complex span tasks should not be dichotomized to simply reflect working memory and short-term memory". According to Bayliss et al. (2005), the reason for the similar correlations between simple and complex tasks with different cognitive tasks is that the key determinant of performance on simple and complex tasks, is storage capacity .

To summarize, though the existing evidence attested that both WM and short-term memory are two different constructs, there is evidence indicating that WM and short-term memory tasks index very similar processes and tap the same underlying systems (Juffs & Harrington, 2011; Unsworth & Engle, 2007). In addition, the findings regarding the correlations between complex and simple memory tasks and other cognitive tasks are mixed (Bayliss et al., 2005), suggesting that the two types of tasks may be underpinned by the same cognitive system.

2.2.3.2 Components of working memory

The original modular model of WM (Baddeley & Hitch, 1974) is composed of three functionally independent but interconnected subcomponents. As displayed in Figure 2.2, this multifaceted model comprises a temporary auditory storage subsystem (the phonological loop), which is claimed to be responsible for dealing with verbal-acoustic inputs. A parallel visuospatial subsystem (the visuospatial sketchpad) is in charge of maintaining visual and spatial information. The real brain of the WM system is the central executive, which is proposed as an attentional system. The central executive allocates and divides attention between the other subsidiary systems, phonological loop and visuospatial sketchpad (Baddeley & Hitch, 1974). In the revised version of the WM model, Baddeley (1986) added a new component, the episodic buffer, which is known as a multidimensional store. The central executive also controls this store. The main role of the episodic buffer is to integrate information from WM with information already in the long-term memory. The different subsets of WM have been found to be dissociable in children and adults. For example, in studies on children, Jarvis and Gathercole (2003) and Alloway et al. (2006) documented a dissociation between verbal and visuospatial components of WM. Engle and Kane (2004) also revealed that verbal and visuospatial WM systems were separable in adults. The following subsections provide a detailed discussion of the three components of WM, phonological loop, visuospatial sketchpad, and central executive, as they are relevant to this investigation.

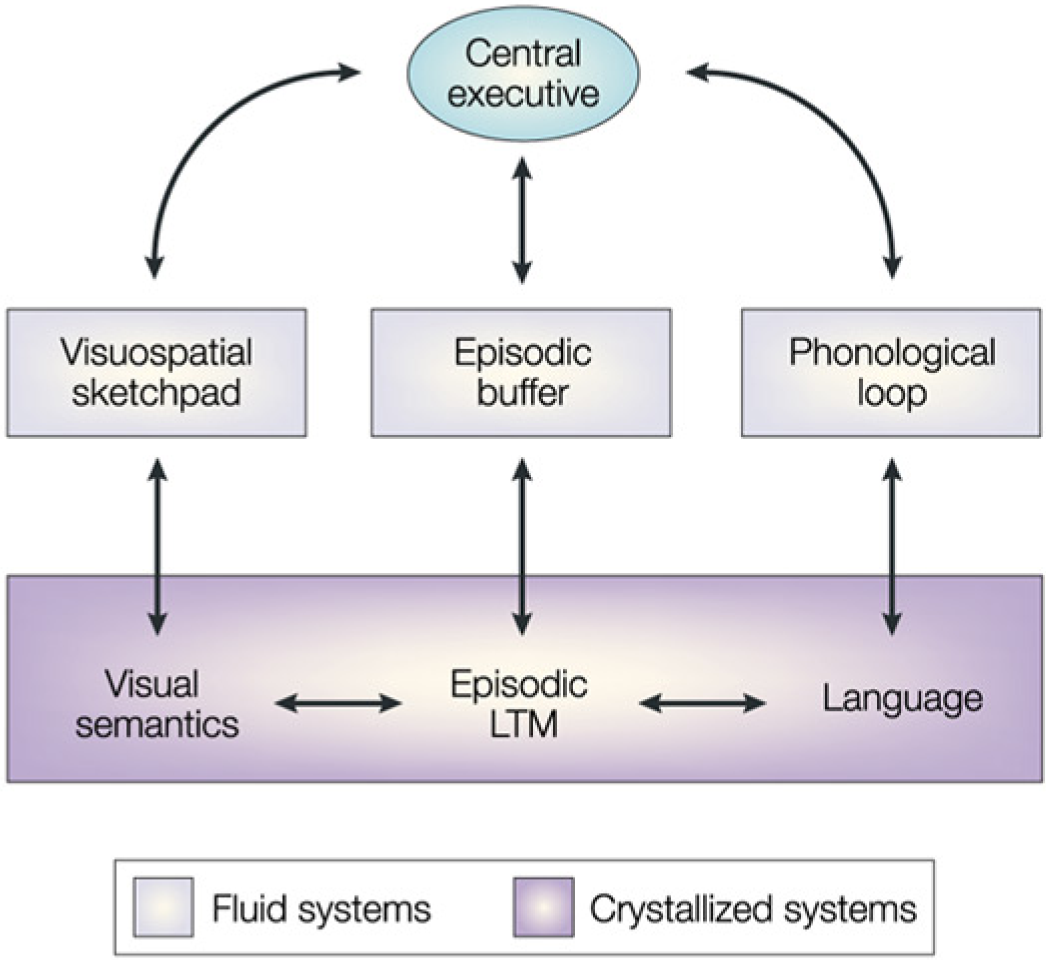


Figure 2.2. Working memory model (adopted from Baddeley, 2003, p. 196)

2.2.3.3 The phonological loop

The phonological loop is of great importance for cognitive, educational, and social development. In the four decades since the publication of Baddeley and Hitch (1974) WM model, this subsystem (phonological loop) has received considerable attention and is by far the best understood system of WM. This system is responsible for storing speech-oriented or auditory information (e.g. words, as in the case of the present study) (Baddeley & Hitch, 1974). The phonological loop is thought to be closely related and linked to language development in general, and vocabulary learning, in particular. Dehn (2008) hypothesised that the phonological loop is developed to facilitate language learning above all. Baddeley, Gathercole, and Papagno (1998) viewed this system as a dual-functioning unit that comprises two further sub-components. The first subcomponent is the phonological store that retains materials in phonological forms. Information in this store fades rapidly after around two seconds unless refreshed by the second subcomponent, the articulatory rehearsal mechanism (Baddeley, 1997; Baddeley, 2003). The articulatory rehearsal mechanism serves to maintain and refresh information held in the phonological store in a way of inner speech or subvocal rehearsal in order to prevent the rapid decay of information. By rehearsing information, an individual can retain it over a longer period of time (Baddeley, 2003; Dehn, 2008).

There is an extensive body of research advocating the basic features of the phonological loop system. The phonological similarity effect is the first piece of evidence that supports the idea of the phonological loop (Lobley, Baddeley, & Gathercole, 2005). It has been found that words and letters that sound similar (e.g., sat, cat, map, bad) are more difficult to recall than words that do not (e.g., pen, lip, car, key) (Dehn, 2008). For example, a study by Conrad and Hull (1964) reported that a recall of lists of letters that all rhymed was more difficult than recalling non-rhyming lists. The study's findings concluded that phonological similarity affected the performance of the verbal/phonological WM. The reason for such results is that storing similar sounding items makes the distinction between memory traces confusing, leading to poor performance on the memory tasks.

Furthermore, another piece of evidence supporting the existence of the two subcomponents of the phonological loop comes from research on the word-length effect. Several studies found that one-syllable words (e.g., hand, wash, cat, car) are better recalled than longer words (e.g., computer, reservoir, auditorium), as one can utter them more rapidly. For instance, Baddeley, Thomson, and Buchanan (1975) used a test of phonological WM in which participants had to remember lists of five words in the correct order. The length of the words varied from one to five syllables. The study found that the longer the words, the harder they were to recall. Baddeley et al. (1975) maintained that word length affects recall, because long words take a longer time for verbal rehearsal. The articulatory suppression serves as additional evidence supporting the assumption of the existence of the phonological loop and to understand the mechanisms underlying it (Baddeley, 1997). The articulatory suppression task refers to situations when participants are asked to repeat an irrelevant word repeatedly whilst being engaged in a remembering task. It has been reported that the performance of subjects on digit span tasks was negatively affected by articulatory suppression (Baddeley, 1997).

The contribution of the phonological loop in vocabulary learning has been studied extensively. A large body of evidence from correlational and experimental studies has shown that the phonological loop is closely related to vocabulary acquisition in both L1 and L2 of normal and cognitively impaired children and adults (Alloway, Williams, Jones, & Cochrane, 2014; Gathercole & Baddeley, 1990; Gupta, 2003; Kaushanskaya, Blumenfeld, & Marian, 2011; Kormos & Sáfár, 2008; Masoura & Gathercole, 2005; Papagno & Vallar, 1995; Service, 1992; Speciale, Ellis, & Bywater, 2004). The general findings of these studies lend support to Baddeley's (2003) theoretical claim that the phonological loop is a language learning device. In light of the widely perceived importance of the verbal WM system, the capacity of the learners' phonological loop in this study is thus measured to determine whether the capacity of this component could be related to vocabulary development and to find out whether the individual differences in the capacity of this system could mediate the effects of the different modalities of audio-visual input.

2.2.3.4 Visuospatial sketchpad

The visuospatial sketchpad is the second subcomponent of Baddeley and Hitch (1974), which is equivalent to the phonological loop. It serves to maintain spatial and visual information and has a capacity for three to four objects (Baddeley, 2003). It comes into play whenever pictorial and spatial information needs maintenance (Dehn, 2008). This system is far less clear than the phonological loop and has received less attention since the introduction of Baddeley and Hitch (1974). Although this subordinate system has received less attention than the phonological loop, there is generally an agreement that this component can be broken down into two further components: the visual and the spatial (Logie, 1995).

The visual store is thought to be a passive store like the phonological store of the phonological loop. This store deals with static visual information, such as shapes and colours. While the spatial store, or known as the inner scribe, is an active repository that is responsible for a rehearsal process. The responsibility of this sub-component is to refresh information maintained in the visual store in order to eliminate it from rapid decay. The spatial store (inner scribe) is also in charge of the manipulation of dynamic information (Logie, 1995). The what and where organization in the visual system is mirrored by this breakdown of visual and spatial subcomponents, which is supported by several studies, suggesting dissociations between memory for visual patterns and memory for spatial movements (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Logie, Zucco, & Baddeley, 1990). The study by Della Sala et al. (1999) that examined patients with brain damage reported that some patients showed considerable challenges in processing visual resources, whereas others had difficulty in dealing with spatial information. Since the verbal system and the visuospatial sketchpad system of WM have been found to be dissociable in children as well as in adults (Alloway et al., 2006) and have different contributions to language learning (Jarvis & Gathercole, 2003), it is thus important to include the visuospatial sketchpad system in the present study to determine its role in learning through different input modalities which involve visual stimuli.

2.2.3.5 The central executive

Although the WM model of Baddeley (1986) has attracted considerable attention during the last few decades, the different components of this model have received markedly differing interest. Baddeley (1996) pointed out that the phonological loop has attained the greatest prominence amongst the other components, while the central executive lagged far behind it. In the original model of Baddeley and Hitch (1974), the exact nature of this key system was left unspecified (Baddeley, 2007). Nevertheless, scholars and researchers unanimously agreed that the central executive has a primary role to play in WM. Some researchers attribute individual differences in WM fundamentally to the central executive process and such variations are the main reason for differences in learning outcomes (Daneman & Carpenter, 1980; Dehn, 2008). Baddeley (1996) described the central executive as a complex system that comes into play when information needs to be manipulated and processed. That is, the central executive is regarded as a processing system that is usually involved in higher-order cognitive tasks (Baddeley, 1996). Furthermore, unlike the other systems of WM (the phonological loop and the visuospatial sketchpad), the central executive is domain-free that means it deals with both auditory and visual-spatial information (Baddeley, 1986).

In their initial conceptualisation, Baddeley and Hitch (1974) assigned several functions to the central executive, including capacity for storage; interaction with the long-term memory; and coordination and allocation of incoming information between the other slave systems of the WM. However, Baddeley (1986) thought that such roles were not fully adequate and needed to be clearly defined. Thus, Baddeley sought to comprehensively specify the executive functioning performed by the central executive by adopting the Supervisory Attention System (SAS) of Norman and Shallice (1986, as cited in Baddeley, 1996). In this way, Baddeley specified a range of new roles to the central executive, including 'selective attention' which refers to the capacity to focus on relevant tasks and suppress irrelevant ones, 'switching attention' which relates to the ability to coordinate different concurrent tasks, and 'allocating resources' to the other two systems of the WM (the phonological loop and visuospatial sketchpad). In addition, Baddeley (2007) added that the central executive is not responsible for linking the other components of WM with long-term memory, however this function has been assigned to the newly added component, labelled as the episodic buffer. More importantly, the central executive has the capacity to determine how to share cognitive resources and how to inhibit the disruptive effects of irrelevant information (Baddeley, 1996). However, since the central executive has many essential roles to play, Baddeley (2003) suggested that the central executive should be fractionated into specific executive functions in order to account for those several functions. To conclude, the central executive of the WM is regarded as a broad attentional control space, which is important for the processing of information during cognitive tasks.

2.2.3.6 The implication of working memory in vocabulary learning

Since learning a language is complex process that draws on different cognitive mechanisms, including WM, this cognitive mechanism has been the subject of a considerable number of studies. Educational and psychological studies sought to explore why there are individual differences in learning in different learning domains. WM has been found to be one of the main sources of individual differences in reading comprehension (Alloway, 2007; Alloway & Alloway, 2010; Berninger, Raskind, Richards, Abbott, & Stock, 2008; Daneman & Carpenter, 1980); vocabulary development (Baddeley et al., 1988; Baddeley & Wilson, 1988; Cockcroft, 2016; Engle de Abreu, Gathercole, & Martin, 2012; Gathercole & Baddeley, 1989; Gathercole & Baddeley, 1990; Gupta, 2003; Kaushanskaya et al., 2011; Martin & Ellis, 2012; Masoura & Gathercole, 2005; Ramachandra, Hewitt, & Brackenbury, 2011; Verhagen & Leseman, 2016); and mathematics (Alloway, 2007; Swanson & Kim, 2007). Since learning requires the simultaneous storage and processing of information, WM underlies the learners' ability to carry out such processes. According to Dehn (2008), WM evidently has a key role to play in learning, as all information has to be stored and processed by it. Similarly, Baddeley (2003) considered WM as an obligatory doorway through which information passes into long-term memory. Accordingly, the ability to learn is determined by the learners' WM resources.

Arguably, vocabulary learning is one of the most crucial skills that WM is clearly implicated in. The assumption that WM has an essential role to play in L1 and L2 vocabulary learning is prompted by a large body of research, which indicated that vocabulary development in children and adults is reliably correlated with indices of WM capacity, like non-word repetition and digit span. One of the first studies that lent support to this assumption is Vallar and Baddeley (1984) who studied a neuropsychological patient, PV who was known to have an impaired phonological short-term memory. Vallar and Baddeley compared PV's ability to learn foreign language words with her ability to make associations between pairs of unrelated words in her mother tongue. They found that the phonological short-term memory constrained new vocabulary learning, but not the native language-paired associate learning. Based on these findings, Vallar and Baddeley (1984) concluded that verbal short-term memory was essential for new word learning. Similarly, Baddeley and Wilson (1988) demonstrated that phonological loop impairment negatively affects learning word-nonword pairings. This finding supports the claim that the phonological loop has a predictive part in novel word learning.

Furthermore, Gathercole and Baddeley (1989) investigated the correlation between vocabulary development in 4 and 5 year old children and their verbal WM. The scholars used a non-word repetition task as a measure of the verbal WM. The study found that performance on the non-word repetition task correlated significantly with the measures of vocabulary learning. In the same vein, Gathercole and Baddeley (1990) provided further experimental evidence for the positive correlations between the verbal WM measure (i.e., nonword repetition) and vocabulary learning. In their study, Gathercole and Baddeley asked a group of children to learn unfamiliar names and map the newly learned words to referents in the real world. The researchers found a correlation between performance on the nonword repetition and the vocabulary development tests. This pattern of findings has also been reported by Baddeley et al. (1998). Moreover, a further study by Masoura and Gathercole (2005) looked at the contributions of the phonological WM and L2 novel vocabulary learning of 40 Greek children. The study reported that the participants' phonological WM capacity had a close relationship with the learning of new vocabulary items.

Similar results have been found by Engle de Abreu et al. (2012) who looked at the relationship between WM measures and a range of language skills, including vocabulary, grammar and reading in a sample of young Luxembourgish children. The children completed four WM tasks, counting recall, backward digit, digit recall, and nonword repetition as well as three language tests, word picture vocabulary test, reception of grammar test, letter decision test and word detection test. The study reported that the different WM tasks did not link in the same way with the different language measures. Simple WM measures, digit recall and nonword repetition, were strongly associated with the word picture vocabulary test, suggesting that vocabulary development in the native language relies heavily on WM resources. On the other hand, complex WM tasks, counting recall and backward digit recall, manifested strong links with the other syntax and reading measures. Engle de Abreu et al. (2012) concluded that simple and complex WM measures play different roles in supporting language acquisition. A recent study by Verhagen and Leseman (2016) examined the relationship between different WM measures and vocabulary acquisition of bilingual (Turkish learners of Dutch) and monolingual Dutch children. WM capacity of the participants was measured through three simple span tasks (word recall, Dutch-like nonword recall, and Dutch unlike nonword recall) and two complex tasks (backward digit and listening span). Vocabulary learning was assessed through a receptive vocabulary test, Diagnostische Toets Tweetaligheid. Simple WM measures significantly predicted vocabulary scores of both groups of participants. Simple and complex measures were differentially associated with language learning of L1 and L2 learners.

The close link between WM and L2 vocabulary learning was also found in studies that involved adult learners. Gupta (2003), for example, carried out a pair of experiments to explore the association between vocabulary learning and the phonological WM of 52 subjects. A non-word repetition task, an immediate serial recall task and vocabulary learning tests were utilised for data collection. Gupta found a strong association between vocabulary learning and the verbal WM abilities of the participants. Similarly, Martin and Ellis (2012) reported positive correlations between vocabulary learning and the WM capacities of adult monolingual native English speakers.

In a pair of experiments, Kaushanskaya et al. (2011) examined the relationship between WM and vocabulary learning in monolingual English speakers and bilingual (English-Spanish) speakers. To obtain a measure of the phonological WM system, a digit span test was administered. In the first experiment, '*early*' bilinguals who learned both L1 and L2 from an early age were recruited as well as monolingual speakers. Participants completed a receptive vocabulary test (Peabody Picture Vocabulary) in English. Scores on the Peabody Picture Vocabulary test of the bilingual participants correlated tightly with their scores on the digit span test. However, in the monolingual sample, the scores on the WM and the vocabulary measures did not correlate. In the second experiment, Kaushanskaya et al. (2011) compared '*late*' bilinguals who started learning the L2 after they fully acquired their mother language with monolingual speakers. As in experiment 1, participants in this experiment completed a digit span test and receptive and expressive vocabulary tests. The findings of this experiment showed that WM capacity of the bilingual participants correlated with their performance on the receptive and expressive vocabulary tests. However, scores on the digit span test of the monolingual participants did not correlate with their scores on the vocabulary tests. Similarly, Cockcroft (2016) compared bilingual and monolingual South African children. The author used four WM tests, digit recall, nonword recall, counting recall and backward digit recall to assess the verbal WM capacity of the participants. To assess vocabulary development receptively and productively, Cockcroft adopted the British Picture Vocabulary Scale (receptive test) and The Boston Naming test (productive test). The findings revealed that scores on the digit recall test were closely linked to scores on the vocabulary tests for both the bilingual and monolingual participants. However, performance on the nonword and counting recall tests associated only with performance on vocabulary tests for the monolingual participants.

Despite their importance in shedding some light on the association between word learning and WM capacity, the studies reviewed above used word-learning tasks that do not mirror incidental vocabulary learning. The vast majority of the support for the relationship between WM and new vocabulary learning, comes from research that examined the role of WM in vocabulary learning through intentional learning conditions. Of less interest, has been the role of WM in vocabulary learning through incidental learning conditions. Malone (2018) is one of the very few incidental vocabulary studies that examined the relationship between incidental vocabulary learning through multiple input modes and WM. 80 ESL learners were appointed to four input conditions: (1) listening while reading with four repetitions of the target words, (2) listening while reading with two repetitions of the target words, (3) reading only with four repetitions of the target words; and reading only with four repetitions of the target words. The participants across the four experimental groups completed two verbal WM tests, a nonword repetition (simple test) operation span (complex test), and a complex visuospatial WM test, shapebuilder. Malone (2018) reported a strong link between the students' scores on the verbal WM tasks, nonword repetition and operational span and the vocabulary test.

Taken together, the highly consistent findings of the studies discussed above, suggest that WM influences vocabulary achievement in that WM capacity is a strong predictor of successful vocabulary learning.

2.2.3.7 The role of working memory in learning through multiple modalities of input

Learning from multimodal input conditions is a complex process that involves the simultaneous maintaining and encoding of verbal and visual information. To successfully accomplish these cognitive processes, there is a need for sufficient WM capacity. It is well established that WM has a pivotal role to play in learning through multiple modalities of input (Schüler et al., 2011). Mayer (2014) argued that WM underlies the most important work of multimedia learning, as multimedia learning occurs in WM. More specifically, since learning from multimedia/multimodal conditions is a demanding cognitive process, which requires maintaining and processing incoming verbal and visual information, WM underlies our capacity to carry out such processes. In particular, the assumptions of limited capacity and the modality-specific systems of WM make it important and relevant to the present study (Schüler et al., 2011).

The first assumption is the limited WM capacity of the subjects, which is thought to adversely affect learning from multimodal input conditions, i.e., each system of WM can deal with a limited amount of information. Since learning from multimodal conditions requires processing information from multiple modalities, limited capacity of WM constrains learning from such conditions (Wiley, Sanchez, & Jaeger, 2014). Such effects are dependent on the system that is affected by the resource limitations. In other words, the limitation of the phonological loop affects the encoding of verbal information, whilst the processing of visual information is affected by the limitations of the visuospatial sketchpad. Similar to its role in any other forms of learning, WM plays a chief role in learning from multiple input modalities, because all the verbal and visual information needs to be processed and stored in WM before it can be transferred to the long-term memory. The second assumption that makes WM relevant to this study is the modality-specific systems (the phonological loop and the visuospatial sketchpad). Given that learning from multimodal conditions involves processing verbal and visual information, the involvement of the WM modality-specific systems for processing incoming information leads to the development of stronger mental representations and deeper learning of the presented information (Mayer, 2014).

Numerous investigations over the past two decades have looked at whether individual differences in WM capacity account for variations in learning through different input modalities (Austin, 2009; Gyselinck et al., 2000; Gyselinck et al., 2008; Kozan et al., 2015; Lusk et al., 2009; Malone, 2018; Pazzaglia, Toso, & Cacciamani, 2008; Varol & Erçetin, 2016). Researchers have been particularly interested in exploring whether WM capacity mediates the effects of different input modalities on different learning domains, and whether individual variations in WM capacity can explain the different learning results from different multiple input modes. These studies yielded somewhat mixed results regarding the role of WM in learning through different input modalities. Gyselinck et al. (2000), for instance, examined the role of WM capacity in learning physics from two input conditions: a condition of written texts only and a condition including both written texts and illustrations. In Gyselinck et al's study, WM capacity was measured through a visuospatial Corsi blocks test. The findings showed that participants with high WM capacity benefited from the presence of illustrations, while participants with low WM capacity did not. This seems to suggest that WM is involved in learning from text and pictures and that learners need to possess sufficient WM capacity to benefit from the pictorial representations. Likewise, Kozan et al. (2015) explored the role of WM capacity in mediating the effects of different input modalities on text comprehension. Kozan et al. (2015) used a complex verbal WM test (reading span) to measure the participants' WM capacities. They assessed the text comprehension through retention and transfer tests. The study found a significant interaction between WM capacity and input modality on the retention test, suggesting that learners' performances on the retention test was dependant on their WM capacity.

Similarly, Lusk et al. (2009) assigned their participants to a segmented instruction condition and a non-segmented instruction condition to watch a multimedia tutorial. Participants were assessed through content recall and application tests. Participants' WM capacity was measured using an operation-span task. Lusk et al. (2009) reported that students with high WM capacity performed better on the content recall and application tests than those with low WM capacity. The results also showed that WM capacity moderated the effects of the multimedia instruction type. In another similar study, Doolittle and Altstaedter (2009, experiment 1) looked at whether individual differences in WM capacity have an impact on learning about lightening from two multimedia conditions: animations and narration (AN) and animation and seductive details (ANSD). Participants' WM capacity was measured by a complex operation-span task and their content learning was assessed by recall and transfer tests. High-span groups significantly outperformed the low span groups on both recall and transfer tests. However, the interaction between WM capacity and multimedia conditions for recall and transfer tests was not significant, indicating that WM capacity did not moderate the effects of the multimedia environments. In their second experiment, using data collection methods similar to those of the first experiment, Doolittle and Altstaedter (2009, experiment 2) engaged their high and low span subjects into two multimedia situations, animation and narration (AN) or animation and narration that included visual signalling (ANAS), to learn about car brake use. Similar to the results of the first experiment, the findings showed that individual differences in WM capacity accounted for differences in scores on both recall and transfer tests. Moreover, the interaction between WM capacity and the multimedia conditions was nonsignificant, suggesting that WM capacity did not have an impact on the effect of the multimedia conditions.

Varol and Erçetin (2016), on the other hand, examined the effects of WM and gloss types on L2 text comprehension and incidental vocabulary learning. WM capacity of the participants was gauged by a backward digit span task. The study observed a significant relationship between WM and reading comprehension but not between WM and incidental vocabulary learning. Varol and Erçetin (2016) concluded that although WM plays an influential role in text comprehension, it did not mediate the effects of the different types of glosses on incidental vocabulary learning. Similar results were found in a study by Chun and Payne (2004) which looked at the relationship between individual differences in WM capacity and the types of multimedia annotations. The researchers examined whether WM affected the participants look-up behaviour and the achievements on comprehension and vocabulary tests. The findings showed that individual differences in WM capacity accounted for the differing amounts of look-up, meaning that low WM span participants looked up significantly more words than those with high WM span. However, contrary to what Chun and Payne (2004) expected, their data revealed that differences in WM capacity did not influence performance on the vocabulary and comprehension tests, indicating that WM capacity does not affect learning through multimedia conditions. Similarly, Austin (2009, experiment 2) examined learning from multimedia conditions after accounting for the effects of WM, which was measured by the operational span task. Participants were assigned to three multimedia conditions, animation and narration; animation, text, and narration; or animation and text, to learn about the formation of lightning. Austin (2009) demonstrated that individual differences in WM accounted for significant variances in learning from the multimedia conditions, which indicates that WM contributed to multimedia learning. However, the effect of multimedia conditions was not mediated by the participants' WM abilities.

To conclude, the findings of the previous studies in terms of the role of WM in learning from multiple modalities of input are contradictory. The vast majority of the studies surveyed above did not focus on the role of WM in incidental vocabulary learning through different modalities of audio-visual input (Varol & Erçetin, 2016, is an exception). The WM capacity of the participants is therefore examined in the present study. It was thus essential to find out whether participants' WM capacity has a role to play in incidental vocabulary learning from the four different input conditions (VAC, VA, CA, and A only). More importantly, since WM is one of the main sources of individual differences, it was therefore important to explore whether such differences would account for the variations in the learners' scores on vocabulary tests.

2.2.4 The integrated (phonological/executive) model of working memory

The integrated (phonological/executive) model of WM was introduced by Zhisheng Wen in 2016. After reviewing the theoretical literature on WM, Wen (2016b) found two distinctive theoretical research camps of WM: the British WM camp (Baddeley & Hitch, 1974) and the North American WM camp (Cowan, 1999). He noticed that there are still a range of controversies and debates over the structure, functions, nature, and the relationship with long-term memory amidst these theoretical camps of WM. Wen (2016b, 2019) thus attempted to develop a conceptual framework through which he can arrive at some unifying characteristics of the WM construct. Those unifying characteristics would allow for applications of WM in various applied fields of human cognition such as second language acquisition (SLA). More importantly, Wen (2019) sought to understand and measure the WM construct in SLA. Building on the insights of the extant theoretical frameworks of WM, Wen (2016b, 2019) proposed his integrated conceptual framework for theorizing and measuring WM in SLA. In the integrated model, which is known as the phonological/executive model, Wen endeavoured to give a more principled approach to specify how WM can be operationalised and conceptualised in SLA research. Wen (2016b) integrated insights from both cognitive psychology and applied linguistics to better guide future investigations looking at the relationship between WM and SLA.

Wen (2016b, p. 80) conceives WM as "multiple cognitive mechanisms and functions implicated in the temporary *maintenance, access,* and *control* of a limited number of pieces of linguistic information to facilitate the *acquisition, representation, processing,* and *development* of various domains and activities in learning a second language". In this conception of WM, Wen views WM as a set of executive functions, such as processing, maintenance, rehearsal, as well as non-executive functions, like chunking, retrieving, and consolidating. This definition includes the unifying characterisations of the WM construct such as the limited capacity, multiple components/functions, and the relationship with long-term memory.

The structure of Wen's model of WM in SLA is inspired by the multicomponent model of Baddeley (1986). The architecture of the model is visually presented in visually presented in Figure 2.3. As portrayed in the figure below, the structure of the model consists of four levels or layers that aim to integrate long-term memory, the different components of WM, and the functions of WM in SLA. The first layer at the bottom of the model is the long-term memory. This permanent repository stores long-term knowledge of L1 and L2 in the bilingual brain. This memory contains "inter alia", the bilingual learner's mother language or L1 competence, which includes, for instance, lexical and grammatical competence. The memory also stores L2 knowledge alongside the L1 competence. The L2 knowledge stored in long-term memory encompasses different domains, including lexis, semantic, phonemes, morphosyntax, and formulaic sequences. Of utmost relevance, the integrated model postulates that long-term memory includes the activated items of bilingual learner's L1 and L2 knowledge in WM. This overlapping part WM and long-term memory is labelled as LT-WM.

At the next level up from the bottom of the integrated model lies the different components of WM. Similar to Baddeley and Hitch's (1974) conception of WM, Wen's (2016b) integrated model of WM posits that WM consists of multiple components of WM. These components include a phonological WM system, an executive system, a visuospatial sketchpad system, and an episodic buffer. Due to the lack of studies and attention on the relationship between the components, episodic buffer and the visuospatial sketchpad, and SLA, Wen (2016b, 2019) excluded them from his integrated model. Instead, his framework focuses on the phonological WM and the executive WM system due to their importance in different L1 and L2 learning activities. According to Wen (2016a), these two systems are distinguished from each other based on the mechanisms associated with each system and the tasks used to measure each one. In addition, these two systems have distinctive effects on specific L2 domains and tasks. Specifically, the phonological WM component is usually implicated in the acquisition and development of L2 lexis. The executive WM system, on the other hand, plays an instrumental role in L2 reading/listening comprehension, and L2 interaction.

The third layer towards the top of the framework is the multiple executive functions and mechanisms of WM associated with the phonological and executive WM components. The phonological WM system has been assigned two main functions, the phonological short-term store and the articulatory rehearsal that are equivalent to the functions associated with the phonological loop of Baddeley and Hitch's (1974) multicomponent WM model. Wen (2019) indicated that the extant studies lend support to the claim that the phonological WM plays a pivotal role in L1 and L2 acquisition and that individual differences in phonological WM capacity is one of the main reasons for variations in language learning performance, such as vocabulary. On the other hand, the executive WM system which is purported to regulate attention control processes, includes a range of executive roles, like updating, switching, and inhibition of information. Wen (2019) observed that little attention has been given to the various functions associated with those two key components of WM. That is, most of the former studies have focused on conceptualizing the phonological and executive WM systems as broad WM systems. However, Wen thinks that it is of vital importance to tap the different subskills or functions of each system of WM to explore their effects on L2 learning.

At the top level of the integrated WM framework of Wen (2016b, 2019) lies the measures that are used to tap each WM component. Wen uses different tasks for measuring the phonological and executive WM systems. The phonological WM is tapped by simple memory span, such as digit span and (non)word span. Whilst, the executive WM system is assessed by complex memory span tasks (for example, the reading/listening span, counting span, etc.). Wen (2019) justified his decision for implementing separate simple and complex span tasks for assessing the two WM components based on the distinctive nature of the two sets of WM tasks, simple and complex. In light of his extensive review of the past research, Wen confirmed that simple memory tasks provide reliable and valid means for assessing the phonological WM system, while complex memory tasks have been proven reliable and valid measures for the executive WM system. For example, Wen found that past research (e.g. Gathercole, 2006) has demonstrated that the simple non-word repetition task approximates the mechanisms (the phonological short-term store and the articulatory rehearsal) associated with the phonological WM system in L1. On the other hand, Wen indicated that several previous studies (e. g. Conway et al., 2005) reported a close relationship between complex memory spans and attention-regulating or executive aspects (i.e. updating, switching, and inhibition) of WM in language learning.

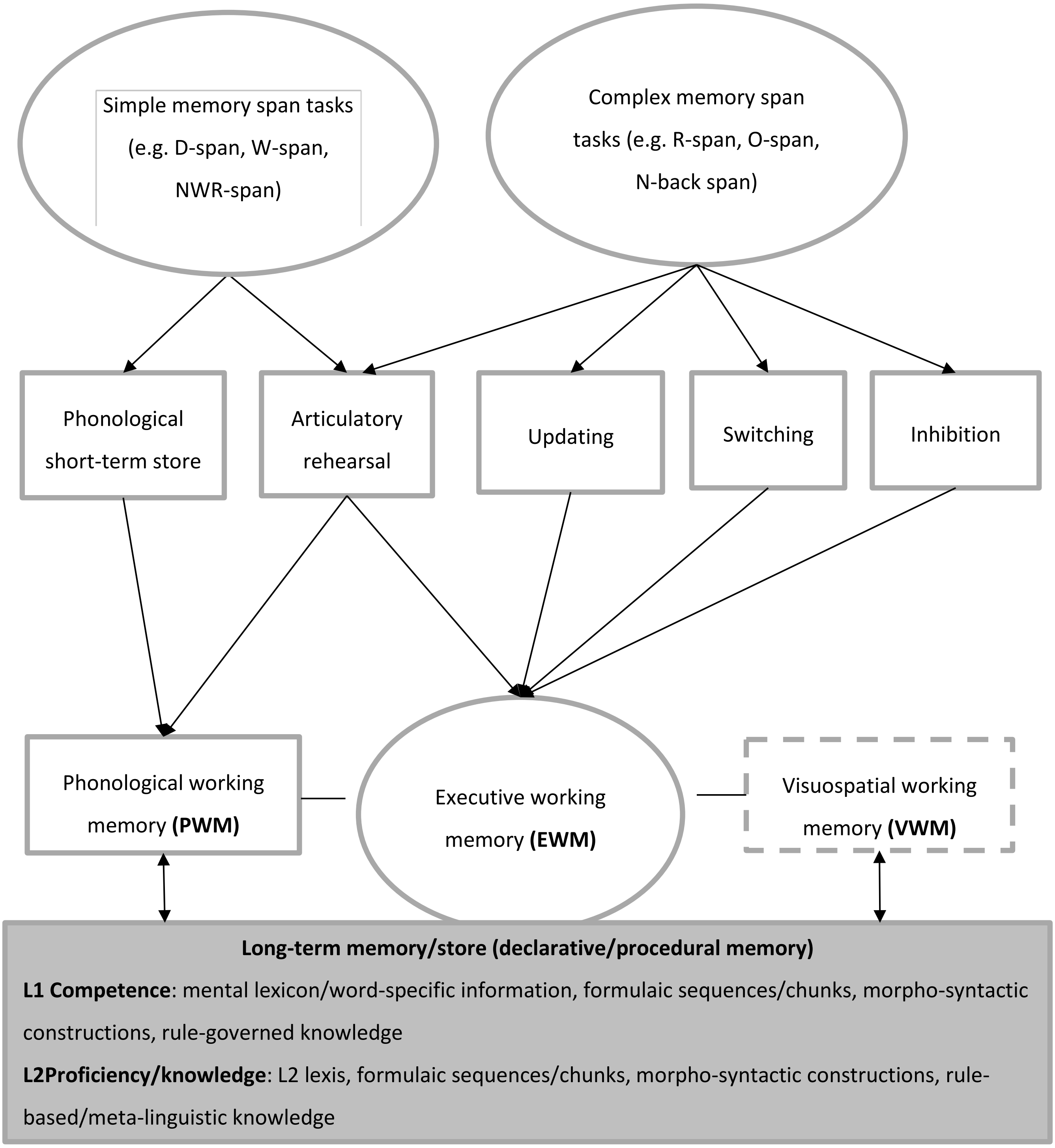


Figure 2.3 The integrated (phonological/executive) model of WM. Adopted from Working memory and second language learning: Towards an integrated approach by Wen (2016, p. 83).

The basic tenet of the integrated model is that SLA researchers should treat the different systems of WM (phonological WM and executive WM) and their distinctive cognitive mechanisms and functions separately. The rationale behind this assumption is that the systems and their associated mechanisms and functions exert distinct effects on SLA activities. In this regard, Wen (2016b) emphasizes that SLA studies should operationalise and measure the different components of WM separately.

2.2.5 Section summary

To sum up, DCT and CTML are two prominent theories that are believed to best explain the effect of different modalities of audio-visual input on language learning in general, and vocabulary learning, in particular. Both theories provide an understanding of how the two cognitive channels, namely the verbal channel and the nonverbal/pictorial channel mediate cognitive activities. According to Paivio (1986) and Mayer (2014), using multimodal input modes (i.e. verbal and visual representational codes) promotes learning and recollection of the presented information better than a single input mode. Since the present study involves multimodal input conditions which present vocabulary information through a variety of modalities that include both verbal information (e.g. caption and audio) and visual information (e.g. dynamic video images), it is felt that both DCT and CTML are helpful in contributing to explanatory framework. In addition, since learning in multimodal input conditions is underpinned by WM, as Mayer (2005, p. 38) argued "the central work of multimedia learning takes place in working memory", it is thus vital to account for the role of the learners' WM abilities in learning through the different modalities of audio-visual input. In addition, considering that the two modality-specific systems of WM are important for an efficient use of memory, they make WM particularly relevant to learning through multimodal input conditions. In short, WM limitations and the modality-specific systems are the two main features of WM that make it especially relevant to the present study. The next section looks at the concepts of modalities of input and presents previous studies that examined the effect of different input modalities on incidental vocabulary learning and retention.

2.3 Modalities of input

Input modality is an issue of great interest in the field of learning psychology, as it offers significant benefits for learning (Mayer & Moreno, 2002). The leading modality of input that has long dominated education is the verbal mode (Mayer & Moreno, 2002; Moreno & Mayer, 2007). However, in recent years, educational materials incorporating different modes/forms of input, such as pictures, illustrations, graphs, videos, and animations have received increasing attention in different educational domains. Moreover, the emergence of multimedia/multimodal learning materials has created key platforms for fostering L2 learning. This section casts some light on the concept of input modality and critically reviews the relevant empirical studies that have been carried out on the effect of different modalities of audio-visual input on L2 incidental vocabulary learning.

A number of researchers attempted to define the concept of multimodalities of input or multimedia. From the information processing perspective, Clark and Paivio (1991), for example, described this concept as presenting information as visual or verbal. Likewise, Guichon and McLornan (2008) viewed multimodality as the use of different semiotic codes (pictures and speech) to present information. Similarly, Moreno and Mayer (2007) defined multimodal learning conditions as conditions that represent information using verbal and non-verbal modes of input. In a broader and more comprehensive definition that covers all possible modes of presentation/input that have been used by researchers in the field of multimedia, Mayer (2005) defined multimedia as:

> Presenting both words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video). By words, [he] mean[s] the material is presented in verbal form, such as using printed text or spoken text. By pictures, [he] mean[s] that the material is presented in pictorial form, such as using static graphics, including illustration, graphs, diagrams, maps, or photos, or using dynamic graphics, including animation or video. (p. 2)

However, it is worth noting that the present study uses different terms, including multiple modalities of input, and multimodal input conditions/modes interchangeably to refer to Mayer's (2005) concept of multimedia. That is to say, the present study uses the terms listed above to refer to the use of on-screen text (captions), audio (soundtrack of the videos), and video (dynamic pictures/moving pictures in the movies).

The values of using different input modes have been widely acknowledged (Moreno & Mayer, 2002). The premise of multimodal learning centres on the idea that the combination of dual presentation modes (pictures and words) augments learning (Mayer, 2003; Mayer, 2014). This assertion is based on the findings of several studies that have been conducted to verify the effects of dual input modes on learning (Akbulut, 2007; Al Seghayer, 2001; Chun & Plass, 1996; Mayer et al., 2001). Mayer (2014) asserted that the simultaneous presentation of verbal and visual information enables students to establish mental connections between the presented information, which thus results in better learning. An additional benefit of multimodal input conditions is that they are developed "in light of how the human mind works" (Mayer, 2005, p. 31). This implies that presenting information using pictures and narrations maps the proposals of

Paivio's DCT (1986) and Baddeley's WM model (1986), which postulate that human information processing system includes dual channels for pictorial and verbal processing. One of the central implications of DCT is that presenting information using both visual and verbal modes enhances the recall of the materials presented (Liu & Todd, 2014).

A further advantage of using multiple modalities of input is that learners with different learning preferences could select the input mode they prefer (Moreno & Mayer, 2002). In this way, those who prefer visual processing of information and those who prefer auditory processing of information can both benefit from the multimodal input (Mayer et al., 2001). The literature is rich with studies that have explored the effects of different modalities of input on different types of learning. However, this section presents a discussion of only the most influential studies that have explored the effects of different modalities of input on incidental vocabulary learning.

2.3.1 Review of previous empirical studies on different modalities of input

Research conducted on the effects of different input modalities on incidental vocabulary learning contributes to our understanding of how the different input modalities effectively enhance learning. An examination of previous empirical studies on the effect of multiple modalities of input reveals two basic strands of research. The first strand contains studies conducted in the realm of CALL that examined the effect of multimodal glosses/annotations on enhancing incidental vocabulary growth from reading (for example, Akbulut, 2007; Al Seghayer, 2001; Boers, Warren, He, & Deconinck, 2017; Chen & Yen, 2013; Chun & Plass, 1996; Marzban, 2011; Yoshii, 2006, to name a few). The majority of these studies reported results in favour of implementing multimodal annotations and provided evidence for the DCT and the CTML. Although they yielded promising findings, their findings are inconclusive regarding the most effective multimodal glosses for enhancing L2 word learning. For instance, Chun and Plass (1996) compared the effects of textual, pictorial, and video media annotations on vocabulary learning. Subjects were placed into three experimental groups: the first group received textual annotations only, the second group received textual and pictorial annotations, and the third group received textual and video annotations. Findings of this study indicated that incidental vocabulary learning was improved as a result of the use of different kinds of annotations. It was also found that the second experimental group who received text paired with picture annotations outperformed the other two groups on the production and recognition vocabulary tests. Chun and Plass's study attested to the advantage of dual modes of text and pictures over the single mode of text in enhancing incidental vocabulary uptake.

A similar study by Al Seghayer (2001) was conducted to compare the efficacy of three annotation conditions, textual annotations paired with still pictures; textual annotations coupled with dynamic video clips; and textual annotation only on fostering L2 incidental vocabulary learning. Participants completed two vocabulary recognition and production tests. The findings of the study revealed that the annotations that contained dual presentation modes were more effective for enhancing incidental vocabulary learning than the single annotation type. However, contrary to the findings of Chun and Plass (1996), Al Seghayer (2001) found that participants who received textual annotations coupled with video clips scored highest on both vocabulary tests.

However, Akbulut (2007) reported contradictory results to those of the above reviewed studies regarding the positive effects of dual presentation modes on incidental word learning. Akbulut (2007) examined the effects of three annotations types: text and picture; text and video; and text only on L2 incidental learning of form recognition, meaning recognition and meaning recall. The study found that the two dual annotations conditions (i.e. text and picture and text and video) boosted incidental word acquisition better than the single annotation condition, text only. However, no significant differences between the dual annotations modes (text and pictures and text and videos) were detected, indicating that the two visual groups were equally effective. These contradictory findings could be ascribed to a number of factors, including the type of vocabulary tests used, the type of subjects each study targeted (proficiency of the learners) and the type of pictures and videos used.

However, recently, Boers et al. (2017) presented contrasting findings to those of the aforementioned studies. In their study, Boers et al. (2017) compared the effects of two gloss conditions: text-only, and text coupled with picture on vocabulary uptake, using a meaning recognition test and form recall test. The findings of this study showed that the multimodal gloss-condition was not more effective than the single-gloss condition, failing to provide evidence for the effects of multimodal input conditions on vocabulary growth. Overall, the vast majority of the existing studies to date lend support to the claim that multimodal annotation conditions stimulate incidental word learning better than single annotation conditions.

The second strand of research includes studies that examined the effects of multiple modalities of audio-visual input on incidental vocabulary learning. The majority of this type of research looked at the effects of different types of captions and the differential effects of L1 subtitles versus L2 captions. These studies developed different treatment conditions, used different vocabulary achievement tests, adopted various types of audio-visual materials, and targeted different L2 speakers. Nevertheless, they confirmed that L2 incidental vocabulary learning is possible through different audio-visual conditions with varying amounts of learning gains due to the distinct input modes used. Table 2.2 demonstrates an overview of the studies carried out on the effects of different modalities of audio-visual input on L2 incidental vocabulary learning.

Table 2.2. Schematic representation of previous empirical studies examining incidental vocabulary learning through different modalities of audio-visual input

study	Treatment conditions	Number of target words	Vocabulary measures	Key findings
Sydorenko (2010)	Group1 (video+ audio+ caption) Goup2 (video + audio) Group3 (video + caption)	28	Form recognition (written and auditory); Translation test; Word knowledge scale (VKS)	Groups with captions scored higher on the written form recognition than the non-captioning group. The non-captioning group outscored the captioning groups on the spoken form recognition test The VAC group significantly outscored the other groups on the translation test.
Winke et al. (2010)	Group1 (video with captions) Group2 (video without captions)	unspecified	Meaning recall test (half items were presented aurally and half were presented in written forms)	The captioning group gained higher scores on the vocabulary tests than the non-captioning group.
Montero Perez et al. (2014)	Group 1 (full captioning) Group 2 (keyword captioning)	20	VKS as a pre-test; Form recognition test;	The captioning conditions resulted in better vocabulary learning gains than the non-captioning condition in the form

	Group 3 (highlighted keyword captioning) Group 4 control group (no captioning)		Clip association test; Meaning recall test; Meaning recognition test	recognition and clip association tests. No significant differences between the captioning groups in the form recognition and clip association tests; Group two and three outscored the other groups on the meaning recognition test
Peters et al. (2016) Experiment1	Group1 (L1 subtitles) Group2 (L2 captions)	39	Spoken form recognition test; Spoken meaning recall test	Group2 performed better than group1 on the spoken form recognition test No statistical significant differences between the two groups in the spoken meaning recall test
Peters et al. (2016) Experiment2	Idem	Idem	Written form recall test; Written form recognition test; Written meaning recognition	The captioning group gained better scores on the form recall and on the form recognition tests than the subtitles group. The subtitles group performed better than the captioning group on the meaning recognition test
Montero Perez et al. (2018)	Group 1 (full captioning) Group 2 (keyword captioning)	18	VKS as a pre-test; Form recognition; Clip association	The captioning groups outperformed the control group on the form recognition and clip association tests; The glossed keyword captioning condition was the most

	Group 3 (glossed keyword captioning) Group 4 control group (no captioning)		Meaning recall; Meaning recognition;	effective captioning type for aiding incidental vocabulary learning

One of the studies that looked at the effects of different input modes on different aspects of word knowledge was Sydorenko (2010). Sydorenko studied the effects of different input modalities on the recognition of aural and written word form, overall vocabulary gains, and attention to various modes of input of 26 novice adult learners of Russian. The learners were assigned three treatment conditions: condition 1, video, audio, and caption (VAC), condition 2, video and audio (VA), and condition 3, video and caption (VC). Sydorenko (2010) used three short videos from a Russian comedy, from which 28 target words were selected based on the probability of their appearance in the participants' textbooks. The study employed three vocabulary tests, a form recognition test, a translation test, and a word knowledge test. Half of the tested vocabulary items were presented aurally and half in a written form. It was reported that participants in the VAC and VC groups gained higher scores on the written form recognition test than those in the non-captioning group (i.e. VA group). Conversely, the VA group significantly outscored the captioning groups (VAC and VA) in the auditory form recognition test. As far as the learning gains of word meanings were concerned, the study indicated that the VAC group outperformed the other two groups. The highest scores were on the form recognition tests across the three experimental groups. Despite its importance in contributing to our understanding about the effects of different modalities of audio-visual materials on the learning of different word knowledge aspects, form and meaning, Sydorenko’s (2010) study is limited due to the fact that learning cannot be considered incidental, as her subjects were informed about the aim of the study at the start of the intervention. Sydorenko forewarned her participants about the implementation of the subsequent tests and asked them to pay attention to the language while watching. This may have encouraged the subjects to focus on the stimuli and eventually affected the learning of the target words. This is a crucial issue, as Wesche and Paribakht (1999) stressed that informing the subjects about the purpose of the study is a critical factor that distinguishes incidental learning from intentional learning. Laufer and Hulstijn (2001) also warned that to ensure the incidental character of learning, learners should not be given advance warning about the implementation of a subsequent test.

The majority of studies on modalities of audio-visual input examined the differential effects of L1 subtitles and L2 captions. However, there is more promising evidence for the effectiveness of adding captions to audio-visual materials for boosting incidental vocabulary learning. Winke et al. (2010) carried out a study to investigate the effects of the presence of captions on vocabulary learning. One hundred fifty learners of Arabic, Spanish, Russian, and Chinese participated in this study in which each participating group watched a series of three documentary videos, two times in their respective target language, once with captioning and once without. The Spanish group was divided into two groups: one viewed the videos twice with captioning and another viewed the videos twice without captions. The videos were translated and then dubbed into the respective languages of the participants. The study administered a meaning recall test in two test presentation formats (a written format and an auditory format) for each of the four target languages. Version A of the test presented half of the test vocabulary items in a written form and version B presented the other half orally. The participants' prior knowledge of the target words was pre-tested using a vocabulary knowledge scale. The findings of the study indicated that the captioning conditions were more effective for aiding vocabulary learning than non-captioning conditions, regardless of the vocabulary test presentation format. The findings also revealed that the Spanish subjects who watched the videos twice with captioning significantly outperformed their counterparts in the other group who viewed the videos twice without captioning on both vocabulary tests versions A and B.

Recently, some studies developed a better-designed methodology to examine the effects of different modalities of audio-visual input on L2 incidental vocabulary learning. Montero Perez et al. (2014) is among the few studies that used multiple vocabulary tests to measure form and meaning at a receptive and productive level of mastery. The researchers used a form recognition test, a meaning recognition test, a meaning recall test, and a clip association test. The study sample consisted of 133 Flemish learners of French who were allocated to three conditional groups and a control group to watch three French video clips twice. The first conditional group viewed the video clip with full captions (FC), the second group watched the video clip with keyword captions (KC), and the third conditional group watched the clip with full captions with highlighted keywords (FCHK). Keywords were chosen based on their importance in understanding the meaning. Participants were pre-tested on their knowledge of the potential target words using the VKS of Wesche and Paribakht (1996). The control group had only watched the film clip without captions (NC). 17 target words were selected from the short French clips.

The results showed that the participants in the captioning groups (FC, KC, and FCHK) gained significantly higher scores on the form recognition test and on the clip association test than the control group did. Yet, no significant differences were reported between the captioning groups in these two tests. In addition, the subjects in the KC and FCHK groups significantly outperformed the control group on the meaning recognition test. No significant differences were found in the learning uptakes among the treatment groups (FC, KC, and FCHK). As far as the meaning recall test was concerned, the findings indicated that contrary to the study's hypothesis, the study failed to detect any significant differences in the scores of the treatment groups and the control group. Similar to the results of Sydorenko (2010), participants across all groups obtained the highest mean scores on the form recognition test, followed by clip association and meaning recognition test. The least developed knowledge aspect was meaning recall.

A recent pair of experiments by Peters et al. (2016) was conducted to explore the effects of different modalities of audio-visual input on incidental learning of two aspects of word knowledge, form recognition and form-meaning mappings. The study also sought to look at the relationship between frequency of occurrence and incidental vocabulary learning. In the first experiment, 28 Dutch-speaking learners of English were appointed to two audio-visual conditions: condition 1, L1 subtitles and condition 2, L2 captions to view a relatively short documentary video twice. The researchers administered two vocabulary achievement tests, spoken form recognition and spoken-meaning recall. The two vocabulary tests were run on two occasions as, pre-and-immediate-post to measure knowledge of 39 target words. The results of the first experiment indicated that the participants in the captioning condition gained higher scores on the spoken form recognition test compared to their counterparts in the subtitles condition. It was also found that there were no major differences in the spoken meaning recall test between the two groups. The study further showed a positive correlation between frequency of word repetition and word learning.

In experiment two, which had developed a similar design to experiment one, Peters et al. (2016) recruited 18 Dutch-speaking learners of English from a vocational school. The subjects were assigned to either a captioning group or subtitles group. Using a cartoon film (Simpsons) the experiment measured vocabulary learning gains using three tests: written form recall and recognition and written meaning recognition. In line with the findings of the first experiment, Peters et al. found that the captions group scored greater on both the form recall test and the form recognition test (21.5 % in the form recall) and (29.1% in the form recognition) than the subtitles group (11.1 % in the form recall) and (25.5% in the form recognition). It was also reported that on the meaning recognition test the subtitles group (31.8%) had higher scores than the captions group (17%). Overall, the study concluded that incidental vocabulary learning occurs from audio-visual materials. A central limitation of the study is that although the authors found a link between frequency of occurrence of words and vocabulary learning, they did not compare the learning uptakes across different frequency groups.

Along similar lines, Montero Perez et al. (2018) explored the effects of two enhancement techniques on L2 incidental vocabulary learning of form and meaning. The first technique consisted of augmenting the audio-visual inputs (videos) with different types of captioning: full captioning (FC); keyword captioning (KC); and glossed keyword captioning (GC). The second technique was related to the announcement of the post-test. Subjects were either forewarned that they would receive vocabulary tests immediately after viewing the videos (intentional condition) or were not informed about the vocabulary tests (incidental condition). The study consisted of 227 Dutch-speaking learners of French who were assigned to eight experimental conditions: (1) full captioning and incidental; (2) full captioning and intentional; (3) keyword captioning and incidental; (4) keyword captioning and intentional; (5) glossed keyword captioning and incidental; (6) glossed keyword captioning and intentional; (7) no captioning and incidental; and (8) no captioning and intentional. Findings revealed that the three captioning groups (FC, KC, and GC) scored significantly higher on the form recognition and clip association tests than the non-captioning condition. On the form recognition and clip association tests, the glossed captioning group (GC) performed significantly better than the FC group and the control group. The GC group also significantly outperformed the other treatment groups (FC, KC), and the control group on the meaning recall test. Montero Perez et al. (2018) concluded that glossed captioning is a more effective technique to use in audio-visual materials than other captioning types in aiding incidental vocabulary learning. The second enhancement technique, test announcement was found not to affect learning gains.

2.3.2 Critique of the previously described studies

Taken together, the surveyed studies above demonstrated that incidental vocabulary learning occurs from audio-visual inputs. However, different input modes have differential effects on the amount of learning gains. Although these studies provided interesting findings regarding the differential effects of different modalities of audio-visual inputs, they are constrained by a myriad of limitations. The first concern in some of these studies (Montero Perez et al., 2014; Montero Perez et al., 2018; Peters et al., 2016, are the exception) is the absence of the measurement of participants' prior knowledge of the target words. Testing participants’ pre-existing vocabulary knowledge prior to the intervention is crucial as to establish the number of words that resulted from the intervention. Since these studies have not pre-tested the participants' prior knowledge of the target word items, it is likely that some of the words reported as learned on the post-tests were already known beforehand. Schmitt (2010) emphasised that the aim of using a pre-test is to ensure that the participants have no prior knowledge of the target words before the start of the research intervention, as the absence of pre-testing can jeopardise the vocabulary gain results.

A further limitation of these studies is the lack of retention data. All the former described studies have not administered a delayed post-test to gauge how long the reported learning gains on the immediate post-test(s) will last. Rather, they focused on initial word learning. Waring and Takaki (2003) pointed out that a delayed post-test can show the real and durable learning of target items, as Schmitt (2010) warned that immediate post-tests do not show long-term acquisition. Nation and Webb (2011, p. 102) indicated “having only an immediate post-test overestimates learning because forgetting inevitably occurs”. In their study, Waring and Takaki (2003) found that only very small amounts of newly acquired vocabulary items were retained over some time later.

Another problem associated with the methodology of some of the former discussed studies (i.e., Montero Perez et al., 2014; Montero Perez et al., 2018) is the use of different types of pre-tests and post-tests. Those studies pretested the participants’ prior knowledge of the target words using test types that were different from the immediate post-test types. That is, they used vocabulary knowledge scale of Wesche and Paribakht (1996) as a pre-test to measure the subjects’ previous knowledge of the target words and used different post-test types, such as form recognition test, clip association test, and meaning recall test to measure learning gains. However, Nation and Webb (2011, p. 103) advised against the use of dissimilar test types, arguing “we do not know if a change in score from one test to another is because of the time lapse or because they are different kinds of tests”. It is thus essential to run the same type of test(s) across the study.

Previous research was also limited by the paucity of the data on the effect of frequency of occurrence on incidental vocabulary acquisition. To the best of the researcher’s knowledge, only three viewing studies examined the role of repetition on incidental vocabulary learning through different audio-visual input modalities (Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013). Nevertheless, despite advancing our knowledge about the relationship between frequency of occurrence and incidental vocabulary learning, those three studies are limited in several ways. Firstly, they did not compare the exact frequency of word occurrence (e.g. 1 – 2 – 3 – 4 repetitions) on the development of different aspects of word knowledge. In other words, they did not pinpoint the exact number(s) of repetitions required for incidental vocabulary learning to take place. They only revealed that the chances of learning a vocabulary item increase when the number of encounters increases. Schmitt (2010) pointed out that the number of exposures is essential to developing the incremental learning of vocabulary items. The studies did not also contrast the effect of repetitions on incidental vocabulary learning across different input modes, which is essential to showing how many encounters are needed in different audio-visual input modes. Although there is no agreed number of repetitions necessary for the development of word knowledge, it is evident from prior reading and listening studies that repetitions of the target words tend to positively affect incidental vocabulary learning (Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003). Examining the frequency of occurrence is thus important in audio-visual studies, as it would enable researchers to determine the number of repetitions needed for a word to be learned in different input modes.

The previous studies above are also subject to a further limitation that concerns the neglect of the role of imagery. The majority of the past research looked at how imagery occurs with the aural input (concurrent presentation of imagery with audio) fostering vocabulary learning, but it is still unknown whether imagery was superior to audio in supporting vocabulary learning. Those studies did not develop a design that would tease apart the effects of imagery on the learning of the materials presented. Unlike the early research of different multimodal glosses, most audio-visual studies mainly focused more on the effects of different types of captioning and the differential effects of L1 subtitles versus L2 captions and discarded the role of imagery in multimodal audio-visual learning. Rodgers (2018) indicated that imagery has a key role to play in learning from audio-visual materials. He urged future researchers to examine the effect of imagery on incidental vocabulary learning.

Lastly, and most importantly, none of the previously discussed studies has empirically tested how WM contributes to incidental vocabulary learning from different modalities of audio-visual input. As has been indicated before, WM plays a pivotal role in learning from different input modalities, especially from text and pictures. It has been indicated that one of the key reasons for individual variations in learning outcomes in many learning fields is related to individual differences in WM capacity. It has also been documented that the effects of multiple modalities of audio-visual input are moderated by the learners' WM capacity. Thus, investigating the role of WM in incidental vocabulary learning and retention through different modalities of audio-visual input is of utmost importance.

To conclude, the previously reviewed studies are limited by a variety of methodological flaws and limitations. Although they have contributed to our understanding of the effectiveness of different modes of audio-visual input on incidental vocabulary learning, there is a need to address their limitations. The present study has a value for the literature, as it expands on earlier studies in many ways, as will be explained below.

2.3.3 Gaps in literature and the focus of the study

The focal aim of the present study is to investigate the effect of different modalities of audio-visual input on L2 incidental vocabulary short-term learning and long-term retention. The secondary aim is to examine the role played by both a learner-related factor (i.e. WM) and an item-related factor (i.e. frequency of occurrence) on L2 incidental learning and retention. The foregoing review of the related literature demonstrates that past studies were restricted by a number of problems attributable to their research methodology and design. The present study is thus motivated to address the limitations of those studies and expands on their designs. Particularly, it attempts to fill the following gaps:

- The absence of pre-test of prior knowledge of the target words;
- The scarceness of long-term retention data (delayed-post-test);
- The dissimilarity of the pre-post-and-delayed-post-test types;
- The lack of data on the effect of frequency of occurrence of the target words;
- The neglect of the role of imagery in multimodal learning conditions;
- Lastly, and most importantly, the absence of WM measurements

The present study is unique in incorporating some key related aspects (i.e., multiple modalities of audio-visual input, DCT, CTML, and WM). The integration of these aspects in this study is justified as follows. Firstly, as has been stated before, the value of using multiple modalities of input is said to have a positive impact on learning and memory of the presented information. The theoretical foundation of this assumption is Paivio (1986) and Mayer (2005). Both DCT and CTML assume that recall and learning is enhanced when information is presented verbally and visually, which, as a result, can facilitate learning of the presented information. Because the multimodal input conditions in this study involves verbal and visual stimuli, examining the corresponding verbal (the phonological loop) and visual (the visuospatial sketchpad) cognitive systems of WM that encode verbal and visual incoming information is of particular interest. In addition to that, the limited WM capacity of the learners can have an effect on the efficacy of the multiple modalities of input. The literature, by and large, suggests that WM capacity affects learning. That is, the higher the WM capacity an individual has, the more successful the learning will be. Learning through multiple modalities of input raises the question of whether learners with different WM capacities can benefit from multimodal audio-visual input modes. In addition, since WM capacity is a key predictor of individual differences in learning, examining this variable in this study is essential.

To sum up, the present study sets out to examine these closely interrelated variables. The basic premise of this study is that the different audio-visual input conditions (VAC, VA, CA and A only) would have differential effects on incidental short-term vocabulary learning and long-term retention. However, the effects of these different conditions might be mediated by the WM capacity of the learners.

2.3.4 Section summary

In summary, the past studies examined a multitude of issues concerning incidental vocabulary learning through audio-visual materials. It should nonetheless be clear that the previous studies suffered from a wide range of methodological flaws that need to be considered in future research. The forgoing literature review was essential in exploring what still needs to be done in the field of incidental vocabulary research through audio-visual input. The present study therefore addresses a range of knowledge gaps found in the previous similar studies. The following chapter presents the research design and methodology of the present study.

Chapter 3 Methodology and research design

The preceding chapter presented the theoretical and conceptual accounts of the topic under investigation. This chapter provides a discussion of the methodological issues related to the design and the execution of the current study. It starts by presenting the philosophical foundations of research, ontology and epistemology. It further describes the research paradigm underpinning the current study. Following that, the research design of the study is discussed in detail. A description of the participants, materials and the selection of the target words is then provided in detail. Next, a detailed account of the treatments and the procedures involved in the research intervention is outlined. The chapter also presents the data collection instruments and highlights the rationale behind using them. Finally, it ends with a description of the ethical considerations, issues of validity and reliability, pilot studies, and the analytical framework.

3.1 Objectives and research questions

As set out in Chapter 1, the present study aimed to achieve the following. Firstly, it aimed to examine the effect of different modalities of audio-visual input (video, audio and caption (VAC), video and audio (VA), caption and audio (CA), and audio only (A only)) on L2 incidental vocabulary short-term learning and long-term retention of spoken form recognition, meaning recall, and meaning recognition. It also intended to compare the effects of these learning conditions (VAC; VA; CA; and A only) on L2 incidental vocabulary learning and retention. Additionally, it aimed to examine the effect of the frequency of occurrence on incidental vocabulary learning from these audio-visual conditions. A further aim was to explore the relationship between incidental vocabulary learning and retention and the capacity of the learners' WM. Finally, the study aimed at examining the role of WM in mediating the effects of the different input modalities on L2 incidental vocabulary learning and retention. To this end, the study adopted an experimental research design and employed a range of quantitative instruments to collect data. The following table presents the research questions addressed and sources of data used in this study.

<table>
<tr><th>Research Questions</th><th>Data collection methods</th></tr>
<tr><td>1. To what extent do audio-visual conditions, VAC, VA, CA, and A only, enhance/promote L2 incidental vocabulary short-term learning of spoken form recognition, meaning recall, and meaning recognition?</td><td rowspan="4">• Spoken form recognition test;
• Spoken meaning recall test;
• Spoken meaning recognition test
• Semi-structured interviews</td></tr>
<tr><td>2. Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary short-term learning of the spoken form recognition, meaning recall, and meaning recognition?</td></tr>
<tr><td>3. Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary long-term retention of the three vocabulary knowledge aspects, spoken form recognition, meaning recall, and meaning recognition?</td></tr>
<tr><td>4. To what extent does the frequency of occurrence affect L2 incidental vocabulary learning from the four experimental conditions?</td></tr>
<tr><td>5. Is there a relationship between WM and incidental vocabulary short-term learning and long-term retention? If so, which subset(s) of WM, phonological loop or visuospatial sketchpad contribute(s) to this relationship?</td><td rowspan="3">• forward digit recall task
• backward digit span recall task
• dot matrix task
• the odd one out task</td></tr>
<tr><td>6. Which of the WM task types, simple or complex, best predicts incidental vocabulary short-term learning and long-term retention?</td></tr>
<tr><td>7. Does WM mediate the effect of the four modalities of audio-visual input on incidental vocabulary short-term learning and long-term retention?</td></tr>
</table>

When reviewing the above research questions, it can be said that these research questions helped to examine issues of interest to the present study. The first research question aimed at exploring the potential of the different modalities of audio-visual input in fostering incidental vocabulary uptake. The second and third research questions were concerned with the differential effects of the four modalities of input on incidental vocabulary short-term learning and long-term retention. The fourth research question was formulated to explore role of frequency of occurrence of the target words in enhancing incidental vocabulary learning in the four different conditions. The last three research questions aimed at investigating the relationship between WM and incidental vocabulary learning and retention. In particular, research questions five and six explored the nature of the relationship between WM to see which system of WM contributed to the relationship, if any, and which test type, simple or complex, best predicted performance on the vocabulary tests. The last research question explored whether individual differences in WM capacity account for the variation in vocabulary learning and retention rates. It also aimed at exploring whether WM mediates the effects of the different audio-visual input modalities.

3.2 Philosophical overview

Research philosophies are of great importance in any research, they are one of the essential characteristics of research. An investigator's research philosophy entails fundamental positions in regards to the way s/he views reality and the world. These positions inform the choice of the research strategy and the methods (Bryman, 2016). Nowadays, there is a plethora of methodological texts on the different research philosophical stances and research paradigms, however scholars have not reached a consensus regarding the classifications and categorisation of the research philosophies. This section attempts to define the concepts ontology and epistemology. It then continues to discuss the theoretical framework or the paradigm that underpins the present study.

In the philosophy of knowledge and reality, both ontology and epistemology are key elements (Grix, 2010). Ontology has been perceived by many authors as "concerns views of reality" (Crotty, 2003, p. 3); "concern[s] the very nature or essence of the social phenomena being investigated" (Cohen, Lawrence, & Keith, 2007, p. 7), or "the nature of social entities" (Bryman, 2016, p. 28). In contrast, epistemology, according to Grix (2010, p. 64), is "concerned with the theory of knowledge". In other words, Crotty (2003, p. 8) defined epistemology as knowing "how we know what we know". Epistemological claims are seen as "concern the very bases of knowledge-its natures and forms, how it can be acquired, and how communicated to other human being" (Cohen et al., 2007, p. 7).

The importance of the ontological and epistemological claims lies in the fact that they give rise to the theoretical paradigms, and the paradigms in turn inform the selection of methodological considerations, which also give rise to the choice of data collection methods (Cohen et al., 2007). In the research literature and methodological texts, research paradigm has received considerable attention, as it is one of the most important elements of the process of any research. Grix (2010, p. 79) perceived research paradigm as “our understanding of what one can know about something and how one can gather knowledge about it”. The three particular paradigms that govern different disciplines are positivism, interpretivism and pragmatic. The current study was situated within the positivism paradigm.

Positivism is one of the most commonly adopted philosophical paradigms in educational and applied linguistic research. Positivism is a term used to reflect the philosophical views a researcher adapts in his/her research (Cohen et al., 2007). It is described by Mertens (2005, p. 10) as “ the social world can be studied in the same way as the natural world, that there is a method for studying the social world that is value free, and that explanations of a causal nature can be provided". It is evident from these definitions that positivist researchers use scientific methods to examine and confirm patterns of behaviours objectively.

One of the basic characteristics of a positivist researcher is that she/he aims to test a hypothesis or set of hypotheses deduced from an existing theory (Grix, 2010). Positivists hold a variety of underlying assumptions. Firstly, they generally believe that the basis of genuine knowledge is sense experience and it is only through observation and experiment they can advance this knowledge (Cohen et al., 2007). Further, positivists assume that social world or reality exists independently from the inquirer which can be studied objectively (Crotty, 2003). By studying knowledge objectively, the researcher’s bias is reduced to the minimum, as s/he is external to the research environment. Her/ his role is to guide and control the research process. The third assumption is that the methods used to study the social world/reality are value-free, ‘neutral’ (Grix, 2010; Mertens, 2005). Positivist researchers generally adopt quantitative methodology, utilize an experimental approach and administer pre-and-post-tests to gauge the outcomes of the study. Since the focus of the positivists is on objectivity, they tend to make themselves external to the study site and only control and guide the process involved in the study (Bryman, 2016).

The present study adopted the positivist paradigm as the theoretical framework for the following reasons. Firstly, since the positivist paradigm operates through quantitative methods, this study used quantitative measures, vocabulary tests and WM tests, as the leading methods in data collection. In addition, seeing as the positivist researchers typically tend to develop experiments to examine the cause-effect relationship, this study developed an experiment to examine the effects of the different modalities of audio-visual input (VAC, VA, CA, and A only) on incidental vocabulary learning and retention.

3.3 Research design

Research design is another central component of the research process. Creswell (2012, p. 20) defined it as “the specific procedures involved in the research process: data collection, data analysis, and report writing”. Similarly, Bryman (2016, p. 40) described it as “a structure that guides the execution of a research method and the analysis of the subsequent data”. Selecting an appropriate research design is a crucial step to be taken in any research, as different aspects of the research process depend on it (Bryman, 2016). Cohen et al. (2007, p. 79) pointed out that research designs are selected based on their "fitness for purpose”. This means that the study’s aim determines the suitable research methodology and design. It is worth noting that the research literature reached no agreement regarding the number of research designs available, however the most oft-mentioned and most commonly used designs are (1) case study, (2) survey design, (3) longitudinal design, (4) experimental design, and (4) comparative design (Bryman, 2016). This study adopted an experimental design and therefore discusses this design in detail.

Experimental research represents the most scientific form of quantitative research which is explained by Dörnyei (2007, p. 115), as “it can establish unambiguous cause-effect relationships”. In this regard, Creswell (2012) remarked that experimental research examines the effects of the independent variables on the dependent variables. The current study was conducted through experimental research because it looked at the effects of the independent variables (different modalities of audio-visual input, frequency of occurrence, and WM) on the dependent variable (incidental vocabulary learning and retention). Furthermore, Cohen et al. (2007) commented that one of the basic characteristics of experimental studies is the control and manipulation of the conditions in order to examine and measure the effects of the outcomes of the intervention. The aim of the manipulation, which is commonly known as a *treatment* is to examine a causal relationship (Mackey & Gass, 2005).

Cohen et al. (2007) outlined three types of experiments: true experiments, quasi-experiments, and natural experiments. To them, the first two types (true experiments and quasi-experiments), widely used in educational research, vary in relation to the random assignment of participants to experimental groups. In true experiments, subjects are randomly assigned to different conditions of the experiment, while quasi-experimental studies adapt a non-random assignment approach, which means that participants are not randomly assigned to the experimental groups (Cohen et al., 2007). Considering the fact that random assignment of participants reduces threats to the internal validity, a true experiment study design has an advantage over a quasi-experimental study design (Dörnyei, 2007), but Dimitrov and Rumrill Jr (2003, p. 160) claimed that using intact classes (as the case in the present study) could lessen "the reactive effects of the experimental procedure and, therefore, improves the external validity of the design". Since, it is sometimes practically impossible for researchers to undertake a true experiment; it is therefore advisable to adopt a quasi-experiment, where the random assignment of participants is quite impractical (Dörnyei, 2007).

The present study, therefore, used a quasi-experimental research design utilising pre-immediate-post-test and delayed-post-tests, as displayed in the figure below. The reason for adopting this design is related to the fact that study was conducted in a university context in which it was impossible to artificially form groups of participants for the current experiment, as the students were already placed in their relative classes by the respective academic departments. In a way to deal with the potential limitations of the absence of non-randomization of participants into experimental groups, a "semi-randomization" procedure was implemented which relates to the practice of arbitrarily appointing genuine/intact classes to the experimental treatments (Mackey & Gass, 2005, p. 143). To this end, the four existing classes were randomly assigned to the four treatment conditions of the study (VAC; VA; CA; and A only).

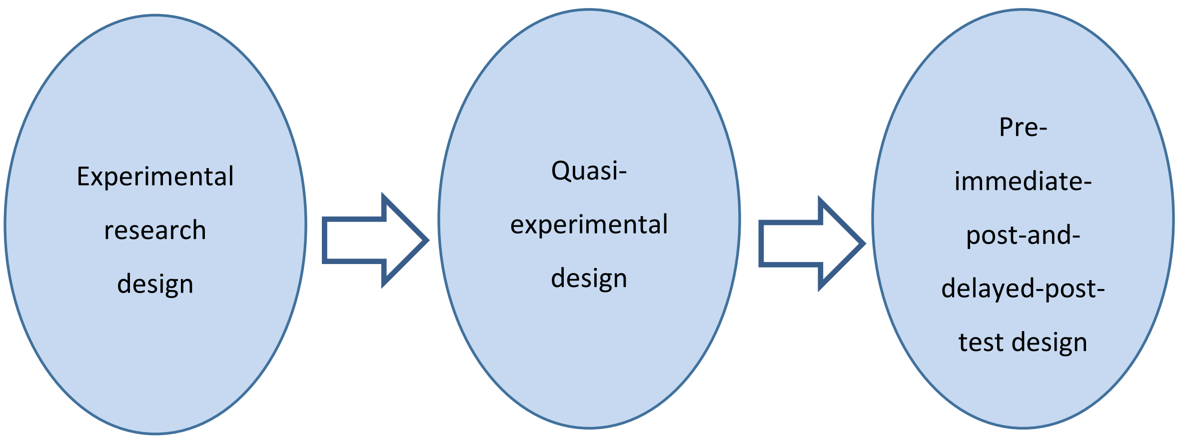


Figure 3.1.1 The study design

To conclude, the present study employed a pre-immediate-post-test-and-delayed-post-test quasi-experimental design to examine the effects of the different modalities of audio-visual input (VAC, VA, CA, and A only) on L2 incidental vocabulary learning and retention. The performance of the four experimental groups was compared to explore whether the different modalities of input have differential effects on vocabulary learning and retention of three vocabulary knowledge types: spoken form recognition, meaning recall, and meaning recognition.

3.4 Approach selected for this study

The process of choosing an appropriate methodological approach is of paramount importance to ensure accuracy in the research design. The distinction between quantitative and qualitative approaches is one of the most widely-acknowledged distinctions in research methodology (Dörnyei, 2007) and the debate over the merits of each one of them has long attained great impetus (Bryman, 1984). The nature of the research questions determines the suitable methodological approach. Seeing as the aim of quantitative research is to determine a relationship between variables (Mackey & Gass, 2005), this study thus employed a quantitative enquiry. Dörnyei (2007, p. 24) viewed quantitative research as "involve[ing] data collection procedures that result primarily in numerical data which is then analysed primarily by statistical methods". Dörnyei (2007, p. 34) posited that one of the main strengths of the quantitative approach rests on the fact that it is "systematic, rigorous, focused, and tightly controlled, involving precise measurement and producing reliable and replicable data that is generalizable to other contexts". The reason for adopting a quantitative approach in the present study was related to the fact that quantitative tests could effectively show whether the intervention improved the participants' vocabulary knowledge.

However, though the present study was purely quantitative in nature, it also employed an additional qualitative data collection method; semi-structured interviews. Qualitative data was used to help explain the results generated by the quantitative tests. Bryman (2006, p. 106) argued that quantitative and qualitative research methods can be fruitfully combined in a study, because "when one generates surprising results" the other can be helpful in explaining such surprising results. One limitation of the quantitative tests was that they could not fully explain certain issues that emerged in this study. More specifically, in some cases the quantitative measures could not provide direct evidence for what helped or hindered learning through the different multimodal input conditions and, thus, the use of a qualitative method was deemed necessary. For example, the quantitative data showed that the participants involved in the VAC input condition performed poorly on the meaning tests compared to their counterparts in the VA condition, although the VAC condition was considered to be the optimal input condition, as it presented the input through three modes; audio, images of the videos, and on-screen text (captions).

3.5 The participants and setting of the study

The selection of appropriate participants for study is crucial to its quality (Cohen et al., 2007). Vogt (2007) stated that the bases of the generalisation of the results to the population are the representativeness and the size of the sample. Representativeness refers to a situation where each individual of a population has an equal chance of being recruited in the study (Mackey & Gass, 2005). In relation to the optimal sample size, Dörnyei (2007) observed that there are no set rules in deciding the sufficient sample size, because according to Bryman (2016), the decision depends on a variety of considerations, such as time and cost. However, some scholars have proposed a rough minimum sample size as a guideline. Fraenkel and Wallen (2006), for example, suggested about 15 to 30 participants for experimental studies, 100 subjects for descriptive studies, and 50 participants for correlational studies. Similarly, Dörnyei (2007) proposed similar numbers to Fraenkel and Wallen's (2006), recommending that correlational studies involve at least 30 subjects and experimental studies should recruit at least 15 participants.

The target subjects for the present study were first-year university students at Jeddah Community College (JCC) at King Abdulaziz University (KAU), Saudi Arabia. This University was chosen through a convenience sampling method. This University was chosen because the researcher had worked for the JCC for about five years (between 2009 and 2014), which made it possible for him to gain access to the English language classes. The researcher could argue for considering the first-year students as the most appropriate participants for this study for two reasons; firstly, these students were undertaking an intensive English course as part of the requirement of the foundation year at JCC. Another reason was that the participants were assumed to have an intermediate English proficiency level, based on their performance on an in-house English language proficiency test conducted by JCC, which is felt it would enable them to comprehend the video clips used in the experiment.

The English language course at the JCC, lasts for a year, and is a fundamental part of the foundation year. It is developed to help students enhance and develop their language abilities in order to secure entry into the different disciplines at the JCC. The programme has two levels of instruction, GRC 111 (beginners to pre-intermediate) and GRC 112 (intermediate). During the foundation year, the intensive English language course is intended to take language learners from beginner level up to the upper-intermediate level. The duration of each level is 14-15 weeks (one academic term), with 20-hours of formal instruction per week. Students enrolled in this programme take three English classes (4-hours) per day 5 times a week as follows: reading/writing (1 hour), listening/speaking (1 hour) and grammar (2 hours). Immediately after admission into the JCC, the GRC department runs the mandatory English language placement test and students are placed in relevant levels (GRC 111 or GRC 112) according to the test scores. Apart from the English language course, students enrolled in the foundation year are also required to enrol in the other core courses, such as mathematics and communication skills, which are all taught in English.

After I secured the approval from the JCC administration, I was allocated 4 intact classes from the intermediate level (GRC 112). Initially, 162 Saudi EFL learners from the four intact classes, who speak Arabic as a first language took part in the study. However for various reasons, including absenteeism and dropping out, the total number of the subjects who completed the different phases of the study dropped substantially to 108 subjects. For meaningful statistical analysis, participants who failed to attend any of the treatment sessions were eliminated from the study. Table 3.1 shows the distribution and size of the participants in each treatment group. The participating students were all adult males aged between 18 and 22 years old. Male learners were chosen, as in Saudi Arabia, males and females are segregated into different educational sites and it was thus never possible for the researcher to recruit female participants. The participants represented students from different provinces of the Kingdom of Saudi Arabia and shared very similar language learning experiences. It is evident from the vocabulary pre-tests that the participants across the four classes had almost similar vocabulary knowledge before they took part in the study intervention (this will be discussed this in section 4.1.1). The learners were compensated for their time by receiving extra credit in the English course from the concerned teachers.

There is a particular reason for choosing Saudi Learners of English as participants in this study. As has been discussed in section 1.2 most Saudi learners of English possess a small vocabulary size. Even after completing 832 hours of formal classroom instruction, a number of studies (e.g., Alsaif & Masrai, 2019; Alsaif & Milton, 2012) repeatedly indicated that Saudi students finish high schools with approximately 1000 of the most frequent words. Thus, it should not be surprising to find that at university level, Saudi students' vocabulary size ranges between 1650 and 3000 of the most frequent words (Masrai & Milton, 2012). To get some indication of the size of vocabulary knowledge of other ESL/EFL learners, Laufer (2010) pointed out that English language learners in Japan, Indonesia, Israel, and Spain acquired between 600 to 3,720 word families after around 1,500 hours of instructed input. Unfortunately, these are indeed small amounts compared to the size of word knowledge required for effective use of English. Nation (2006) concluded that ESL/EFL learners should know about 8,000 – 9,000 word families in order to comprehend 98 % spoken and written discourse. In contrast to the small figures of words known by Saudi learners of English, it is estimated that the vocabulary knowledge of native English speaking students at a university level is approximately 20,000 word families (Goulden, Nation, & Read, 1990).

These small figures therefore stress that Saudi students fall far short from gaining the level of vocabulary knowledge needed to use English efficiently. Inevitably, limited vocabulary knowledge would have negative impacts on students' progress in the other language skills as well as on their higher educational study. Direct evidence has been found for the close link between lexical knowledge and the development of other language skills, including reading comprehension (Al-Homoud & Schmitt, 2009) as well as academic achievement (Milton & Treffers-Daller, 2013). This raises the question of the factors behind this small level of vocabulary knowledge.

There are a number of factors contributing to this problem. Firstly, Alsaif and Milton (2012) analysed the vocabulary input presented in the English language textbooks used in some Saudi schools and found that only about 2500 most frequent words are presented in such textbooks. Though, there are potentially several sources for language learning, including audio-visual materials, stories, novels, internet, and smartphone applications, in Saudi Arabia, the main source for vocabulary learning is the language of textbook and the classroom. However, the poor exposure to and insufficient coverage of vocabulary in Saudi language textbooks appears to have serious implications on word learning. In addition, Alhazmi and Milton (2015) associated Saudi students deficient vocabulary knowledge with poor reading skills and comprehension. Saudi learners of English have been found to experience a unique difficulty in developing their L2 reading skills. Masrai and Milton (2018) indicated that the reading process in Arabic and English are considerably different which makes it truly difficult for Arabic learners of English.

A probability or a random sampling method was used in the process of recruiting participants for completing WM tests and attending the semi-structured interviews. The participants were chosen randomly through simple random sampling in Excel. On a voluntarily basis, 12 participants from each of the four experimental groups were invited to complete the four WM tests. Although the original plan was to include all the participating students, it was not practically feasible to invite more than 12 participants due to the time constraints of the study and the limited access the researcher was given to the participating groups. Also, three participants from each group were chosen randomly to attend the interviews.

Table 3.1. Number of subjects in each treatment group, completing the WM tests, and attending the interviews

Group	Starting size of participants	Final number of participants	Participants completed the WM tests	Participants attended in the interviews
Experimental group 1: video, audio and caption (VAC)	35	26	12	3
Experimental group 2: video and audio (VA)	37	27	12	3
Experimental group 3: caption and audio (CA)	46	30	12	3
Experimental group 4 audio only (A only)	44	25	12	3

As mentioned above, since it was inevitable to work with intact classes, the four intact classes were randomly assigned to the treatment conditions. That is, GRC 112 class 0A was assigned to the first treatment group (VAC), GRC 112 class 0B was assigned to group 2 (VA), GRC 112 class 0C was assigned to group 3 (CA), and GRC 112 class 0D was allocated to group 4 (A only). The researcher, who had five years of EFL teaching experience, instructed the four experimental groups.

3.6 Materials

Choosing suitable materials for any research is of great importance, as unsuitable materials would have been counterproductive to meet the aims of the study. During the initial stages of the present study, the intention was to adopt educational/instructional video clips designed specifically for ESL/EFL learners that may best suit the participants' language level. However after a meticulous screening process and thorough examination, I realised that such instructional video clips may be inappropriate for this study for different reasons. Firstly, most of those clips were developed for educational purposes and were centred on some themes, such as greeting people, making orders, compliments, making reservations, which would not serve the aims of the study. That is, vocabulary items used in such video clips were usually amongst the most common ones and hence the chances of encountering mid-frequency vocabulary and infrequent words were low. According to Rigg (1991), simplified materials (e.g., ESL/EFL video clips in this case) tend to provide insufficient exposure to the target language, greatly manipulate the language and distort appropriate pragmatic use. Additionally, instructional videos were different from the videos that students were likely to view outside the classrooms. Based on my intuition, which was later confirmed in both the pilot study and the main study, the majority of Saudi students would prefer to watch authentic videos/films, as this is one of the most popular leisure activities among them. Therefore, it was felt that it would be more suitable to choose authentic audio-visual materials.

The selection of suitable authentic materials when used with L2/FL learners is very critical and requires deep consideration of their levels of linguistic capacity, as choosing difficult materials can render the study a failure (Kweon & Kim, 2008). Therefore in the process of selecting suitable materials (videos) for this study, a number of criteria, which were inspired by a number of previous studies (for example, Baltova, 2000; King, 2002; Montero Perez et al., 2014), were taken into consideration. These criteria were:

1. Shared level of difficulty;
2. Compatibility with the proficiency level of the learners;
3. Suitability of the content;
4. Authenticity;
5. Watch-ability of the videos in terms of pace of delivery speech and clarity of speech;
6. The length of the videos;
7. Appealing to the participants' interest

In light of the criteria outlined above, four authentic documentary videos were chosen which were available on YouTube. Table 3.2 below represents an overview the videos. Authentic materials were preferred for this study, as they could increase students' interest, motivate them to learn (Pellicer-Sánchez & Schmitt, 2010), and provide promising results (Kweon & Kim, 2008). Similarly, Rodgers (2018, p. 192) indicated that viewing authentic videos has several benefits, including "increase[ing] [learners] motivation towards language learning". The chosen video clips were all about the life of different animals, such as lionesses, elephants, dogs, and kangaroos. These types of documentary videos (about animals) were chosen, as they do not require any background knowledge to be understood. That is, learners did not need to have any specific culturally related knowledge to comprehend the content, as learners' familiarity with the content will affect the rate of learning (Al-Homoud, 2007). Furthermore, it was important to choose culturally neutral materials in the context of Saudi Arabia, where there are some restrictions on some types and contents of audio-visual materials, as they may be considered to threaten traditional values.

A number of procedures were carried out to ensure that the selected materials satisfy the aforementioned criteria. In regards to the language difficulty of the videos, three major steps were applied to determine the difficulty of the materials. Firstly, two experienced EFL teachers from the institution that hosted this study (JCC) were asked to view the videos and comment on the linguistic level of difficulty of the videos, suitability of the content, and the compatibility of the materials with the proficiency of the learners enrolled in GRC 112. The two teachers confirmed that the videos were suitable in matching the target students' language level. The teachers also reported that the content of the videos was acceptable. After transcribing the four videos, they were fed into the vocabulary frequency-profiling software, Vocabprofile/Range (see Appendix A, for a sample of the output of Vocabprofile analysis of one of the videos). This freely available online software was developed by Nation and Heatley (1994) to perform lexical text analysis. It classifies words in texts into four different frequency bands or lists: (1) the 1.000 most frequent words (K1); (2) the 2.000 most frequent words (K2); (3) the academic word (AWL); and (4) off-list words or infrequent words. The analysis of the four videos showed that most of the running words were from the first category (K1).

One of the key factors that affects the understanding of the discourse is the lexical coverage (Schmitt, 2008), which refers to "[the] percentage of known words in a piece of discourse" (van Zeeland & Schmitt, 2012, p. 457). Nation and Webb (2011) pointed out that generating sufficient comprehension enables the learner to derive the meaning of the unknown words from the surrounding context. Researchers have thus suggested different estimates for the required lexical coverage for comprehension, ranging from 90% to 98% based on the type of discourse (Laufer & Ravenhorst-Kalovski, 2010; van Zeeland & Schmitt, 2012). Peters and Webb (2018) claimed that the TV programme that they used in their study was suitable for their target population of EFL learners with an intermediate proficiency level, as 90.3% of vocabulary items in that TV programme were from the K2 list. Therefore, in the present study, it could be safely presumed that the videos were suitable given that the participating learners' proficiency level was classified as intermediate and most of the running words were from the K1 list (see Table 3.2). In addition, the chosen video clips were piloted in the first pilot study. In the first pilot study, participants were asked to evaluate the videos in terms of the contents, language, and watch-ability. The majority of the subjects reported that the videos were interesting and suitable for their language level. They also reported that the delivery speech pace and clarity of the videos were acceptable.

Table 3.2. Overview of the videos and the results of Vocabprofile analysis of the texts of the four videos

Title	theme	K1 words	K2 words	AWL words	Off-list words
Kangaroo Comeback	This video is about the story of a mother (Kangaroo) and her joey (baby Kangaroo) and how he comes into his own.	81.54%	7.27%	1.19%	10%
Elephant Queen	The video is about a herd of elephants who had to leave their homeland looking for a safer place. The film tells the viewers about the hardships the elephants faced in their journey towards the new place.	81.24%	8.46%	2.27%	8.03%
The power of lionesses	The film focuses on a pride of lionesses and portrays how another group of lionesses from the north attacked them to take control of their territory.	80.08%	7.87%	1.71%	10.34%
The wonderful dogs	This video is about dogs and their relations with humankind. It shows the different services the dogs offer to humans and demonstrates how people train dogs to do different works.	80%	6.42%	3.15%	10.43%

Having chosen suitable videos, the next important step was to design the four treatment conditions (video, audio, and caption (VAC); video and audio (VA); caption and audio (CA); audio only (A only). This was done over different phases, in the first phase, after obtaining a permission from the publishers of the videos to reuse the videos for the research purposes, the four videos were downloaded from YouTube and saved as separate files on a computer. The original videos were shortened to a duration of 15 minutes. Next, the soundtrack of each video was transcribed. To ensure their accuracy, transcriptions and written texts were then reviewed by two fellow PhD students (one was a native English speaker and the other was an EFL teacher). After that, in order to make the first treatment condition (VAC), written texts (captions) were added to the relevant videos using software called Windows MovieMaker (WMM). This software was chosen, because it allows the adjustment of all features necessary for this experiment, such as the font colour, size, and position. Concerning the development of the third condition, caption and audio (CA), firstly the soundtracks of the videos were exported in the form of sound files (MP3) instead of video files and each audio file was imported to the WMM software. Following this, captions were added to each soundtrack file on a black background. In order provide maximum contrast, the classic caption format of white font was used. The captions were carefully matched to their corresponding audios. For the fourth treatment condition (audio only), the videos were only converted into MP3 files using a website specifically designed to extract audio tracks of the videos and convert them to MP3 files (Online video converter). The last phase of the treatment condition design was checking the materials for any flaws. This was done by running the developed conditions several times. In addition, two fellow PhD students were asked to evaluate and check the edited materials and conditions. No major concerns arose.

3.7 Target words

One of the most fundamental steps in vocabulary research is the selection of target words, for if they do not match the aims of a study; the results of the study will render unreliable (Schmitt, 2010). One key factor likely to affect the learning of target words in incidental vocabulary research is the concurrent language learning which refers to other sources of language learning such as ESL/EFL courses that are taking place at the same time as the intervention (Nation & Webb, 2011). It is thus essential to eliminate the effects of factors that may influence the learning of target words. One of the widely used techniques by several past studies to circumvent these challenges was the use of non-sense words or substitute words (Brown et al., 2008; Van Zeeland & Schmitt, 2013; Waring & Takaki, 2003). Substitute/nonsense words are words that "look like plausible English words and take on English spelling conventions" (Waring & Takaki, 2003, p. 136). For example, in their study, Brown et al. (2008) changed the spelling of some real target words to make substitute words (e.g., '*happy*' is rendered, '*mird*'). Although this method proved useful in ascribing any development in vocabulary knowledge of the target words solely to the treatment adopted, applying this approach in the present study was not practically feasible because the selected videos were authentic and therefore it was particularly challenging to manipulate them.

The selection of the target words in the present study was informed by several criteria adopted by a range of former studies (for example, Sonbul & Schmitt, 2010; Van Zeeland & Schmitt, 2013). The first criterion was that words should have high frequency of repetitions in the materials (frequently repeated in the input). This criterion is important, as choosing words with different numbers of occurrences in the input would allow the researcher to examine the effects of frequency of occurrence (which is a variable investigated in this study) on incidental vocabulary learning. The second key criterion was the probability of knowing the target words. That is, the target words should be very unlikely to be known by the majority of the participants. Additionally, following Montero Perez, Peters, and Desmet (2015) and Sydorenko (2010), target words were chosen based on whether they were visually supported by the images in the videos. In light of these criteria, the texts of the selected videos were analysed using the Vocabprofile software of Nation and Heatley (1994) to find the words that the learners were highly unlikely to know based on the researcher's intuition. Words from the K2 list and the off-list, which had more than two repetitions in the texts, were chosen as potential target word candidates. Words from the K1 list and the AWL list were not included as target word candidates due to their high frequency of use. More importantly, words were particularly selected from the two frequency groups, the K2 and off-list, because they were unlikely to be met outside the treatment. An initial list of the potential target words was thus compiled (60 words).

Several procedures were followed to ensure that the selected target words satisfied the previously-mentioned criteria. Firstly, the list of the potential target words was sent to two experienced EFL instructors (both are PhD holders) at the institution (JCC) that hosted the present study in Saudi Arabia, to examine the probability of encountering the potential target words in the textbooks and to check whether the chosen words were likely to be known to the students enrolled in GRC 112. The feedback received from the two teachers indicated that some words from K2 list (e.g., eagle, pride, flood, hunt, protect) were likely to be known by some participants, but the rest of the words were unlikely to be known or met in the language prescribed textbooks. The words that were likely to be known by some of the target participants were excluded from the list. However, relying on the teachers' judgement alone did not seem sufficient. Therefore, vocabulary pre-tests were conducted to guard against the possibility of prior knowledge of the target words (Schmitt, 2010). The pre-tests included a larger number of new words to allow the researcher to exclude words that were unknown by 30% of the participants following Montero Perez et al. (2014). In light of the results of the pre-tests, words known by 30 % of the participants, were excluded from the list. Furthermore, to ensure that the selected target words satisfied the last criterion, which concerned the usefulness of the images of the videos for word learning, two fellow PhD students were invited to view the videos and were given the list of the potential target words. They were asked to evaluate how well the selected target words were supported by the visual images of the videos. The comments received from these evaluators confirmed that the images of the videos were generally supportive in the depiction of the chosen words' meanings.

Therefore, in total, 36 target words were retained as they were confirmed to meet the above-mentioned criteria. 86 % of the 36 target words were drawn from the off-list frequency band and 14 % were from the K2 list as illustrated in Table 3.3. The initial intention was to include all word classes in the research. However, as has been found in the study of Pigada and Schmitt (2006), it was not possible to include word classes, such as adjectives and adverbs in the list of the target words, as they did not satisfy the criteria discussed before. In fact, it proved extremely difficult to find adjectives and adverbs that meet the criteria specified above in such relatively short videos. Consequently, the target words consisted of concrete nouns and verbs only. In addition, the use of authentic materials made it impossible to take into account other word-related factors, such as the position of the words in a sentence and the salience of the words, which were found to play a part in fostering incidental vocabulary learning (Nation & Webb, 2011; Vidal, 2011).

Table 3.3 Target words, their number of repetitions, and their frequency lists

Target words	Frequency of repetition in-text	Frequency list
Bush	2	K2
Reign	3	Off-list
ignite	5	Off-list
Cub	14	Off-list
Invaders	5	Off-list
detest	3	Off-list
Territory	10	Off-list
Scent	5	K2
Materialize	4	Off-list

Herd	6	Off-list
dehydration	2	Off-list
chase	2	Off-list
Swamp	6	Off-list
Trunk	5	K2
thrive	6	Off-list
tendons	5	Off-list
suckle	2	Off-list
patrol	6	Off-list
ambush	4	Off-list
Shepherd	7	Off-list
breed	7	Off-list
flock	9	Off-list
string	2	K2
Mesmerise	10	Off-list
Prey	4	Off-list
Graze	3	Off-list
Pup	8	Off-list
Predators	7	Off-list
Hop	10	Off-list
Joey	22	Off-list
Pouch	10	Off-list
Shade	2	K2

brim	2	Off-list
Claws	4	Off-list
lurk	2	Off-list
Ancestors	4	Off-list

In order to account for the role of frequency of occurrence, the target words were assigned to three frequency groups based on their repetitions in the input as displayed in Table 3.4. The strategy of assigning the target words into different frequency groups has been widely applied in a number of incidental vocabulary studies (for example, Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Waring & Takaki, 2003). However, it is worth mentioning that there is no unified rule for assigning the target words into frequency bands. The past studies allocated the target words into frequency groups that would allow for a meaningful analysis. The present study thus developed the following three frequency groups: (1) 2-4 occurrences, (2) 5-7 occurrences, and (3) 8+ occurrences. It was not possible to keep the number of target words in each frequency group constant, owing to the fact that the selected videos had unequal distribution of words across the three frequency groups. Therefore, each of the three frequency groups consisted of a different number of words, as displayed in the table below.

Table 3.4. Target words and their frequency of occurrences in the materials

Frequency list	Number of target words	Number of words in each frequency range		
		(2-4)	(5-7)	(8+)
Off-list	31	13	10	8
K2	5	3	2	0

3.8 The treatment

Study design is one of the primary research components. According to Mackey and Gass (2005), typical experimental research adopts 'comparison group design' or 'control group design'. In the former, the researcher compares between two or more groups of participants who receive different treatments and take pre-and-post-tests. Conversely, in the latter design, the researcher compares between two or more groups, but one of these groups (control group) receives no treatment and all groups (the treatment and the control) take pre-and-post-tests. The present research favoured the ‘comparison group design’ over the other type due to the fact that one of its chief aims was to compare the effects of different modalities of audio-visual input on L2 incidental vocabulary learning and retention. This design was also applied in a number of similar studies, such as those of Sydorenko (2010) and Peters et al. (2016). Another reason for not including a control group was related to the fact that at the time of data collection (second semester of the academic year 2016-2017), there were only four classes of GRC 112, which thus made it not feasible to include a fifth intact class as a control group. However, to control for the lack of a control group, Mackey and Gass (2005) suggested that researchers should make sure that the syllabus used during the research intervention should not include a particular language focus related to the experiment. In the present study, as has been highlighted before, efforts were made to ensure the choice of target words that were highly unlikely to be met in the classroom syllabus.

As has been said before, to account for the lack of non-random assignment of the participants, the four intact classes were assigned arbitrarily to the four treatment conditions as illustrated in Table 3.5. The first treatment group (VAC) watched the videos while listening to the English audios and reading the full-on-screen texts (captions) (for examples of the VAC computer screens, see Appendix B). The second group (VA) viewed the videos and listened to the English audios, without reading the captions (see Appendix C, for examples of the VA computer screens). The third group (CA), however, listened to the audios of the videos while reading the captions that appeared on a black screen background, without viewing the video images (for examples of the CA computer screens see, Appendix D). In the last treatment condition (A only), participants listened to the soundtracks of the videos, without viewing the video images or reading the captions. All the four treatment groups received each of the treatment materials only once and sat the three vocabulary tests on three separate occasions as: pre-tests, immediate post-tests, and one month delayed tests. All experimental groups were also treated equally and received the same materials but in different modalities.

Table 3.5. The treatments of the study

Groups	Treatment conditions
Group 1	video, audio, and caption (VAC)
Group 2	video and audio (VA)
Group 3	caption and audio (CA)
Group 4	audio only (A only)

3.9 Procedures of the experiment

The present study was carried out over a period of 10 weeks during the second semester of the 2016-2017 academic year. The four groups of subjects took part in the four main phases of the study. The four groups followed the same procedures and received similar instructions to avoid any problem that may arise owing to the use of different procedures and instructions. Table 3.6 provides an overview of the different phases of the experiment.

The first phase began a week prior to the commencement of the treatment in which I visited the four classes which I had been given access to and introduced a general idea about the study to the subjects. I also gave them the participant information sheet and consent form to read and sign (see Appendix E). The participants were informed about their rights for confidentiality, anonymity, and safety. Upon agreement to participate in the study, each participating group was administered the three vocabulary pre-tests: spoken form recognition, spoken meaning recall, and spoken meaning recognition. A short and simple demonstration was provided on how to respond to the tests to ensure that the participants understood exactly what was required.

The second phase consisted of four treatment sessions and the vocabulary immediate post testing. This part took place over four consecutive weeks (once a week), from week two to week five. In each week, each experimental group was met separately in a computer lab and was exposed to one of the videos in their relevant conditions. During each treatment session, the learners were encouraged to pay attention to the video and to watch or listen for pleasure. They were also not allowed to take any notes nor consult a dictionary. After the end of the four-week treatment sessions, the four groups took the three vocabulary immediate post-tests. In the third phase (week 6), the semi-structured interviews were conducted upon the completion of the four-week treatment and the immediate post-tests. Three participants from each group were invited to take part in the interviews, which were conducted in the same computer lab while it was empty. Each participant was interviewed individually for about 10 minutes.

In the last phase, which lasted for four successive weeks (from week 7 to week 10), the WM tests and the delayed vocabulary post-tests were conducted. 12 participants from each experimental group were selected to complete the four different WM tests. Each participating student was tested individually for about 30 minutes in a vacant teacher’s office to avoid any possible distractions form the surrounding environment. During the final week of this study (week 10), all participants in the four treatment groups completed the one month vocabulary delayed post-tests. Participants across the four groups were provided with a full explanation of the purpose of the study in both Arabic and English.

Table 3.6. Overview of the research procedures

Week	Group 1 (VAC)	Group 2 (VA)	Group 3 (CA)	Group 4 (A only)
1	Introduction + consent form + vocabulary pre-tests			
2-5	Video 1 (the power of lionesses) Video 2 (The wonderful dogs) Video 3 (Kangaroo) Video 4 (Elephant queen) + Vocabulary immediate post-tests			
6	Semi-structured interviews			
7 - 10	WM tests vocabulary one-month delayed post-tests			

3.10 Research instruments

This study used a variety of instruments in the process of data collection. Quantitative data was collected mainly through three vocabulary tests and four working memory tests. Qualitative data was generated using semi-structured interviews. The study developed a range of data sources to address the different research questions posed. A detailed discussion of each data source is provided below.

3.11 Vocabulary tests

Vocabulary knowledge entails knowledge of interrelated, yet separate aspects (Nation, 2001). Measuring all knowledge aspects in a single study is not possible, as Schmitt (2010) pointed out. The present study therefore tested the participants' receptive and productive knowledge of *form* and *meaning* of the target words. This is because *form* and *meaning* are the first dimensions that need to be learned before other dimensions. As Sonbul and Schmitt (2010, p. 254) stressed, "the first step in learning, a new word is likely to entail mainly acquisition of its form–meaning link". In the same vein, Pellicer-Sánchez and Schmitt (2010) maintained that testing receptive and productive mastery of word forms and meanings provides sensitive measures to small uptakes in word knowledge as well as demonstrates the incremental nature of vocabulary knowledge.

As the study adopted Nation's (2001) framework of vocabulary knowledge as a theoretical basis to guide the design of vocabulary tests, it developed three vocabulary tests: (1) spoken form recognition, (2) spoken meaning recall, and (3) spoken meaning recognition. The main aim of using multiple assessments of word knowledge was to give a wider picture of what word knowledge entails and to accurately assess the extent of the learning that had resulted from the audio-visual inputs (Webb, 2005). Every vocabulary test was particularly appropriate for certain sources of data. That is, the spoken form recognition test was designed to measure spoken form receptive knowledge of the target words. The spoken meaning recall test was used to measure productive knowledge of meaning. While the spoken meaning recognition test was aimed at testing receptive knowledge of meaning. To put it differently, each test was developed to enable the subjects to productively or receptively demonstrate a particular dimension of word knowledge. An additional value for using multiple measures of the target words is to explore whether the different modalities of audio-visual input (VAC, VA, CA and A only) lead to different kinds of word learning. A number of previous studies have reported that different input modalities resulted in different kinds of vocabulary learning (Brown et al., 2008; Sydorenko, 2010).

The three vocabulary tests were administered over three different occasions as: pre-tests (a week before the start of the intervention), as immediate post-tests (immediately upon the completion of the four-week treatment sessions), and as delayed post-tests (one month after the end of the four-week treatment). The reason for conducting the pre-tests a week before the start of the treatment was to counteract the risk that learners may make connections between the pre-tests and the treatment. Nation and Webb (2011, p. 277) warned that a pre-test may alert the participants "to the fact that an experiment is being conducted. This awareness can then change learners' behaviour". This in turn could result in deliberate vocabulary learning. The rationale for administering the delayed post-tests one month after the end of the treatment was based on Schmitt's (2010) argument that the longer time interval between the immediate post-test and delayed post-test, the better it showed how long a word could be retained. Schmitt (2010, p. 157) argued that there is not a standard period of delay, but he suggested that three weeks or more "should be indicative of learning which is stable and durable".

The test items used in the three test administrations were the same except the inclusion of a larger number of potential target words as well as some non-target words in the pre-tests. The reason for including a larger number of potential target words was to allow the researcher to choose the least known words as target words, while the inclusion of some non-target items was meant to divert the learners' attention from the target words in the experiment (Nation & Webb, 2011).

In addition, the three same sets of tests (form recognition, meaning recall, and meaning recognition) were used across the three test administrations (pre-immediate-post-and-delayed-post-tests). This was done to avoid the risk that differences in test scores across the different test times could be due to of the use of different test formats across the study (Nation & Webb, 2011). In this case, a researcher cannot ascertain whether the differences in the pre-immediate-and-delayed-post-tests scores are a result of the treatment or due to of the use of various test formats. Furthermore, the prompts or tested items were presented orally one by one and the participants were instructed to report their answers on a provided sheet of paper. An English native speaker recorded the test items. Following a number of listening and audio-visual studies, including Peters et al. (2016); Peters and Webb (2018); Rodgers (2013); and Van Zeeland and Schmitt (2013), the aural form of the test items was employed because the subjects who were involved in the two non-captioning conditions (VA and A only) did not see the written forms of the words. This necessitated the importance of providing a test format that is congruent with the way the participants received the input, as this would maintain the reliability of the data (Brown et al., 2008). In light of the findings of the first pilot study, each item was repeated two times with a pause of about 10 seconds between the items presentations. At the end of each test, the participants were also given time to check their answers.

The three tests were presented in a strict order: 1) the form recognition test, 2) the meaning recall test and 3) the meaning recognition test. The tests were sequenced in this way to prevent the potential effect of the earlier tests affecting the performances on the later tests (Nation & Webb, 2011; Van Zeeland & Schmitt, 2013; Webb, 2005). Moreover, the three tests were in decontextualized formats because learning cannot be confidently ascribed to the experiment conducted if there is a chance that context facilitates the guessing of the meaning of the target stimulus during test time (Waring & Takaki, 2003). Participants were given clear verbal and written instructions in English and in Arabic on the test procedure. They were also explicitly instructed not to attempt to randomly guess their responses.

To ensure validity of the three tests, a number of steps were taken. The first step was to have the tests evaluated by two experts (both native English speakers). It was ensured that the tests were effective for measuring the vocabulary knowledge aspect that they were purported to measure. Further, the evaluators were asked to examine the following issues concerning the multiple-choice test: (1) all options of tested items were from the same part of speech; (2) the options were not semantically related; (3) they were consistent in terms of their structures. The evaluators were also asked to check the clarity of the instructions of the tests. The comments received from the evaluators helped improve the tests. The second step taken to examine the validity of the tests was to pilot test them with participants who had similar language proficiencies and L1 backgrounds to the target participants (this will be described in greater detail in section 3.15). The following is a detailed description of the three tests.

3.11.1 Spoken form recognition test

The spoken form recognition test, also known as a yes/no or checklist test, has become popular since 1982 and is found to be a reliable and valid measure (Nation, 2001). It measures the participants' recognition of the phonological form of the target words. The spoken form recognition test format has been used by a number of similar studies (for example, Montero Perez et al., 2014; Montero Perez et al., 2018; Peters et al., 2016; Sydorenko, 2010). It is an unprompted recognition test that was administered by presenting the participants with an oral recording of a random list of target words that they had been exposed to in the videos, plus some additional non-words as distractors. The inclusion of the non-words served to measure the accuracy of the subjects' responses, because if subjects chose a non-word as known, then this shows that they "[were] overstating their vocabulary knowledge" (Nation, 2001, p. 346). Additionally, non-words were used to control for guessing (Nation & Webb, 2011; Pellicer-Sánchez & Schmitt, 2012). Following the suggestion of Vidal (2011) and (Nation & Webb, 2011), the nonwords amounted to 33 % of the total of the number of target words (*n* = 36) in this research. This translated to 12 nonwords for the 36 target words. The 12 non-words were taken from the study of Waring and Takaki (2003), for the reason that they have been validated and adopted in different studies (the list of the non-words are provided in Appendix F). For example, the word '*year'* has been rendered '*jurg'* in its substitute form.

In this test, participants were presented with the test items on the recording, each preceded by a number, and they were given an answer sheet/grid that showed the numbers that preceded the test items (see Appendix G, for an example of the test answer sheet). They were asked to tick either '*yes'* if they heard the words in the videos or '*no'* if they did not. Each test item was repeated twice with about a 10-second silence between the item presentations. The test required the subjects to determine if they had heard the words when listening to the videos by ticking 'Yes' or 'No'. 'Yes' responses to target words reflected the participant's vocabulary knowledge while 'Yes' responses to non-words reflect the participant's tendency to guess.

One of the serious limitations of this test is random or blind guessing, that is, it is possible that some subjects did overstate their vocabulary knowledge by ticking 'yes' to vocabulary items that they did not know (Mochida & Harrington, 2006; Pellicer-Sánchez & Schmitt, 2012). It was thus crucial to control for any blind guessing for the results to be valid (Harrington & Carey, 2009). In the literature, researchers have identified four different methods that can be used to correct for guessing in a yes/no test or a checklist test (the spoken form recognition) (Huibregtse, Admiraal, & Meara, 2002; Pellicer-Sánchez & Schmitt, 2012). However, Pellicer-Sánchez and Schmitt (2012) remarked that researchers did not reach a consensus on the best way to correct for guessing.

After considering the available existing methods, I decided to adopt the method of Anderson and Freebody (1983), which is known as 'correction for guessing' (cfg). This method of correction was chosen because it has been applied in some similar incidental vocabulary studies (for example, Feng, 2017). Table 3.7 illustrates an example for how the formula was applied to correct for guessing in this study. According to this formula, a 'yes' response to a real word is given one point and is labelled as a 'hit'. While a 'yes' response to a non-word is also given one point and marked as a 'false alarm'. The following formula was used to calculate the proportion of truly learned words p (K):

P (K) = (hit) – p (false alarm) / 1 – p (false alarm)

Table 3.7 Example of the scores on the spoken form recognition test before and after applying the cfg formula

participant	Unadjusted score	Adjusted score
1	14	11
2	18	17

3.11.2 Spoken meaning recall test

The spoken meaning recall test was run to measure the participants' productive knowledge of meaning. This test measured their ability to recall rather than recognize the meanings of the target words. The test format has been used in many incidental vocabulary studies, including Van Zeeland and Schmitt (2013), Peters et al. (2016), Montero Perez et al. (2014), and (Peters & Webb, 2018) and has been found appropriate for measuring productive knowledge of meaning. In this test, students were presented with each target word on the recording, each preceded by a number, and were required to report any kind of knowledge about the meaning either in English or Arabic on the given answer page. The knowledge may include a synonym, an explanation, or an L1 equivalent. An "I don't know" option was purposively included to reduce random guess work. Like the yes/no test, each prompt/target word was repeated twice with a 10-second pause before moving to the next one. The subjects were provided a sheet of paper with a table of three columns. The first column on the left shows the digits that preceded the target words in the recording. The middle column is allocated for writing the meaning of the target words. The last column was for the “I don’t know” option (see Appendix H, for an example of the test answer sheet).

3.11.3 Spoken meaning recognition test

In this receptive test, meaning recognition was tested through multiple-choice items. This multiple-choice test was a prompted recognition test. The choice of alternatives was informed by the guidelines of Nagy et al. (1985). Nagy et al. identified three levels of difficulty for the distractors in a multiple-choice test, namely, the highest level; the intermediate level; and the lowest level. These three levels differ with regards to the degree of similarity between the tested items and the concepts represented by the alternatives. This study opted for the intermediate level of difficulty, as it was sensitive to partial meaning knowledge. This meant that although distracters were from the same part of speech, they were carefully drawn from fairly diverse semantic sets to detect partial knowledge of meaning. The distracters were also chosen from higher-frequency levels (K1 list and K2 list) to avoid the problem of some distracters being harder than the tested items (Nation & Webb, 2011). The order of the correct answers was spread roughly equally amongst the distractors in order to avoid creating a pattern that the learners could easily detect. The test-takers were presented with an oral list of the target words, each preceded by a number and were instructed to circle the words with the nearest meaning to the stem (target word) on the answer page provided (see Appendix I, for an example of the test answer sheet). Following Rodgers (2013), an “I don’t know” choice was added to control for the effect of guessing. This test format has been employed by a number of incidental vocabulary studies, such as Brown et al. (2008), Montero Perez et al. (2014), Peters and Webb (2018) and Rodgers (2013).

3.11.4 Scoring scheme

The three vocabulary tests were scored dichotomously. That is, (0) point was given for incorrect and black responses and (1) point was awarded for correct responses (as long as the meaning was not fully distorted). This scoring criterion was based on Peters et al. (2016); Montero Perez et al. (2014); and Van Zeeland and Schmitt (2013). An inter-rater reliability check was carried out on the meaning recall test; this will be discussed in section 3.14.1.

3.12 Measures of working memory

In this study, WM was measured by four memory span tasks drawn from Automated Working Memory Assessment (AWMA) (Alloway, 2007). This computerised test battery AWMA, consists of

12 automated tests, is designed specifically to address WM performance according to the multicomponent WM model of Baddeley and Hitch (1974). Some of the tests are considered simple tasks (e.g. forward digit recall, dot matrix), as they only involve sequential recall of presented items, while others that require processing of the "to-be-recalled" presented information are referred to complex tasks (e.g. backward digit, the odd one out) (Alloway et al., 2008). Due to time restrictions, the study administered four tests only.

Two simple span tasks (forward digit, the dot matrix) and two complex spans (backward digit and the odd one out) were used. The verbal/phonological WM capacity was measured by two digit tasks: the simple verbal forward digit recall test and the complex verbal backward digit recall test. The justification for using these two digit tasks rather than the other tasks, such as reading span or listening span, is related to the fact the digit tasks require less knowledge of the language (English) than the other verbal tasks (Juffs & Harrington, 2011). Juffs and Harrington (2011, p. 141) argued that performance on tasks, such as reading and listening span requires participants to have prior knowledge of the language, which "thus confound[s] to some extent [verbal WM] capacity and language knowledge". The visuospatial WM capacity was tested through the dot matrix and the odd-one-out tests. The inclusion of these visuospatial measures was intended to provide a comprehensive investigation of whether the visuospatial WM system was associated with incidental vocabulary learning from the multiple modalities of audio-visual input. This study adopted multiple measures to gauge both verbal and visual WM components in order to provide a reliable assessment of the participants' WM capacities for each WM system

There are a number of reasons this battery test (AWMA) was chosen. Firstly, the measures of this test battery have been found to be valid and reliable (Alloway et al., 2008; Alloway et al., 2006; Dehn, 2008) and have been used in many studies that involve typical and atypical individuals. In addition, the different measures of this test battery are suitable for measuring WM capacity of individuals aged between 18-22 years, which was the age range of the participants of the present study. Lastly and more importantly, since administration and scoring of the tasks were fully automated which provide consistency in presenting the stimuli across learners; this minimizes the experimenter errors and inconsistency (Alloway et al., 2008).

As explained in the procedure, the study randomly selected 12 participants from each treatment group to voluntarily complete the four memory tests. Each participant was tested individually in an empty teacher's office. Each individual testing session lasted for approximately 30 minutes. Following the administration guidelines of the test battery developer, the tests were sequenced in the following order: forward digit recall test, backward digit recall test, dot matrix test, and the odd one out test. The tests were scored based on the scoring guidelines of the test battery, as will be highlighted below. Participants were permitted practice trials in order to familiarise themselves with the tests. Instructions on how to respond to the tests were given in both Arabic and English to ensure that the participants fully understood what they have to do. All the test items were presented on the screen or spoken by the computer to the participants in English. During the tests, participants were given a break if they needed. Below is a brief description of the four memory tasks.

3.12.1 Forward digit recall

At the beginning of this test, the experimenter (the researcher) explained to the participant what he has to do and how to respond to the test. In this test, the computer verbally presented a sequence of digits and the participant had to recall them in the correct sequence without manipulation. Test trials started with a block of one digit and progressed to a block of nine digits. The test automatically ended when the participant was no longer able to accurately repeat the digits in their correct order or when he made more than three errors within a block of trials. The examinees received a correct score for each trial if they recalled each number in the correct order of presentation.

3.12.2 Backward digit recall

This task started with practice trials to ensure that the participants understand how to respond to the test. In this task, the participants were required to recall a series of random digits in a reverse/backward order of presentation. The test stopped when they made more than three errors within a block of trials. This test differs from the forward digit test in the sense that it imposes a substantial processing demand on the participants, as they have to recall the digits in the reverse sequence (Alloway & Alloway, 2010). The participants were given a score if they correctly recalled the digits in the correct backward order of presentation of the test items.

3.12.3 Dot Matrix

In this task, the participant was shown a matrix of 16 boxes with a red dot in one box. The red dot moved around a series of 4x4 matrices one after the other. At the end of the test trial, empty boxes were shown and the participant had to remember where the red dot was and point to it, in exactly the same order as he saw them. Test trials started with a block of one red dot and increased to a block of nine dots. The test automatically ended when the participant made more than three errors within a block of trials. Similar to the forward digit recall test, the test-taker was given a correct score for each trial when each dot was recalled in the correct order of presentation of the dots.

3.12.4 The odd-one-out

In the odd-one-out task, the participants were shown three shapes presented in a horizontal row of boxes and were required to identify the odd-one-out shape. At the end of each test trial, the participants were asked to recall the location of the odd shapes in the correct order of the presentation by pointing to the blank boxes on the computer screen. The test started with a block of one set of shapes and increased to a block of seven sets of shapes. Like the previous tests, the test came to an end when the subject made more than three errors within a block of trials. The participants received a correct score when he recalled the position of each odd one out shapes in the correct order.

3.13 Semi-structured interviews

Interview is a very popular data collection method in qualitative research (Dörnyei, 2007; Richards, 2009). Research interview has been described as “a data collection method that offers different ways of exploring people’s experience and views” (Richards, 2009, p. 183). Similarly, Cohen et al. (2007) contended that interview is perceived as an auditory conversation between the interviewer and the interviewee, employed to elicit information from the interviewee about a certain target theme. Dörnyei (2007, p. 134) added that interviews are used “to obtain descriptions of the life world of the interviewee with respect to interpreting the meaning of the described phenomena”. In qualitative research, there are three main forms of research interviews: structured interviews, semi-structured interviews, and unstructured interviews (Dörnyei, 2007; Mackey & Gass, 2005; Richards, 2009). These three types of interviews differ in their degree of structure. Semi-structured interview is the most commonly adopted type of interview in applied linguistic research, which Dörnyei (2007) described as a ‘compromise’ between structured and unstructured interviews. It needs to be reiterated that the aim of using this qualitative instrument in this purely quantitative study was to provide explanations for the results generated by the quantitative tests. In particular, there were instances where the quantitative instruments could not provide insightful explanations for some unexpected results; however, the interviews were useful in explaining such results.

With the intention of uncovering the subjects’ experiences on learning through different modalities of audio-visual materials, as a qualitative method, semi-structured interview has been given preference over other types of interviews in this study for the following reasons. Firstly, although in semi-structured interview the themes and topics are somewhat pre-determined and a number of pre-set questions are prepared in advance, the interviewer is flexible to allow the interviewee to open up new areas. Richards (2009) pointed out that semi-structured interviews are used when the researcher has a good overview of the issues being studied and that he/she wants to look at them in more depth. A further reason is that though semi-structured interview is flexible in its structure, which means the interviewer can change and modify the questions, this type of interview can provide reliable and comparable data. Moreover, using a semi-structured interview gives the researcher the freedom to digress and probe for further details. This means that the researcher can elicit additional information using follow-up questions to either clarify or add on vague or incomplete responses. In this research, semi-structured interview was used to serve the following aims: to explore the participants' perspectives towards incidental vocabulary learning through the different modalities of audio-visual input and to uncover the difficulties the subjects faced during the running of the treatment.

Although Dörnyei (2007) suggested that researchers conducting interview studies should recruit between six to ten participants, in this study, due to the time limit, it was not possible to invite more than three participants from each experimental group. Therefore, a total 12 subjects from the four groups were randomly chosen to take part in interviews. A semi-structured interview was held on a one-to-one, face-to-face basis. The choice of one-to-one interviews was based on the significance of this format in establishing a relaxed environment where every participant has the chance to speak freely. In addition, this type of interview was preferred over the focus group to ensure that the elicited information was not influenced by other participants' responses. In addition, this format allowed the researcher to probe the interviewee's point of view. Interviews were conducted in both English and Arabic, i.e., interviewees had the choice to use English or Arabic to respond to the interview questions. The interviewees were allowed to use their first language, Arabic to remove any barriers or difficulties encountered resulting from the use of the target language, English. The beginning of the interviews was devoted to reflecting on the materials presented to refresh the participants' memories and to establish a good rapport. An interview guide was developed (see Appendix J). The questions consisted of both close-ended and open-ended questions, which revolved around the topics and themes related to the topic under examination.

3.14 Validity and reliability

The relationship between reliability and validity is complex. Cohen et al. (2007) argued that though reliability is essential in research, it is not a sufficient condition for validity. In contrast, validity is not essential, however it is a sufficient condition for reliability. In other words, Bryman

(2016, p. 162) pointed out that the two measures, validity and reliability are "analytically distinguishable" but they are related. Dörnyei (2007) contended that validity and reliability are considered the quality criteria in quantitative research. They are of great importance in any research as without them, research findings may be rendered inauthentic (Cohen et al., 2007). Although no research can utterly eliminate the threats to reliability and validity, evaluating and assessing them can reduce such threats (Cohen et al., 2007). In the present study, a number of steps have been taken to ensure the validity and reliability of the data collection methods and the study design, including the two pilot studies, which were mainly carried out to ensure that the study is valid and reliable (see section 3.15). In what follows, a discussion of the steps taken to establish validity and reliability is outlined.

3.14.1 Reliability

Reliability is a key concept that has attracted considerable attention from several methodological discussions. Bryman (2016, p. 156) described the term reliability as "fundamentally concerned with issues of consistency of measures". A reliable measure is a one that gives similar outcomes when it is administered on another occasion to a similar group of subjects (Cohen et al., 2007). There are three main approaches for establishing reliability in quantitative research: test-retest reliability, internal reliability, and inter-rater reliability (Bryman, 2016). Test-retest reliability is a measurement of the stability of a method, which involves administering the method twice over a time span, to a similar group of participants. For the method to be reliable, it must show a high correlation between the two administrations. Cohen et al. (2007) stressed that choosing an appropriate time period between the two test administrations is a critical factor to be considered as it affects the performance on the test administrations. That is, the time span between the two test administrations must be neither too short nor too long. One of the key problems associated with this approach is that according to Schmitt (2000), participants may not always be keen on sitting the test twice.

The second type of reliability is the internal consistency that is tested through "the spilt-half method" (Bryman, 2016, p. 157). The split-half method means that a test is divided into two equal halves and administered only once. The split-half reliability is obtained by calculating the correlation between the total scores of the two-halves. Scoring high on both halves means that the test is reliable. Despite the easiness of a single administration, this procedure is problematic in relation to making closely equivalent forms (two-half) of each test. Since vocabulary tests are item-based, whereby each item represents different constructs, knowledge of one test item does not necessarily mean that the other test items are known (Schmitt, 2010).

The inter-rater reliability is the third type of reliability that relates to the degree to which two or more test-raters agree on their assessment judgement (Bryman, 2016). The type of method is used when subjective decisions in some cases can take place. Due to the limitations associated with the first two types of reliability (test-retest and internal reliability), the present study assessed the third type, inter-rater reliability. The inter-rater reliability was carried out on the scoring of the vocabulary meaning recall test. The other vocabulary two tests (form recognition and meaning recognition) were not a matter of concern, as they can only be objectively marked as a right or wrong answer. To check inter-rater reliability, two experienced Arabic-speaking English language teachers were given random exam sheets from each treatment group to mark them. A percent agreement between the three raters (including the researcher) was done and a high inter-rater reliability was found (93%).

3.14.2 Validity

Validity is the other prominent quality criterion of quantitative research. It is a requirement for successful research because if a study is invalid, it is thus worthless (Cohen et al., 2007). Creswell (2012, p. 159) defined validity as "the degree to which all of the evidence points to the intended interpretation of test scores for the proposed purpose". In other words, it concerns “whether a measure of a concept really measures that concept” (Bryman, 2016, p. 158). Evaluating validity is crucial as it gives integrity to the results of the inquiry (Bryman). As with the reliability, there are different ways of testing validity. Cohen et al. (2007) identified 18 distinct types of validity, examining them all would be beyond the scope of any research. In this study the main relevant types of validity were assessed namely, content validity, face validity and construct validity, and these have also been evaluated in similar studies,.

The first type is content validity which relates to whether the study instrument comprehensively covered the items that meant to be covered (Cohen et al., 2007). This can be done by ensuring that the items selected for the investigation are fairly representative and that they are deeply addressed. In this study, to ensure this type of validity, it was decided that four authentic videos would be chosen and that the vocabulary items used would be drawn evenly from the four videos.

The study also examined face validity which is considered the basic type of validity that concerns the correspondence between a study instrument and the concept under investigation (Bryman, 2016). Bryman explained that a researcher can establish face validity by asking experts in the field of investigation to determine whether the measure is appropriate for testing the studied concept. In this research, a fellow PhD researcher (at the University of Southampton) was asked to look at the study materials and instruments and determine whether on the face of them they reflected the investigated concept (vocabulary). She confirmed that the instruments and materials (videos) seemed valid for measuring vocabulary learning.

The third type of validity that was examined in this study is construct validity, which concerns "the degree to which the research adequately captures the construct of interest" (Mackey & Gass, 2005, p. 108). In the present study, construct validity was enhanced by using multiple vocabulary tests designed specifically to assess the different vocabulary knowledge constructs. In addition, the different components of WM were assessed through multiple tests that have already been proven to be valid.

3.15 Pilot studies

The pilot study is of special importance to any research study. It plays a significant role in establishing the reliability and validity of research instruments and research design. Mackey and Gass (2005, p. 43) defined a pilot study as "a small-scale trial of the proposed procedures, materials, and methods, and sometimes also includes coding sheets and analytic choices". The values of conducting a pilot research study are stressed in the following points. Firstly, a pilot study is useful in identifying the appropriate levels of the materials adopted in a study (videos in the current study). In addition, it is through a pilot study, that researchers could explore the adequacy and feasibility of the data collection methods. Another tangible point for conducting a pilot study is that it brings to light any potential faults in and threats to the materials, tasks, procedures, and instruments in order to address them before running the main study (Mackey & Gass, 2005; Nation & Webb, 2011). The present study therefore conducted two pilot studies; the first pilot study was carried out in the UK at the University of Southampton, three months before the main study and the second one took place in Saudi Arabia at King Abdulaziz University, three weeks prior to the main study (see Table 3.8 below).

Table 3.8 An overview of the pilot studies

Pilot study	Number of participants	Time	Place
Pilot study 1	17	October 2016	Southampton
Pilot study 2	20	January 2017	Jeddah

3.15.1 Pilot study one

In light of the values of carrying out a piloting study, the aim of conducting the first pilot study was threefold. First, it aimed to decide on the appropriateness of the materials (videos) selected for the main study in terms of matching the target learners' language proficiency level, watch-ability of the videos (the clarity of the videos, pronunciation, and the pace), and the enjoy-ability and the suitability of the content. This pilot study also aimed to check the feasibility and the level of difficulty of the data collection methods, the three vocabulary tests and WM tests. Finally, it aimed to explore the timeframe needed for completing the different stages and tasks of the study and to refine the main study procedures.

This pilot study consisted of 17 Saudi EFL learners with English language competence similar to the participants in the actual treatment who were recruited by word of mouth. The participants' language proficiency level was regarded as intermediate based on the relevant institutions' language proficiency tests. The participants were undertaking intensive language courses at different language institutes across Southampton, UK (see Table 3.9). The participants were divided into four groups and were randomly assigned to the four videos (Kangaroo Comeback: N = 3, The Wonderful Dogs: N = 5, Elephant Queen: N = 5, The Power of Lionesses: N = 4). This uneven assignment of subjects was based on the fact that the participants were drawn from different educational establishments, and it was thus difficult to find time that suited everyone. Two of the 17 participants were also invited to complete the working memory tests.

Table 3.9. The number of participants involved in the first pilot study

Language schools	Number of participants
Lewis Language School	8
Language School at Southampton University	7
Southampton Language College	2

This pilot research has yielded several useful and interesting findings. As far as the materials were concerned, three key issues emerged from this pilot study. First, the vast majority of the participants felt that the length of the videos (30 minutes each) was too long. Although most of the subjects confirmed that videos were interesting, they felt that they could not maintain their concentration because of the length of the videos. Unfortunately, this might have let some of them to experience a moment where they felt bored with long viewing. Based on this, it was concluded that reducing the watching-time to 15 minutes was an essential step to be considered. In addition, the findings of the study revealed that the materials were suitable for the participants' current proficiency level. The subjects also confirmed that they understood the videos, as the videos were watchable in terms of clarity and speed of delivery of speech. Regarding the three vocabulary tests, the pilot study revealed some concerns. One of the key issues that have emerged in this piloting was that the participants asked me to repeat the tested items two times before moving forward to the next item. This had affected the flow of the tests. Another point that the pilot study brought to light was the allotted time needed for answering each tested item. A pause of about 10 seconds between the presentations of test items was deemed necessary. Thus, it was decided that the test items will be repeated twice with a 10-second silence between the item presentations.

Overall, this pilot study was useful in casting some light on the aforementioned concerns. Based on the above findings, it was decided to address these issues and modify them accordingly. It was thus felt important to carry out a further pilot study after addressing the concerns raised in this pilot study to ensure the success of the main study.

3.15.2 Pilot study two

Three weeks prior to the start of the actual experiment, a second pilot study was conducted in Saudi Arabia with a group of students from the English Language Institute (ELI) at the King Abdulaziz University that hosted the main study. This group of 20 students was not involved in the main study. This purpose of this small pilot study was to explore whether the updated length of the videos (15 minutes) was appropriate. The time interval between the test items presentation was piloted to make sure that the updated time was sufficient for completing the tests. It also intended to identify any potential problems in the design of the treatment conditions (VAC, VA, CA, and A only). The participants were divided into four groups of 5 and randomly assigned to a treatment condition. Each group of participants was asked to watch and listen to one of the four videos, which was assigned randomly to that treatment condition. They were also asked to complete the three vocabulary tests for that video. The participants reported that the length of the videos was suitable for them. They also confirmed that repeating the test items twice with a 10-second silence between the items presentations was enough for them to complete the test. Finally, no major concerns were found regarding the design of the treatment conditions.

3.16 Ethical issues

Ethics is a crucial issue in any type of research. Cohen et al. (2007) mentioned that ethical concerns may surface at any phase of the study because of the instruments used, the subjects targeted, and the procedures of the study. The present study complied with the ethical guidelines of the University of Southampton. The guidance explicitly stresses that the participants should be made aware of their rights to confidentiality, anonymity and protection from harm. In line with this, at the outset of study, participants were provided with Information Sheet and Consent Form to read and sign. It was ensured that they understood their rights. They were also assured that their data will be absolutely protected, that is, it will only be used by the researcher. It should be mentioned that in keeping with the incidental learning protocol, participants were not informed about the true purpose and nature of the study (i.e. incidental vocabulary learning) at the beginning of the research intervention, instead they were told that the study aimed to examine the opportunities of improving their general English skills through viewing and listening to different audio-visual materials. This was done to avoid alerting them to the unknown/new words in the materials and to maintain the element of surprise for the vocabulary post-tests. However, upon the completion of the treatment, participants were provided with a full account of the nature and purpose of the study.

One of the ethical concerns was that since intact classes were used for the study, it was particularly important to ensure that the participants were aware that their participation was voluntarily and they had the right to not partake in the research. Participants' self-determination to participate in the study without coercion was completely respected and protected. Further, they were not requested to provide their names in order to maintain anonymity. However, for those who decided to give their names, it was ensured that their names would be replaced with pseudonyms when dealing with the data. In addition, subjects were notified about their right to withdraw from the study at any stage of the experiment without repercussion. Moreover, subjects were reassured that their participation in the study would not affect their course grades. Additionally, it was explained that none of the study procedures or tasks would bear any major risks for the subjects. Rather, participating in the study could be considered as a benefit, as the study would provide additional opportunities for language learning. However, in order to account for fatigue, which was to be expected, the researcher provided the participants with free refreshments during the experiment.

3.17 Data analysis

Data analysis is a crucial stage in any research, which is defined by Denscombe (2007, p. 247) as "the search for things that lie behind the surface content of the data". It is indeed the most important and difficult part of research process. This section presents the procedures followed in preparing, organising, analysing, representing, and interpreting the obtained data.

There are two types of statistics: descriptive statistics and inferential statistics (Creswell, 2012; Dörnyei, 2007). The present study conducted both types of statistical analysis and provided a rationale for using each type. Descriptive statistics refers to the description of trends in data. Descriptive statistics aims to provide an indication of general tendencies of data, the spread of scores and a comparison of the relationship between the scores (Creswell, 2012). According to Dörnyei (2007), the value of using descriptive statistics rests on the fact that it gives a summary of sets of numerical data in order to save time and space. Yet, Dörnyei contended that researchers using descriptive statistics cannot draw general conclusions from their sample to the population.

On the other hand, inferential statistics comes into play when, as Creswell (2012, p. 182) maintained " compar[ing] two or more groups on the independent variable in terms of the dependent variable". Inferential statistics allows researchers to generalize the results of sample to the whole population (Dörnyei, 2007). Creswell (2012) proposed three procedures for conducting inferential statistics: (1) testing hypothesis and calculating the *p* values, (2) setting the confidence interval in order as Dörnyei (2007) commented, to provide a range of scores that can be deemed significant for the population, and (3) computing effect sizes to provide information about the strength of an observed phenomenon. Prior to conducting the inferential statistics, the study checked the key assumptions of the parametric tests (ANVOA, T-Test, Pearson) to see whether the data allowed for parametric or non-parametric tests.

In analysing and interpreting quantitative data, Creswell (2012) identified a number of steps, namely, preparing and organising data (scoring data and checking/cleaning data), selecting a statistical software, inputting data into the chosen software, analysing data, representing data (using tables, figures etc.) and validating data. The current study followed all the stages of Creswell's guidelines for analysing quantitative data. Firstly, all vocabulary tests were scored. Next, the obtained scores were initially put into Excel spreadsheets and later transferred to the SPSS software. The inputted data were then checked and cleaned for any errors. This was done through a visual inspection of the data grid and conducting frequency distribution. Finally, the study carried out an appropriate analysis for each research question.

There were two key purposes of conducting statistical analysis in the current study. The first purpose was to examine the mean differences between variables. In other words, this statistical analysis was conducted to examine the effects of the independent variables (different modalities of audio-visual input and frequency of occurrence) on the dependent variable (incidental vocabulary learning and retention). The second objective was to look at the relationship between WM of the participants and their incidental vocabulary learning and retention through the different modalities of input. The study used SPSS software (versions 24/25) for running the required statistical analysis to answer the seven research questions.

As stated above, the study used three vocabulary tests (spoken form recognition test, spoken meaning recall test, and spoken meaning recognition test) as primary methods for quantitative data collection to answer research questions one to four. It also administered four WM tests: forward digit recall, backward digit recall, dot matrix, and the odd one out. The term '*learning*' was adopted to refer to the participants' learning outcomes on the immediate post-tests. While the term *'retention'* was used to describe the participants' outcomes on the delayed post-tests. Mean gain was calculated by subtracting the scores on the pre-tests from the scores on the immediate post-tests.

The following chapter presents the findings from the vocabulary tests

Chapter 4 Vocabulary findings

This chapter draws upon the quantitative data gathered from the three vocabulary tests, the spoken form recognition test, the spoken meaning recall test, and the spoken meaning recognition test. These were conducted at three different times as, pre-tests, immediate-post-tests, and delayed-post-tests. It is recalled that the first part of the present study focused on the effects of the independent variables (different modalities of audio-visual input and frequency of occurrence) on the dependent variable (vocabulary development). The different modalities of input are regarded as a nominal independent variable. This nominal variable has four levels, or categories: (1) video, audio, and caption (VAC), (2) video and audio (VA), (3) caption and audio (CA), and (4) audio only (A only). The dependent variable, vocabulary development, is treated as a continuous variable. This chapter reports the findings of the statistical analysis of the three vocabulary tests. It is divided into four sections that correspond to the research questions pertaining to vocabulary learning and retention. Where applicable, the qualitative data collected through semi-structured interviews are considered to support the quantitative findings.

4.1 Section one: research question 1

R-Q 1: To what extent do audio-visual conditions, VAC, VA, CA, and A only, enhance/promote L2 incidental vocabulary short-term learning of spoken form recognition, meaning recall, and meaning recognition?

The purpose of this section is twofold. The first purpose is to examine the differences between the scores achieved by the four experimental groups (VAC, VA, CA and A only) on the three vocabulary pre-tests, to see whether the four groups differed significantly in their level of pre-existing knowledge of the 36 target words. In order to achieve this aim, inferential tests were conducted on the three vocabulary pre-tests. The second purpose of this section is to examine the effects of the different modalities of audio-visual input on L2 incidental learning of spoken form recognition, meaning recall and meaning recognition knowledge of the target words to explore whether the four input conditions resulted in incidental vocabulary uptakes of these three vocabulary knowledge constructs. To this end, a within-subject analysis was carried out to compare the participants' scores on the pre-tests with their scores on the corresponding immediate post-tests in each treatment group, VAC, VA, CA, and A only. Descriptive statistics was also carried out, in order to provide an initial insight into the nature of the data and to determine the learning of the three vocabulary knowledge aspects, spoken form, meaning recall, and meaning recognition, by the four treatment groups. This was to show the general trend of changes in the participants' performance from the pre-tests to the immediate post-tests. The analyses of the data of the first research questions are presented in the following sub-sections.

4.1.1 Differences in vocabulary scores on the pre-tests (between-groups comparison)

The aim of this analysis was to find out whether the four experimental groups (VAC, VA, CA, and A only) differed significantly in their prior knowledge of the spoken form recognition, meaning recall, and meaning recognition of the target words. This was to ensure that the four groups were equivalent with regards to their pre-existing knowledge of the selected target words. It was necessary to carry out this analysis to find out whether the participants' pre-existing knowledge would be a confounding variable in the results of the dependent measures. It was also considered necessary to examine whether the four experimental groups had the same level of the target word knowledge before the treatment, owing to the fact that the present study adopted a quasi-experimental design, wherein there was no random assignment of the participants into the experimental groups. Before running the inferential statistics, the data was examined to ensure whether it allowed for parametric or non-parametric statistics. One of the key assumptions of the parametric analysis of variance (ANOVA) is the normality of the distribution of the data, which concerns whether the distribution of sample means (across the independent samples) is normal (Field, 2009). There is a variety of ways of to assess this assumption, however the tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk) make a more objective decision on this assumption (Lowie & Seton, 2012). The Kolmogorov-Smirnov test was thus run, as it is the most suitable and reliable normality test when the sample size is larger than 50 subjects. If the test produces $p > .05$, then the normal distribution of data is assumed (Field, 2009). The results of the Kolmogorov-Smirnov test (reported in Table 4.1) showed that the data did not meet the normality of distribution assumption of the parametric one-way ANOVA test. Therefore, the non-parametric equivalent, Kruskal-Wallis H test was carried out. This test was chosen as it is the most suitable test for comparing the performance of three or more groups on a continuous variable (Pallant, 2016). The results of the Kruskal-Wallis H tests on the three vocabulary pre-tests are presented below.

Table 4.1 The Kolmogorov-Smirnova normality test for the three vocabulary pre-tests

		Kolmogorov-Smirnova		
	Groups	Statistic	df	Sig.
spoken form recognition pre-test	VAC	0.199	26	.010
	VA	0.171	27	.040
	CA	0.184	30	.011
	A only	0.231	25	.001
spoken meaning recall pre-test	VAC	0.378	26	.000
	VA	0.409	27	.000
	CA	0.289	30	.000
	A only	0.279	25	.000
spoken meaning recognition pre-test	VAC	0.297	26	.000
	VA	0.194	27	.010
	CA	0.212	30	.001
	A only	0.275	25	.000

Three Kruskal-Wallis H teste were conducted for the comparison of the participants' pre-tests scores across the four experimental groups. The outcomes of the Kruskal-Wallis H tests are presented in Table 4.2. The findings of the Kruskal-Wallis H tests revealed that across the four groups the participants' scores on the spoken form recognition pre-test were not statistically significantly different, $H(3) = 4.801$, $p > .187$. Similarly, the findings showed that the groups' scores on the spoken meaning recall pre-test and on spoken meaning recognition pre-test were not significantly different, $H(3) = 7.164$, $p > .067$ and $H(3) = 1.280$, $p > .734$, respectively. These results could be taken to suggest that the four experimental groups had approximately similar level of pre-existing knowledge of the 36 target words before the start of the experiment. Therefore, it can be safely assumed that the participants' performance on the post-tests will not be affected by their prior knowledge of the target words.

Table 4.2 Kruskal-Wallis H tests for the three vocabulary pre-tests

	spoken form recognition pre-test	spoken meaning recall pre-test	spoken meaning recognition pre-test
Kruskal-Wallis H	4.801	7.164	1.280
df	3	3	3
Asymp. Sig.	.187	.067	.734

4.1.2 Scores on the pre-tests and on the immediate-post-tests (within-subjects comparison)

This part presents the descriptive statistics that show the participants' scores on the three vocabulary pre-tests and on the immediate post-tests. Differences in scores within each group from the pre-tests to the immediate post-tests were tested by a Wilcoxon Signed Rank test (as the data was not normally distributed, see Table 4.1). This nonparametric test, which is regarded as an alternative to the parametric repeated measure T-test, was employed because it compares the participants' scores at two-test times (pre and immediate post) (Mayers, 2013; Pallant, 2016). The effect size was calculated using the Pearson's r method. It was worked out as follows: r = z score divided by the square root of the sample size over the two test time points ($r = Z/\sqrt{N}$) (Mayers, 2013; Pallant, 2016). The study followed Cohen's (1992) guidelines for the interpretation of effect size (small .10, medium, .30, large .50). In what follows, the results of the spoken form recognition test are presented first, followed by the results of the spoken meaning recall and spoken meaning recognition tests. Figure 4.1 presents the mean scores of the four experimental groups on the pre and immediate post spoken form recognition test. The data in Table 4.3 shows the mean scores, the standard deviations for the pre-and-immediate-post spoken form recognition tests per the four treatment groups as well as the Wilcoxon Signed Rank test.

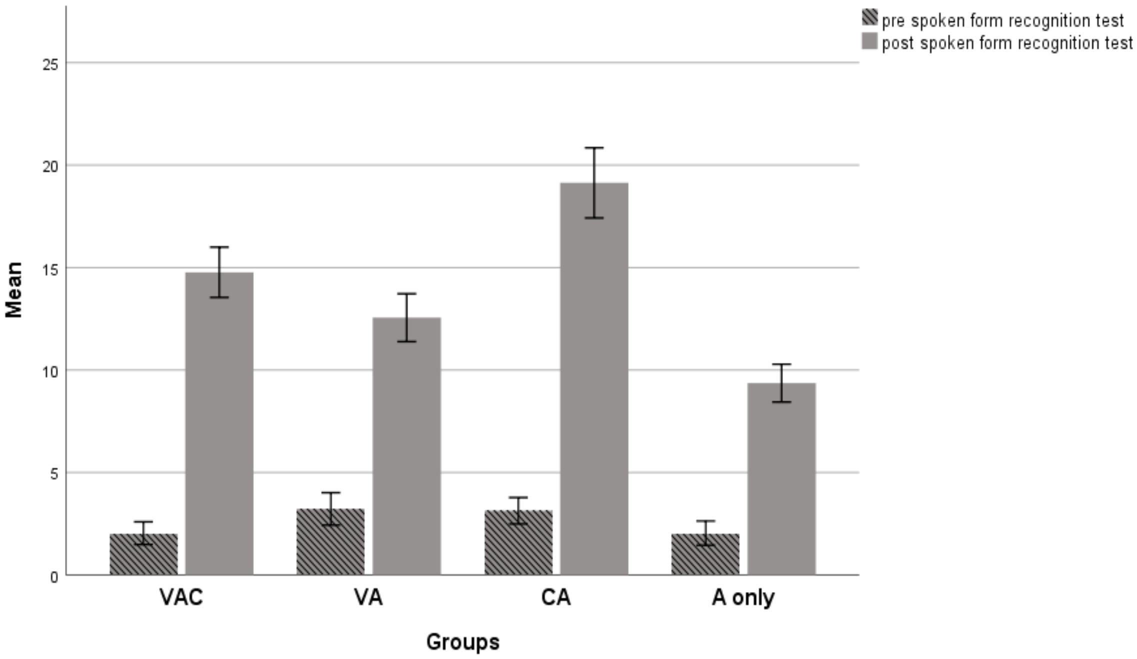


Figure 4.1 Mean scores for the spoken form recognition test of the four treatment groups over pre-and-immediate-post-test administrations

Table 4.3 Mean scores for the spoken form recognition test of the four treatment groups over pre-and-immediate-post-test administrations

	pre-test		immediate-post			Sig. (2-tailed)
	mean	SD	mean	SD	z	
VAC	2.04	1.371	14.77 (41.02 %)	3.037	-4.471[b]	.000
VA	2.70	1.489	12.56 (34.88%)	2.953	-4.550[b]	.000
CA	2.43	1.194	19.13 (53.13%)	4.592	-4.789[b]	.000
A only	2.04	1.428	9.36 (26%)	2.234	-4.386[b]	.000
Total	2.31	1.378	14.18 (39.38%)	4.916		

Note. Maximum score is 36

As shown in the figure and table above, the scores of the four experimental groups were low on the pre-test. This suggests that the subjects in the four groups had minimum knowledge about the target vocabulary items before the treatment. However, after being involved in the four treatment sessions, their knowledge improved, as evidenced in the immediate post-test. Comparing the four groups' overall scores on this test (M = 14.18) with the scores on the spoken meaning recall and spoken meaning recognition tests (M = 6.57%, M = 10.81, respectively), it is clearly evident that the spoken form knowledge aspect was the best developed. This may indicate that this receptive knowledge facet (spoken form recognition) is easier to acquire than the other two facets of word knowledge (meaning recall and meaning recognition). The Wilcoxon Signed Rank test showed that there was a significant increase in the scores from the pre-test to the immediate post-test of the VAC group, $z = -4.471$, $p < .000$, $r = -.62$. A similar pattern of results was also found for the VA group, $z = -4.550$, $p < .000$, $r = -.61$; for the CA group, $z = -4.789$, $p < .000$, $r = -.61$; and for the A only group, $z = -4.386$, $p < .000$, $r = -.62$. The strength of the effect in all cases was large. These results seem to suggest that the four audio-visual input conditions contributed significantly to incidental short-term learning of the spoken form recognition knowledge aspect of the target words.

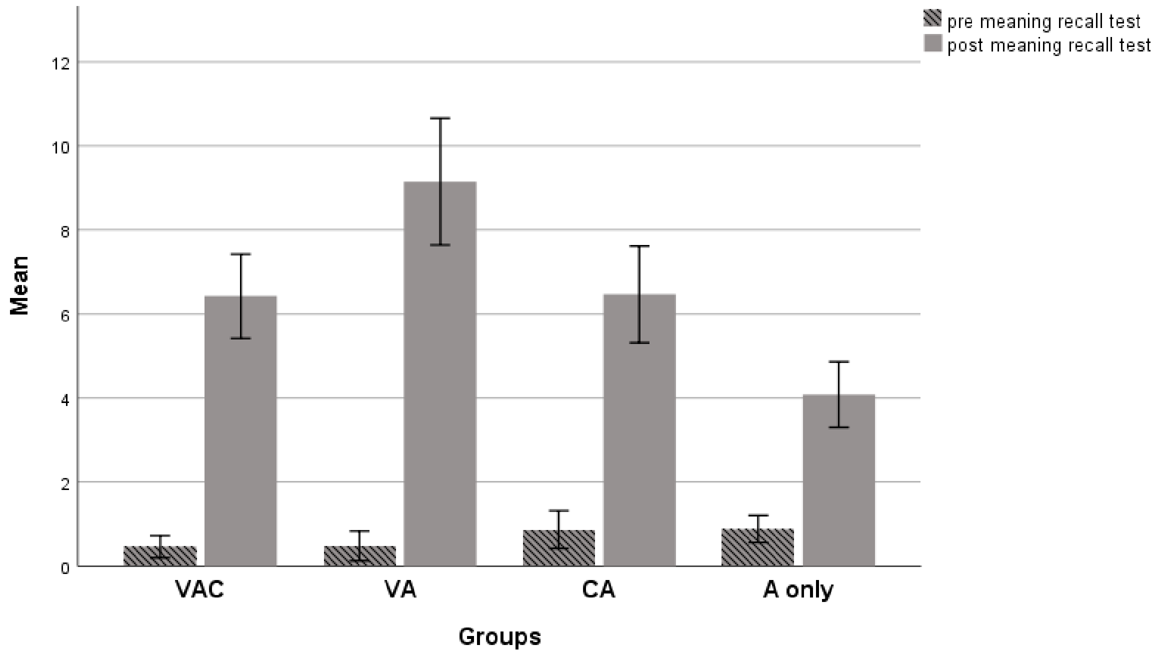


Figure 4.2 Mean scores for the spoken meaning recall test of the four treatment groups over pre-and-immediate-post-test administrations

Table 4.4 Mean scores for the spoken meaning recall test of the four treatment groups over pre-and-immediate-post-test administrations

groups	pre-test		immediate-post			Sig. (2-tailed)
	mean	SD	mean	SD	z	
VAC	0.46	0.647	6.42 (17.83%)	2.485	-4.468[b]	.000
VA	0.48	0.893	9.15 (25.41%)	3.81	-4.548[b]	.000
CA	0.87	1.196	6.47 (17.97%)	3.082	-4.792[b]	.000
A only	0.88	0.781	4.08 (11.33%)	1.891	-4.246[b]	.000
Total	0.68	0.926	6.57 (18.25%)	3.388		

Note. Maximum score is 36

Table 4.4 and Figure 4.2 illustrate the participants' scores on the spoken meaning recall pre-and-immediate-post-tests. Similar to the findings of the spoken form recognition test, the data in data the table and figure above showed that the participants had little prior knowledge of the target word meanings. However, their knowledge improved significantly after the four-week treatment. Contrary to spoken form recognition knowledge, which was the best improved knowledge, across the four groups meaning recall knowledge was the least acquired. This suggests that productive word knowledge (meaning recall) was the hardest to improve in comparison to the other two receptive knowledge dimensions (spoken form recognition and meaning recognition). The table above also shows that the scores on the pre-and-immediate-post-tests for the participants involved in the VAC group, $z = -4.468$, $p < .000$, $r = -.61$ were significantly different. Similar patterns of results were also found for the participants involved in the VA group, $z = -4.548$, $p < .000$, $r = -.61$; in the CA group, $z = -4.792$, $p <.000$, $r = -.61$; and in the A only group, $z = -4.246$, $p < .000$, $r = -0.60$. The effect sizes were large in all cases. These data suggest that meaning recall knowledge of the 36 target words of the four groups increased significantly after the research treatment.

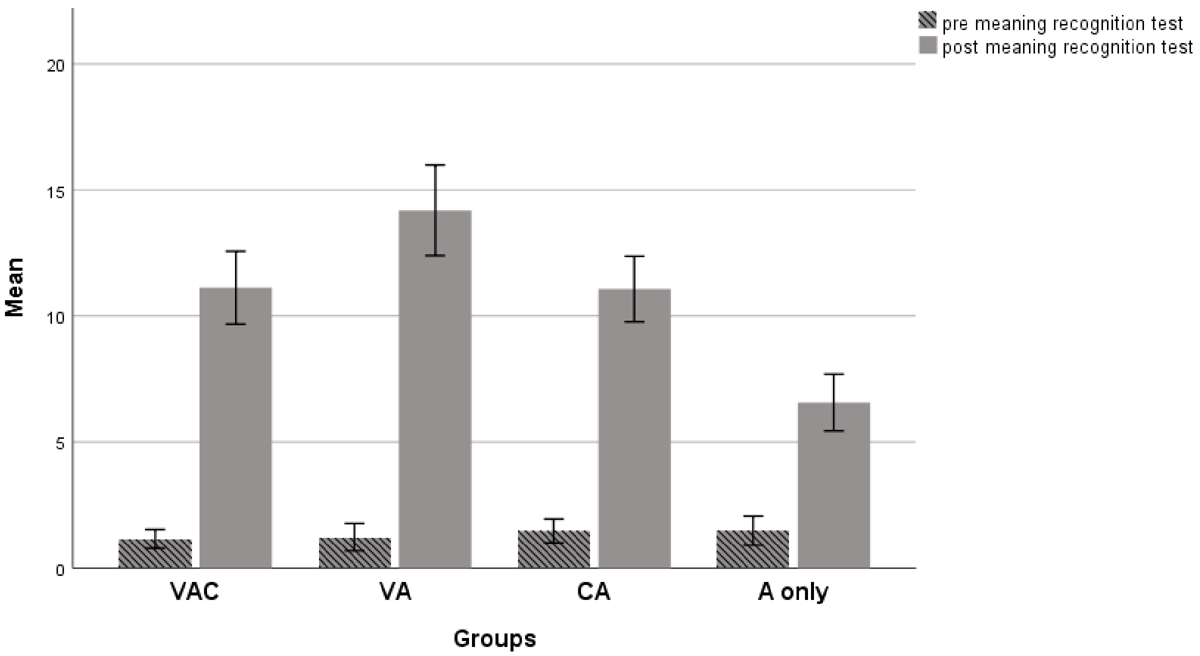


Figure 4.3 Mean scores for the spoken meaning recognition test of the four treatment groups over pre-and-immediate-post-test administrations

Table 4.5 Mean scores for the spoken meaning recognition test of the four treatment groups over pre-and-immediate-post-test administrations

groups	pre-test		immediate-post		z	Sig. (2-tailed)
	mean	SD	mean	SD		
VAC	1.15	0.925	11.12 (30.88%)	3.57	-4.463[b]	.000
VA	1.22	1.368	14.19 (39.41%)	4.541	-4.544[b]	.000
CA	1.47	1.252	11.07 (30.75%)	3.473	-4.794[b]	.000
A only	1.48	1.388	6.56 (18.22%)	2.725	-4.355[b]	.000
Total	1.33	1.238	10.81 (30.02%)	4.479		

Note. Maximum score is 36

Table 4.5 shows the mean scores, standard deviation, and the Wilcoxon Signed Rank test. This is also represented visually in Figure 4.3. The descriptive data represented in Table 4.5 and Figure 4.3 demonstrated that the participants across the four groups had little pre-existing knowledge of meaning recognition, as reported in the pre-test. However, the participants' scores significantly improved after the research intervention. The results also showed that meaning recognition knowledge was the next best acquired dimension of the three knowledge dimensions tested, with an overall mean of ($M = 10.81$) for the four groups. These outcomes may suggest that acquiring receptive meaning knowledge of words (recognition of meaning) is easier than acquiring productive knowledge (recall of meaning). The outcomes of the Wilcoxon Signed Rank test showed that the scores of the participants in the VAC group on the pre-test were significantly smaller than their scores on the immediate post-test, $z = -4.463$, $p < .000$, $r = -.61$. Similar results can be seen for the VA group, $z = -4.544$, $p <.000$, $r = -.61$; for the CA group, $z = -4.794$, $p <. 000$, $r = -.61$; and for the A only group, $z = -4.355$, $p <.000$, $r = -.61$. In all cases the size of the effect was strong. These data suggest the four treatment groups significantly improved their meaning recognition knowledge of the 36 target words after the four-treatment sessions.

4.1.3 Summary of key findings

These above results presented further evidence that incidental vocabulary learning occurs through the different modalities (VAC, VA, CA, and A only) of audio-visual input. All groups improved significantly their vocabulary knowledge to a certain degree. The key findings revealed in this sections were:

- There were no significant differences between the four groups' pre-tests scores; suggesting that the four groups had similar level of pre-existing vocabulary knowledge before the start of the research intervention;
- Participants' vocabulary knowledge improved significantly from the pre-tests to the immediate post-tests;
- All the four input conditions resulted in incidental vocabulary learning;
- Receptive knowledge of words (spoken form and meaning recognition) improved better than productive knowledge (meaning recall).

The following section presents the findings of the second research question, which addressed the differential effects of the different modalities of audio-visual input on incidental vocabulary short-term learning.

4.2 Section two: research question 2

R-Q2: Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary short-term learning of the spoken form recognition, meaning recall, and meaning recognition?

After establishing that the four different modalities of audio-visual input promoted L2 incidental learning of the spoken form recognition, meaning recall, and meaning recognition, in order to determine whether the four audio-visual input conditions had differential effects on short-term learning of these aspects, it is important to compare the effects of the four experimental conditions on the learning of the three vocabulary knowledge aspects. To this end, inferential statistical analyses were carried out. The analysis of variance (ANOVA) is the most feasible test to use, in order to explore whether the differences found in the mean scores of the four groups were statistically significant, as the scores on the three immediate post-tests are taken as a single dependent variable and the four experimental conditions are taken as a single independent variable (Mayers, 2013). However, prior to conducting this test, the key assumptions of this test, the normality of distribution of data and homogeneity of variances, were checked. To examine the normality of distribution assumption, the Kolmogorov-Smirnov normality test was employed. The results (reported in Table 4.6) showed that the data was approximately normally distributed for all the four groups (VAC, VA, CA, and A only) across the three vocabulary immediate post-tests ($p > .05$). Therefore, this assumption was fulfilled.

Table 4.6. Tests of Normality

Tests	groups	Kolmogorov-Smirnova		
		Statistic	df	Sig.
spoken form recognition immediate post-test	VAC	.146	26	.164
	VA	.108	27	.200*
	CA	.134	30	.180
	A only	.093	25	.200*
Spoken meaning recall immediate post-test	VAC	.143	26	.183
	VA	.145	27	.150
	CA	.157	30	.056
	A only	.153	25	.133
Spoken meaning recognition immediate post-test	VAC	.102	26	.200*
	VA	.132	27	.200*
	CA	.126	30	.200*
	A only	.116	25	.200*

The second assumption of the parametric ANOVA test is the equality of variance, which refers to the equality of the spread of scores in different groups of the study (Field, 2009). To meet this assumption, the *p* value should be larger than .05. Levene's test, the most widely used test to assess this assumption, was therefore run (Field, 2009). The findings of the Levene's test reported in Table 4.7 showed that the assumption of homogeneity of variances was not met in the spoken form recognition test ($p < .001$) nor in the spoken meaning recall test ($p < .010$), yet it was met in the spoken meaning recognition test ($p > .117$). Based on these observations, this assumption was not assumed in this part of analysis, however violation of this assumption could be overlooked for the following two reasons. Firstly, ANOVA test produces post-hoc tests (for example, Games-Howell) that can be used when the data is non-homogeneous. In addition, robust tests of equality of means (i.e. Welch ANOVA and Brown-Forsythe) can be consulted when the assumption of homogeneity is violated (Field, 2009; Mayers, 2013; Pallant, 2016). Since the assumption of homogeneity of variances was not ensured, the current study examined the Welch ANOVA test to assess the potential bias that the violation of this assumption would have had on the findings. The Welch ANOVA test yielded the same conclusions of the standard one-way ANOVA regarding the statistical significance of the mean differences between the groups in the three vocabulary tests (the outputs of the Welch ANOVA tests of the three vocabulary tests are given in Appendix K). Therefore to answer the second research question, it was decided to run the standard one-way ANOVA, as it is the mostly widely known and used test. To follow up on a significant ANOVA, a Games-Howell post-hoc test was run.

Table 4.7. Test of Homogeneity of Variances

Test	Levene Statistic	df1	df2	Sig.
Spoken form recognition immediate post-test	5.61	3	104	.001
Spoken meaning recall immediate post-test	4.015	3	104	.01
Spoken meaning recognition immediate post-test	2.01	3	104	.117

However, to control for the risk of Type I error that may result from running multiple comparisons on the same dependent variable, a Bonferroni adjustment was applied (Field, 2009; Mayers, 2013; Pallant, 2016). This was performed by dividing the alpha rate (.05) by the number of the statistical analyses the study had to make. In this study, since there were four treatment groups, the post-hoc test was carried out six times and the alpha value (.05) was thus divided by 6 (.05/6). Thus, any observed *p*-value below the adjusted *p*-value .008 ($p < .008$) is statistically significant.

To draw a more valid conclusion from those results, the present study assessed the magnitude of the effect of the experiment, as the *p*-value in itself does not provide information about the strength of the effect (Lowie & Seton, 2012). This study, therefore, used the partial eta squared (ηp^2) as an effect size measure, because it is the default effect size measure produced by ANOVA tests in SPSS. In addition, this effect size measure (ηp^2) has been used in several similar studies (for example, Montero Perez et al., 2014; Peters & Webb, 2018). The study also consulted the thresholds of Cohen (1988) for interpreting the strength of the effect size. According to Cohen, ηp^2= .010 - .039 is a small effect; .050 - .110 is a medium effect; and .140 – 200 is a large effect. In the following sections, the results of the spoken form recognition test are presented first, followed by the results of the spoken-meaning recall test and the spoken-meaning recognition test.

4.2.1 Findings of the spoken form recognition immediate post-test

This part analyses participants' performance on the spoken form recognition immediate post-test under the four treatment conditions. Table 4.8 presents the findings of the one-way ANOVA test on the spoken form recognition immediate post-test. As seen in Table 4.8, the participants' scores on the spoken form recognition test were statistically significantly different, $F (3, 104) = 40.752$, $p < .000$, with a moderate effect size ($\eta p^2 = .054$). These results suggest that the four input conditions have differential effects on the learning of spoken form recognition knowledge. A Games-Howell post-hoc test was carried out in order to find out where the differences between the four groups laid specifically. As mentioned above, this post-hoc test was preferable than the conventional measures such as Bonferroni's and Turkey's, as the assumption of equality of variance was not satisfied (Field, 2009; Mayers, 2013).

Table 4.8. A one-way ANOVA test for the spoken form recognition immediate post-test

<table>
<tr><th colspan="2"></th><th>Sum of Squares</th><th>df</th><th>Mean Square</th><th>F-value</th><th>Sig.</th><th>Partial Eta Squared ηp^2</th></tr>
<tr><td rowspan="3">spoken form recognition test</td><td>Between Groups</td><td>1397.15</td><td>3</td><td>465.716</td><td>40.752</td><td>.000</td><td>.054</td></tr>
<tr><td>Within Groups</td><td>1188.51</td><td>104</td><td>11.428</td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td>2585.66</td><td>107</td><td></td><td></td><td></td><td></td></tr>
</table>

Summarised data from the Games-Howell post-hoc test is provided in Table 4.9. The findings of the Games-Howell post-hoc test showed that significant group mean differences were found between the four groups, with the CA group gaining the highest mean score, whereas the A only group obtained the lowest score. The participants' scores under the CA condition were significantly higher than the scores of the participants involved in the VAC condition (MD = 4.36, $p < .001$); the VA condition (MD = 6.58, $p < .000$); and the A only condition (MD = 9.77, $p < .000$). These findings suggest that this input condition (CA) was the most effective for enhancing spoken form recognition knowledge compared to the other three input conditions. It was also found that students involved in the VAC condition gained a significantly higher mean score than those in the A only group (MD = 5.41, $p < .000$). Similar results were further revealed in the comparison between the mean differences between the VA group and the A only group (MD = 3.19, $p < .000$). To summarise, these results imply that the multiple input conditions (VAC, VA, and CA) were more effective for aiding short-term learning of the spoken word forms than the single input mode, A only. The findings also suggest that the captioning groups (CA and VAC) were more effective than the non-captioning groups (VA and A only) for fostering incidental learning of spoken form recognition knowledge.

Table 4.9. Summarised data from the Games-Howell post-hoc test for the spoken form recognition test

Groups	CA - VAC	CA - VA	CA - A only	VAC - A only	VA - A only
Mean Difference (MD)	4.36	6.58	9.77	5.41	3.19
Sig	.001	.000	.000	.000	.000

4.2.2 Findings of the spoken meaning recall immediate post-test

It was established that the four experimental groups performed differently on the spoken meaning recall immediate post-test, however it was necessary to examine whether the differences found between the groups' means were statistically significant. For this purpose, a one-way ANOVA test was conducted (reproduced in Table 4.10). The findings of the ANOVA test indicated that there were statistically significant differences among the four groups in the mean scores on the spoken meaning recall immediate post-test, $F(3, 104) = 13.017$, $p < .000$. The strength of the treatment effect was small, ηp^2 =. 027. These results seem to suggest that the four different modalities of audio-visual input have differential effects on learning word meanings at a productive level. In order to examine which group was superior, a post-hoc test was needed. A Games-Howell post-hoc test was thus run.

Table 4.10. A one-way ANOVA test for the spoken meaning recall immediate post-test

		Sum of Squares	df	Mean Square	F-value	Sig.	Partial Eta Squared ηp²
spoken meaning recall test	Between Groups	335.347	3	111.782	13.017	.000	.0273
	Within Groups	893.06	104	8.587			
	Total	1228.41	107				

The Games-Howell test results, displayed in Table 4.11, revealed that the mean scores of the VA, CA, and VAC groups were significantly higher than the mean scores of the A only group (MD = 5.06, $p <.000$; MD = 2.38, $p <.005$; and MD = 2.34, $p <.002$, respectively). This outcome signifies that the multimodal input conditions, VAC, VA, and CA, were superior to the single input mode, A only in promoting meaning recall knowledge. Different from the findings of the spoken form recognition test, the participants in the audio-visual group (VA) gained significantly higher scores than the non-visual input condition (CA) on the spoken meaning recall immediate post-test (MD = 2.68, $p < .006$). This seems to suggest that the visual images of the videos were useful for learning about the meanings of the target words. This finding is endorsed by the qualitative data as one of the interviewees from the visual group (VA) explicitly stated: "*I learned most of the new word meanings from pictures of the videos*" (S1). Unlike this experimental condition (VA), the participants in the CA condition may have had to infer the meanings of the novel words from contextual clues. Qualitative data showed that students in the CA condition might have resorted to contextual clues to deduce the meanings of the novel words. For example, when asked how they dealt with the target lexical items they met while reading while listening, the interviewees stated:

> I can learn the word '*joey*' based on the position of the word in the context, to analyse what sort of definition this word has. I thought it's the position where the word was that gave me the idea or the concept of what does the word mean (S1).

"What comes before and after the unknown word helps me to learn about the meaning of that word" (S2).

"I understood the meanings of the new words from the context of the words. I focused on the sentence to help me learn the meanings" (S3).

Another interesting finding that emerged was that the non-captioning visual VA condition was more effective for meaning recall than the other captioning visual condition (VAC) (MD = 2.72, $p < .007$). This finding may indicate that the non-captioned videos are more effective for enhancing the learning of meaning. This finding is also echoed by the qualitative data. Some interviewees in the VAC group explicitly stated that they focused on the caption first and then looked at the images in the videos, while the audios were almost neglected. For example, some interviewees said: *"The availability of the caption in the videos has pros and cons, as it has distracted me from focusing on the images of the video*" (S1). *"My focus was paid to captions, and then the videos and then the audios; however audios were a bit fast*" (S2).

Table 4.11. Summarised data from the Games-Howell post-hoc test for the spoken meaning recall test

Groups	VA - VAC	VA – CA	VA - A only	CA - A only	VAC - A only
Mean Difference (MD)	2.725	2.681	5.068	2.387	2.343
Sig	.007	.006	.000	.005	.002

4.2.3 Findings of the spoken meaning recognition immediate post-test

A one-way ANOVA was employed to examine whether the four audio-visual groups differed significantly in their learning rates of meaning recognition knowledge of the target words (Table 4.12). As can be seen from the table below, the four treatment groups were significantly different from each other in their gain scores on the spoken meaning recognition immediate post-test, $F(3, 104) = 19.143$, $p < .000$, $\eta p^2 = .035$. The size of the effect was small. This finding points out that the four different input conditions have a differential effect on the recognition of word meanings. However, to explore where exactly the significant differences existed, a Games-Howell post-hoc test was utilized.

Table 4.12. A one-way ANOVA test for the spoken meaning recognition immediate post-test

		Sum of Squares	df	Mean Square	F-value	Sig.	Partial Eta Squared ηp^2
spoken meaning recognition test	Between Groups	763.542	3	254.514	19.143	.000	.0356
	Within Groups	1382.76	104	13.296			
	Total	2146.3	107				

Comparing between the four groups, the findings reported in Table 4.13 showed that the learning gains of the multimodal learning conditions, VA, VAC, and CA, in the spoken meaning recognition test were significantly greater than those of the single input condition, A only, (MD = 7.62, p <.000; MD = 4.55, p <.000; and MD = 4.50, p <.000, respectively). This implies that the single input condition was the least effective of the three multimodal input conditions for promoting the meaning recognition knowledge. In addition, the mean score of the VA group was significantly higher than that of the VAC group (MD = 3.07, p < .001) and that of the CA group (MD = 3.11, p < .003). These outcomes suggest that students in the VA condition improved their receptive knowledge of word meanings better than their counterparts in the other treatment groups. Hence, it can be said that this audio-visual input condition (VA) seems to be the most effective condition for enhancing the meaning recognition knowledge of the 36 target words.

Table 4.13. Summarised data from the Games-Howell post-hoc test for the spoken meaning recognition test

Groups	VA - VAC	VA – CA	VA - A only	VAC - A only	CA - A only
Mean Difference (MD)	3.070	3.119	7.625	4.555	4.507
Sig	.001	.003	.000	.000	.000

4.2.4 Summary of key findings

This section analysed and addressed the second research question, which investigated the differential effects of the four modalities of audio-visual input on incidental short-term learning of spoken form recognition, meaning recall, and meaning recognition of the target word. The most interesting results found were:

- The multimodal input conditions (VAC, VA, and CA) were more effective for promoting incidental vocabulary learning of the three knowledge aspects than the single input condition, A only;
- The captioning conditions (CA and VAC) were more effective for enhancing spoken form recognition knowledge than the non-captioning conditions (VA and A only);
- The visual input condition (VA) proved superior to the other conditions in fostering meaning recall knowledge and meaning recognition knowledge.

The following section presents the findings of the third research question, which looked at the differential effects of the four modalities of audio-visual input on incidental vocabulary long-term retention.

4.3 Section three: research question 3

R-Q 3: Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary long-term retention of the three vocabulary knowledge aspects, spoken form recognition, meaning recall, and meaning recognition?

This section has two aims. Firstly, it aims to examine the degree to which the initial gains reported in the immediate post-tests were retained until four weeks after the end of the treatment. To achieve this aim, a within-subjects analysis was run to compare the participants' scores on the immediate post-tests with their scores on the corresponding delayed post-tests in each treatment group (VAC, VA, CA, and A only). As explained before, the participants' scores on the one-month delayed post-tests were treated as long-term retention, while their scores on the immediate post-tests were interpreted as short-term learning. The second aim is to compare the effects of the four modalities of audio-visual input on the long-term retention of the spoken form recognition, meaning recall, and meaning recognition knowledge aspects of the target 36 words. To this end, a between-groups comparison was conducted to determine whether the students' retention scores across the four groups were statistically significantly different from each other, and if so, which input condition was more effective for the retention of each word knowledge aspect. Relevant assumptions (i.e. normality of distribution, homogeneity of variances) were tested prior to running inferential tests. The effect size was calculated using the Pearson's r method and was interpreted according to Cohen's (1992) guidelines (small .10, medium, .30, large .50).

4.3.1 Scores on the immediate post-tests and on the delayed post-tests (within-subjects comparison)

In this section, descriptive statistics showing the participants' scores on the immediate post-tests and on the delayed post-tests are presented. Differences in the scores within each experimental group from the immediate post-tests to the delayed post-tests were tested through a Wilcoxon Signed Rank test, because the data was not normally distributed (Pallant, 2016).

4.3.1.1 Findings of the spoken form recognition delayed post-test

In order to examine the effects of the different modalities of audio-visual input on long-term word retention, students' scores on the spoken form delayed post-test was compared to their scores on the spoken form immediate post-test for each group. Figure 4.4 displays the trend of changes in the students' scores over the two test times, the immediate post and the delayed post-test.

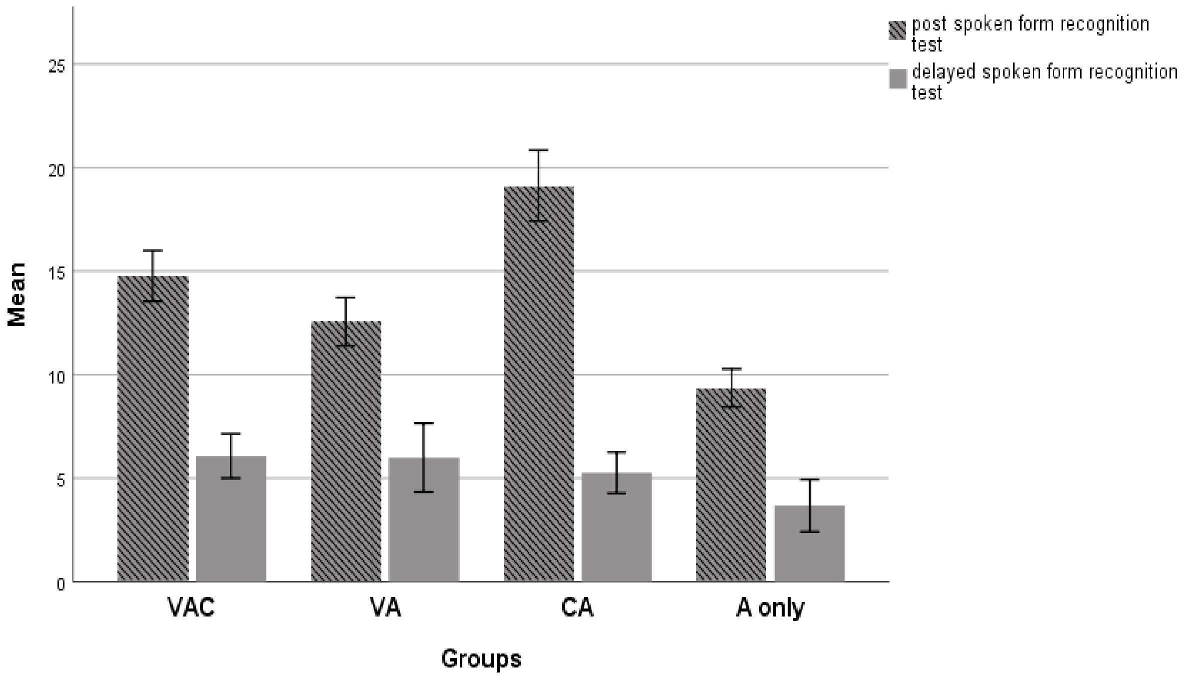


Figure 4.4. Mean scores for the spoken form recognition test of the four treatment groups over immediate-and-delayed-post-test administrations

Looking at Figure 4.4, it is obvious that there was a change in the students' scores on the spoken form recognition test over the two test times, immediate post-test and delayed post-test, irrespective of the treatment conditions. That is, students across the four groups showed a considerable decrease in their mean scores from the immediate post-test to the delayed post-test. Precisely, one month following the completion of the treatment, more than 60 % of their initial learning gains were lost on average. These findings seem to suggest that the spoken form knowledge aspect is not stable and susceptible to attrition. A Wilcoxon Signed Rank test was thus carried out to further explore whether these differences in the scores on the immediate and delayed post-test for each group were statistically significant.

Table 4.14. A Wilcoxon signed rank test for the spoken form recognition test

Groups		spoken form recognition delayed post-test – spoken form recognition immediate post-test
VAC	Z	-4.467[b]
	Asymp. Sig. (2-tailed)	.000
VA	Z	-4.320[b]
	Asymp. Sig. (2-tailed)	.000
CA	Z	-4.788[b]
	Asymp. Sig. (2-tailed)	.000
A only	Z	-4.293[b]
	Asymp. Sig. (2-tailed)	.000

The Wilcoxon Signed Rank test (reported in Table 4.14) revealed that there was a significant decline from the immediate post-test to the delayed post-test in the scores of the VAC group, $z = -4.467$, $p < .000$, $r = .62$. Significant decreases were also found for the VA group, $z = -4.320$, $p < .000$, $r = .58$; for the CA group, $z = -4.788$, $p < .000$, $r = .61$; and for the A only group, $z = -4.293$, $p < .000$, $r = .60$. The effect sizes were strong in all cases. Overall, these data suggest that large attrition of the initial gains of spoken form receptive knowledge has occurred after one month of the end of the treatment.

4.3.1.2 Findings of the spoken meaning recall delayed post-test

Figure 4.5 below presents the mean scores for the spoken meaning recall test over two times: immediate post-test and the one-month delayed post-test of the four experimental groups.

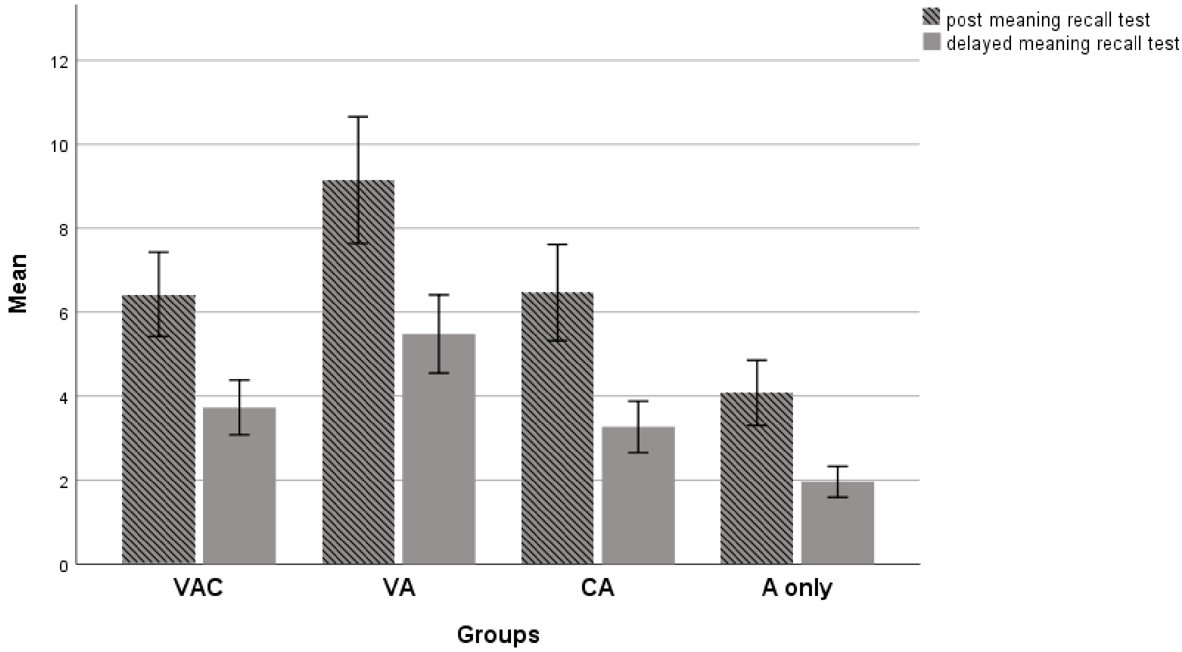


Figure 4.5. Mean scores for the spoken meaning recall test of the four treatment groups over two immediate-and-delayed-post-test administrations

What Figure 4.5 shows is that the vocabulary scores on the immediate post-test were greater than those on the delayed post-test across the four experimental groups. Similar to the findings of the spoken form recognition test above, it is evident that across all groups, the gain rates dropped noticeably, that is, nearly half of the learned target words have been forgotten, in a period of one month. For example, the scores on the immediate post-test of the VAC group, M = 6.42, went down to M = 3.73, which translated into an attrition rate of about 42%. Similar patterns were also found for participants in the VA and CA groups who lost an average of 41% to 50% of the learned words. On the whole, the data showed that an average of 50 % of the learned meanings at the immediate post-test were lost at the one-month delayed post-test. These results could be taken to suggest that even productive knowledge (meaning recall) is not decay-resistant. A Wilcoxon Signed Rank test was employed to test the students' changes in scores on the meaning recall test across the two test times, from the immediate post-test to the delayed post-test to see if they were significantly different for each group of participants.

Table 4.15. A Wilcoxon signed rank test for the spoken meaning recall test

Groups		spoken meaning recall delayed post-test – spoken meaning recall immediate post-test
VAC	Z	-4.486[b]
	Asymp. Sig. (2-tailed)	.000
VA	Z	-4.567[b]
	Asymp. Sig. (2-tailed)	.000
CA	Z	-4.811[b]
	Asymp. Sig. (2-tailed)	.000
A only	Z	-4.053[b]
	Asymp. Sig. (2-tailed)	.000

As can be seen from the table above, significant differences between the immediate post-test scores and the delayed post-test scores were found for the VAC group, $z = -4.486$, $p < .000$, $r = .62$. Likewise, the participants' scores under the VA condition on the delayed post-test were significantly smaller than their scores on the immediate post-test, $z = -4.567$, $p <. 000$, $r = -.62$. Similar results were also found in the comparison between the scores on the immediate post-test and the delayed post-test of students involved in the CA condition, $z = -4.811$, $p < .000$, $r = - .62$ and in the A only condition, $z = -4.053$, $p < .000$, $r = - .57$. Across all cases, the effect sizes were strong. These findings suggest that across the four groups, the participants' scores on the immediate post-test were statistically significantly higher than their scores on the delayed post-test, indicating that significant attrition of the initial learning gains of meaning recall knowledge has taken place over a one-month period.

4.3.1.3 Findings of the spoken meaning recognition delayed post-test

Descriptive statistics are provided in Figure 4.6 to assess the retention rates of meaning recognition knowledge of the 36 target words by the four experimental groups.

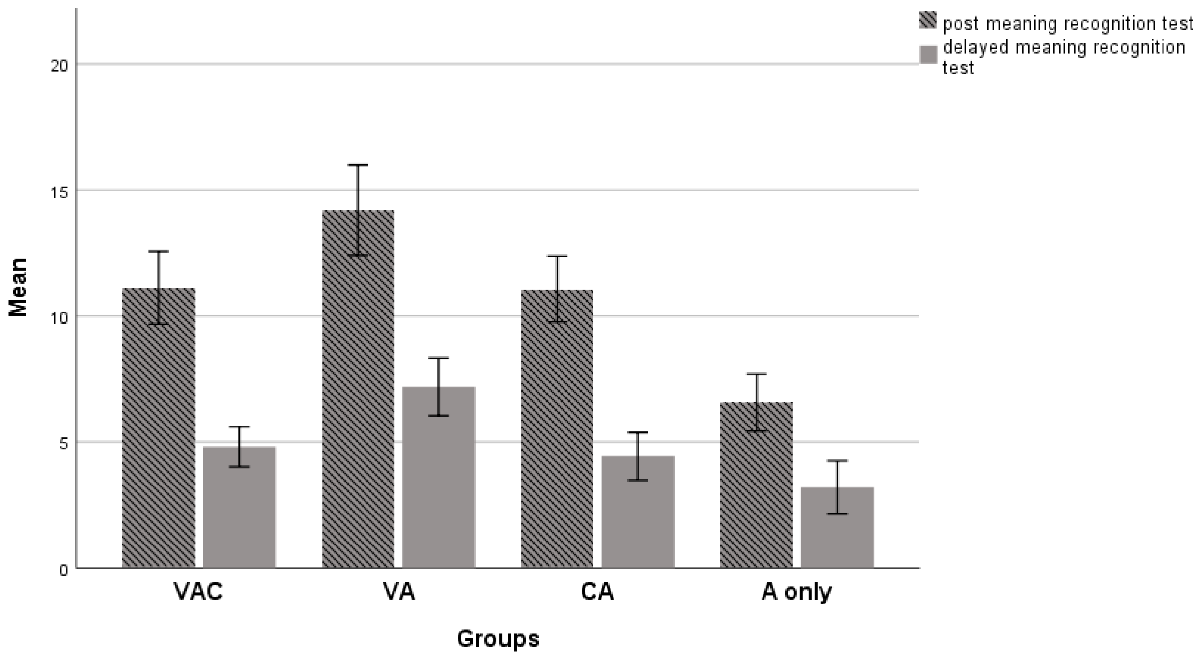


Figure 4.6. Mean scores for the spoken meaning recognition test of the four treatment groups over immediate-and-delayed-post-test administrations

A closer inspection of Figure 4.6 shows that the groups' scores on the immediate post-test were greater than their scores on the delayed post-test. In fact, almost more than half of the initial gains of the VAC, VA and CA groups (M = 11.12; M = 14.19; M = 11.07) have dropped to (M = 4.81; M= 7.19; M= 4.43, respectively), after a four-week period. These findings imply that meaning recognition knowledge is not also resistant to attrition. A Wilcoxon Signed Rank test was applied, to explore whether the differences found between the participants' scores in each treatment group on the immediate post-test and on the delayed post-test were significantly different.

Table 4.16. A Wilcoxon signed rank test for the spoken meaning recognition test

Groups		spoken meaning recognition delayed post-test – spoken meaning recognition immediate post-test
VAC	Z	-4.467[b]
	Asymp. Sig. (2-tailed)	.000
VA	Z	-4.556[b]
	Asymp. Sig. (2-tailed)	.000
CA	Z	-4.791[b]
	Asymp. Sig. (2-tailed)	.000
A only	Z	-4.171[b]
	Asymp. Sig. (2-tailed)	.000

According to the Wilcoxon Signed Rank test (found in Table 4.16) the scores on the immediate post-test were significantly higher than those on the delayed post-test of the VAC group, $z = -4.467$, $p <.000$, $r = -.61$, the VA group, $z = -4.556$, $p <.000$, $r = -.62$, the CA group, $z = -4.791$, $p < .000$, $r = -.61$, and the A only group, $z = -4.171$, $p <.000$, $r = -.58$. Across all cases, large effect sizes were found. The findings may suggest that the four groups were unable to retain the meanings of a large number of the learned words over a one-month time interval.

4.3.2 The differential effects of the different modalities of audio-visual input on long-term retention of the target words (between-groups comparison)

This section provides a comparison of the four different modalities of audio-visual input with regards to their differential effects on long-term retention of the initial gains of the 36 target words. The participants' scores on the delayed post-tests across the four groups were compared. As stated earlier, the aim of carrying out this analysis was to explore which input condition was the most or least effective for maintaining the acquired word knowledge aspect. Therefore, a Kruskal-Wallis test was employed, as the data was not normally distributed. Since the Kruskal-Wallis H test does not produce a z-score, used to estimate the effect size, it is difficult to convert the test statistic H or Chi-square statistic to a z-score (Mayers, 2013), as “it’s summarising a general effect” (Field, 2009, p. 570). It is instead advisable to compute the effect size for each pair of the Mann Whitney U tests that are run to follow up on the analysis of the significant Kruskal-Wallis H test (Field, 2009; Mayers, 2013). However, running a series of the Mann-Whitney U tests will cause an inflation in the Type I error rate, which was corrected for by using a Bonferroni adjustment (Field, 2009; Mayers, 2013; Pallant, 2016). The Mann-Whitney U test was carried out six times and the alpha value (.05) was divided by 6 (.05/6). Thus, the difference between the groups was considered significant only if the *p* was less than .008 (p <.008).

4.3.2.1 Findings of the spoken form recognition delayed post-test

A Kruskal-Wallis H test was run to detect group differences with regards to their scores on the delayed spoken form recognition post-test.

Table 4.17. A Kruskal-Wallis H test for the spoken form delayed post-test

Total	108
Test statistic H	8.114
Degree of freedom	3
Asymptotic Sig. (2-sided test	.044
The test statistic is adjusted for ties	

The results of the Kruskal-Wallis H test are illustrated in Table 4.17. The results of the Kruskal-Wallis H test showed that participants’ scores on the spoken form recognition delayed post-test were statistically significantly different, H (3) = 8.114, p < .044. These results suggest that the different modalities of audio-visual input had differential effects on the retention of spoken form recognition knowledge. Therefore, Mann-Whitney U tests were conducted to specify which of the groups were different from each other.

Table 4.18. Mann-Whitney U post-hoc test for the spoken form recognition delayed post-test

	A only - CA	A only - VA	A only - VAC	CA - VA	CA - VAC	VA - VAC
Mann-Whitney U	257	226.5	180.5	371.5	342	344
Z	-2.012	-2.047	-2.745	-0.539	-0.799	-0.126
Asymp. Sig. (2-tailed)	.044	.041	.006	.590	.424	.900

The findings of the Mann-Whitney U post-hoc test (reported in Table 4.18) showed that the retention scores of the A only group and the VAC group were statistically significantly different, $U = 180.5$, $z = -2.745$, $p < .006$, $r = .038$, with a medium effect. However, the retention scores of the other experimental groups were not significantly different (all $p > .008$).

4.3.2.2 Findings of the spoken meaning recall delayed post-test

The Kruskal-Wallis H test in Table 4.19 was applied to compare the effects of the four treatment conditions on the retention scores on the spoken meaning recall delayed post-test.

Table 4.19. A Kruskal-Wallis H test for the spoken meaning recall delayed post-test

Total	108
Test statistic H	39.705
Degree of freedom	3
Asymptotic Sig. (2-sided test	.000
The test statistic is adjusted for ties	

The outcomes of the Kruskal-Wallis H test demonstrated that the four experimental groups differed significantly in their retention scores on the spoken meaning recall delayed post-test, H (3), = 39.705, $p < .000$. This suggests that the four input conditions had differential effects on the retention of the meaning recall knowledge dimension. A series of Mann-Whitney U tests were utilised to examine where the difference between the groups existed.

Table 4.20. Mann-Whitney U post-hoc test for the spoken meaning recall delayed post-test

	A only - CA	A only - VA	A only - VAC	CA - VA	CA - VAC	VA - VAC
Mann-Whitney U	190.500	37.000	111.500	179.000	339.000	188.500
Z	-3.196	-5.582	-4.148	-3.658	-.858	-2.935
Asymp. Sig. (2-tailed)	.001	.000	.000	.000	.391	.003

A detailed look at the Mann-Whitney U tests (see Table 4.20) revealed that the students' score of the A only group was significantly lower than that of the CA group, $U = 190.500$, $z = -3.196$, $p < .001$, $r = -0.43$; the VA group, $U = 37.000$, $z = -5.582$, $p < .000$, $r = -0.77$; and the VAC group, $U = 111.500$, $z = -4.148$, $p < .000$, $r = -0.58$. In all of these cases, the effect sizes were found to be between moderate and large. These findings seem to suggest that multimodal input conditions (VA, VAC, and CA) were more effective for the retention of the meaning recall knowledge aspect than the single input condition (A only). The Mann-Whitney U tests also showed that the scores of the participants involved in the CA and the VAC groups were significantly lower than those of the VA groups, $U = 179.000$, $z = -3.658$, $p < .000$, $r = -0.48$ and $U = 188.500$, $z = -2.935$, $p < .003$, $r = -0.40$, respectively. The effect sizes in these cases were medium. The findings seem to indicate that the audio-visual modality (VA) was the most effective for the retention of word meanings compared to the other modalities (VAC, CA, and A only).

4.3.2.3 Findings of the spoken meaning recognition delayed post-test

In order to compare the differences between the four experimental groups' scores on the spoken form recognition delayed post-test, a Kruskal-Wallis H test was used.

Table 4.21. A Kruskal-Wallis H test for the spoken meaning recognition delayed post-test

Total	108
Test statistic H	26.488
Degree of freedom	3
Asymptotic Sig. (2-sided test	.000
The test statistic is adjusted for ties	

The results of the Kruskal-Wallis H test are shown in Table 4.21. The data demonstrated that the VAC, VA, CA, and A only treatment groups were significantly different in their retention scores on the spoken meaning recognition delayed post-test, H (3) = 26.488, $p < .000$. This may imply that the four modalities of audio-visual input affected the long-term retention of meaning recognition knowledge differently. A series of Mann-Whitney U tests were conducted to follow up on these significant findings.

Table 4.22. Mann-Whitney U post-hoc test for the spoken meaning recognition delayed post-test

	A only - CA	A only - VA	A only - VAC	CA - VA	CA - VAC	VA - VAC
Mann-Whitney U	279.500	89.000	204.000	176.000	350.000	174.000
Z	-1.631	-4.593	-2.301	-3.696	-.664	-3.186
Asymp. Sig. (2-tailed)	.103	.000	.021	.000	.506	.001

The outcomes of the Mann-Whitney U tests are reported in Table 4.22. From the table above, it can be found that the VA group was able to retain more meanings until the one-month delayed post-test than the A only group, U = 89.000, z = -4.593, $p < .000$, r = -0.63; the CA group, U = 176.000, z = -3.696, $p < .000$ r = 0.48; and the VAC group, U = 174.000, z = -3.186, $p < .001$, r = -0.43. The magnitude of the effect ranged between moderate and strong in all cases. This would seem to indicate that the audio-visual input condition (VA) is the most effective for the retention of meaning recognition knowledge compared to the other input conditions.

4.3.3 Summary of key findings

In this section, the third research question, which examined the differential effects of the four modalities of audio-visual input on long-term retention of spoken form recognition, meaning recall, and meaning recognition of the target 36 words, was analysed and addressed. The following are the key findings revealed:

- The data showed a significant decline in the participants' mean scores from the immediate post-tests to the delayed post-tests, irrespective of the different input conditions;
- The audio-visual input modality (VA) was the most effective condition for the retention of receptive and productive knowledge of word meanings.

The following section presents the findings of the fourth research question pertaining to the role played by the item-related variable (frequency of occurrence) in incidental vocabulary short-term learning.

4.4 Section four: research question 4:

R-Q 4: To what extent does the frequency of occurrence affect L2 incidental vocabulary learning from the four experimental conditions?

This section aims to provide an answer to the fourth research question, which looked at the repetition effect on incidental vocabulary short-term learning of the 36 target words. As discussed in the Methodology Chapter, the 36 target lexical items were arranged into three frequency groups: 2-4 repetitions; 5-7 repetitions; and 8+ repetitions based on their repetitions in the videos. Assigning the target words into different frequency groups was necessary as the target words occurred at different levels of frequency. Thus, this subgrouping is a practical strategy to study the effect of the repetitions of the target words. Each of the three sub-sections below opens with descriptive statistics for the gains of the 36 target words in each frequency group for the four experimental groups. Next, inferential tests were conducted to find out whether the three frequency groups (2-4; 5-7; 8+) differed significantly from each other with regards to their effects on the learning of the three vocabulary knowledge dimensions for each experimental group. A non-parametric method was followed because the data did not meet the assumption of normality of distribution. As had been explained earlier, a Kruskal-Wallis H test was the most suitable non-parametric test to run to compare the scores of three and more groups (Pallant, 2016). To follow up on the outcomes of a significant Kruskal-Wallis H test, a post-hoc test, Mann-Whitney U test, was employed. The significance level was adjusted to account for the multiple comparisons. The adjusted significance level was .016 (.05/3). As explicitly explained in the preceding section, the effect size was calculated for each Mann-Whitney test conducted.

4.4.1 Findings of the spoken form recognition test

The descriptive data represented in Figure 4.7 and Table 4.23 displays the mean numbers of vocabulary items in each of the three frequency groups (2-4, 5-7, 8+) answered correctly on the spoken form recognition immediate post-test by the four experimental groups.

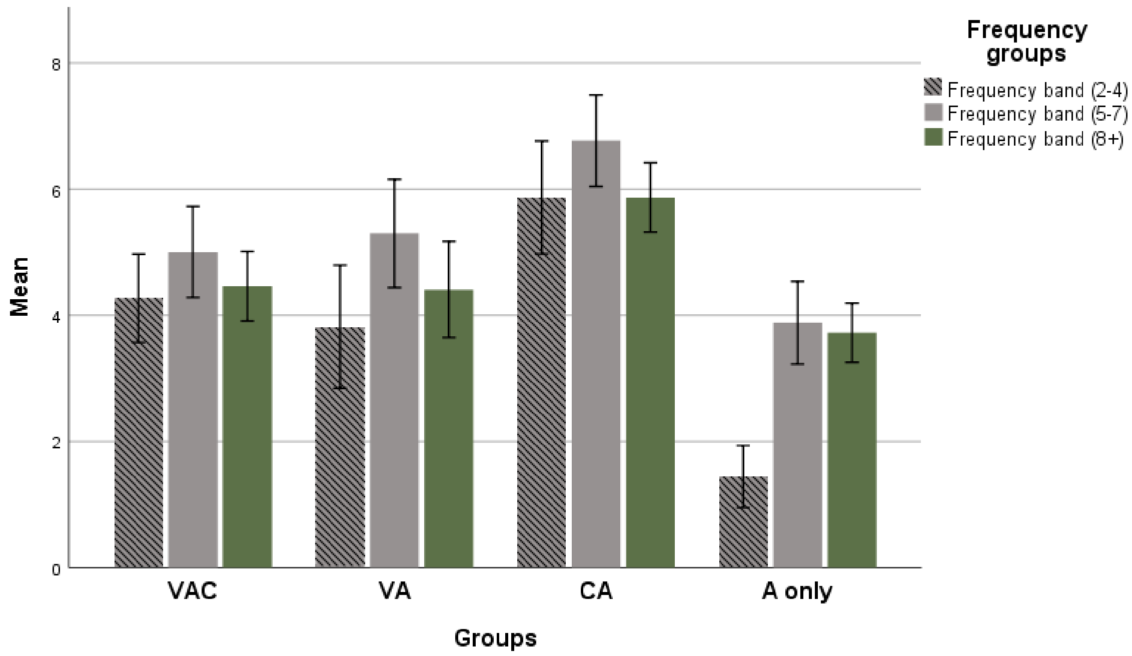


Figure 4.7 The mean scores of words of the three frequency groups answered correctly on the spoken form recognition immediate post-test by the four experimental groups

Table 4.23. The mean scores of words of the three frequency groups answered correctly on the spoken form recognition immediate post-test by the four experimental groups

Groups	Frequency bands	(2-4)	(5-7)	(8+)	H	Sig. (2-sided test)
VAC	Mean	4.27	5.0	4.46	2.392	.302
	Std. Deviation	1.733	1.789	1.363		
VA	Mean	3.81	5.3	4.41	8.095	.017
	Std. Deviation	2.466	2.163	1.927		
CA	Mean	5.87	6.77	5.87	4.211	.122
	Std. Deviation	2.389	1.942	1.479		
A only	Mean	1.44	3.88	3.72	31.228	.000
	Std. Deviation	1.193	1.59	1.137		

As it can be seen from the table and figure above, the effect of repetition was generally found on the learning of the spoken form knowledge type across the four treatment groups. That is, across the four experimental groups, there was a noticeable rise in the mean scores of the first frequency group (2-4) to the second group (5-7). Surprisingly, as Table 4.23 predicts, the participants' scores on items of the third frequency group (8+) were lower than their scores of the second group (5-7). However, there were no significant differences between the scores of three frequency groups for the VAC group, $H(2) = 2.392$, $p > .302$ and for the CA group, $H(2) = 4.211$, $p > .122$. However, statistically significant differences were detected between the scores of the three frequency groups for the VA group, $H(2) = 8.095$, $p < .017$ and for the A only group, $H(2) = 31.228$, $p < .000$. Three Mann-Whitney U tests were run on the three frequency groups to find out where the differences existed.

Table 4.24. Mann-Whitney U tests for frequency of occurrence in the spoken form recognition immediate post-test

Frequency bands	Mann-Whitney U	Z	Sig. (2-tailed)	r
VAC				
(2-4) - (5-7)	262	-1.411	.158	-0.1957
(2-4) - (8+)	314	-0.453	.651	-0.0628
(5-7) - (8+)	275.5	-1.171	.242	-0.1624
VA				
(2-4) - (5-7)	207.5	-2.75	.006	-0.3742
(2-4) - (8+)	283.5	-1.422	.155	-0.1935
(5-7) - (8+)	274	-1.585	.113	-0.2157
CA				
(2-4) - (5-7)	349.5	-1.505	.132	-0.1943
(2-4) - (8+)	434	-0.24	.810	-0.031
(5-7) - (8+)	317.5	-2.001	.045	-0.2583
A only				
(2-4) - (5-7)	76	-4.652	.000	-0.6579
(2-4) - (8+)	61.5	-4.959	.000	-0.7013
(5-7) - (8+)	275	-0.752	.452	-0.1063

The outcomes of the Mann-Whitney U post-hoc tests (reproduced in Table 4.24) showed that the scores of the first frequency group (2-4) were statistically significantly lower than those of the second frequency group (5-7) for participants under the VA condition, $U = 207.5$, $z = -2.75$, $p < .006$, $r = -0.37$ and for students in the A only group, $U = 76$, $z = -4.652$, $p < .000$, $r = - 0.65$. Significant differences were also found between the scores of the first group (2-4) and of the third group (8+) for the A only group, U = 61.5, z = -4.959, $p < .000$, r = -0.70. The effect sizes in all cases were found to be between medium to large. These findings suggest that the effect of frequency of occurrence on the learning of the spoken form recognition knowledge aspect varies based on the modalities of audio-visual input.

4.4.2 Findings of the spoken meaning recall test

The descriptive data showed in Figure 4.8 and Table 4.25 displays the mean numbers of vocabulary items in each of the three frequency groups (2-4, 5-7, 8+) answered correctly on the spoken meaning recall immediate post-test by the four experimental groups.

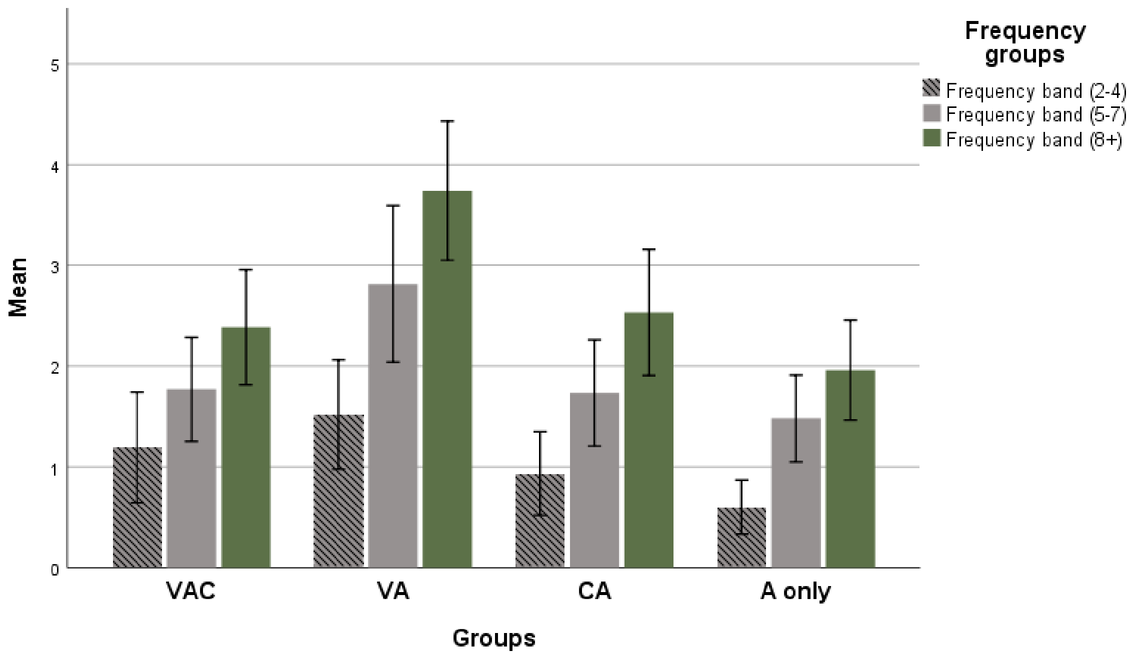


Figure 4.8 The mean scores of words of the three frequency groups answered correctly on the spoken meaning recall immediate post-test by the four experimental groups

Table 4.25. The mean scores of words of the three frequency groups answered correctly on the spoken meaning recall immediate post-test by the four experimental groups

Groups	Frequency bands	(2-4)	(5-7)	(8+)	H	Sig. (2-sided test)
VAC	Mean	1.19	1.77	2.38	10.558	.005
	Std. Deviation	1.357	1.275	1.416		
VA	Mean	1.52	2.81	3.74	20.209	.000
	Std. Deviation	1.369	1.962	1.745		
CA	Mean	0.93	1.73	2.53	15.422	.000
	Std. Deviation	1.112	1.413	1.676		
A only	Mean	0.6	1.48	1.96	19.138	.000
	Std. Deviation	0.645	1.046	1.207		

Data in the figure and table above indicated that meaning recall knowledge of the target words was positively affected by repetition. Noticeably, there was a gradual increase in knowledge of meaning from the first frequency group (2-4) to the higher two frequency groups (5-7 and 8+), regardless of the treatment conditions. As demonstrably shown in Figure 4.8 and Table 4.25, the gain rates of the first frequency group were the lowest in comparison to those of the two higher frequency groups. While the mean numbers on items in the third frequency group (8+) were the highest across the four experimental groups. The findings of the Kruskal-Wallis H test showed that there were significant differences between the scores of the three frequency groups for the VAC group, $H(2) = 10.558$, $p < .005$; for the VA group, $H(2) = 20.209$, $p < .000$; for the CA group, $H(2) = 15.422$, $p < .000$; and the A only group, $H(2) = 19.138$, $p < .000$. A series of Mann-Whitney U post-hoc tests were employed to determine where the differences between the frequency groups laid specifically.

Table 4.26. Mann-Whitney U tests for frequency of occurrence in the spoken meaning recall immediate post-test

Frequency bands	Mann-Whitney U	Z	Sig. (2-tailed)	r
VAC				
(2-4) - (5-7)	235.5	-1.937	.053	-0.2686
(2-4) - (8+)	174.5	-3.059	.002	-0.4242
(5-7) - (8+)	250	-1.661	.097	-0.2303
VA				
(2-4) - (5-7)	213.5	-2.669	.008	-0.3632
(2-4) - (8+)	117	-4.338	.000	-0.5903
(5-7) - (8+)	246.5	-2.069	.039	-0.2816
CA				
(2-4) - (5-7)	296.5	-2.355	.019	-0.304
(2-4) - (8+)	201	-3.775	.000	-0.4874
(5-7) - (8+)	327	-1.854	.064	-0.2394
A only				
(2-4) - (5-7)	162	-3.079	.002	-0.4354
(2-4) - (8+)	104.5	-4.194	.000	-0.5931
(5-7) - (8+)	245	-1.357	.175	-0.1919

The analysis in Table 4.26 showed that across the four experimental groups scores of the first frequency group (2-4) were significantly lower than those of the third frequency group (3+), for the VAC group, $U = 174.5$, $z = -3.059$, $p <. 002$, $r = -0.42$; for the VA group, $U = 117$, $z = -4.338$, $p <.000$, $r = -0.59$; for the CA group, $U = 201$, $z = -3.775$, $p <.000$, $r = -0.48$; and for the A only group, $U = 104$, $z = -4.194$, $p <.000$, $r = -0.59$. The effect sizes in these cases ranged between moderate and strong. In addition, the data demonstrated that scores on words in the second frequency group were significantly greater than those in the first frequency group for the VA group, $U = 213.5$, $z = -2.669$, $p < .008$, $r = -0.36$ and for the A only group , $U = 162$, $z = - 3.079$, $p < .002$, $r = -0.43$. The effect sizes in both cases were medium. These suggest that knowledge of meaning was affected by repetitions. It requires between 2 and 8 repetitions to develop regardless of the modality of input.

4.4.3 Findings of the spoken meaning recognition tests

The descriptive data displayed in Figure 4.9 and Table 4.27 shows the mean numbers of vocabulary items in each of the three frequency groups (2-4, 5-7, 8+) answered correctly on the spoken meaning recognition immediate post-test by the four experimental groups.

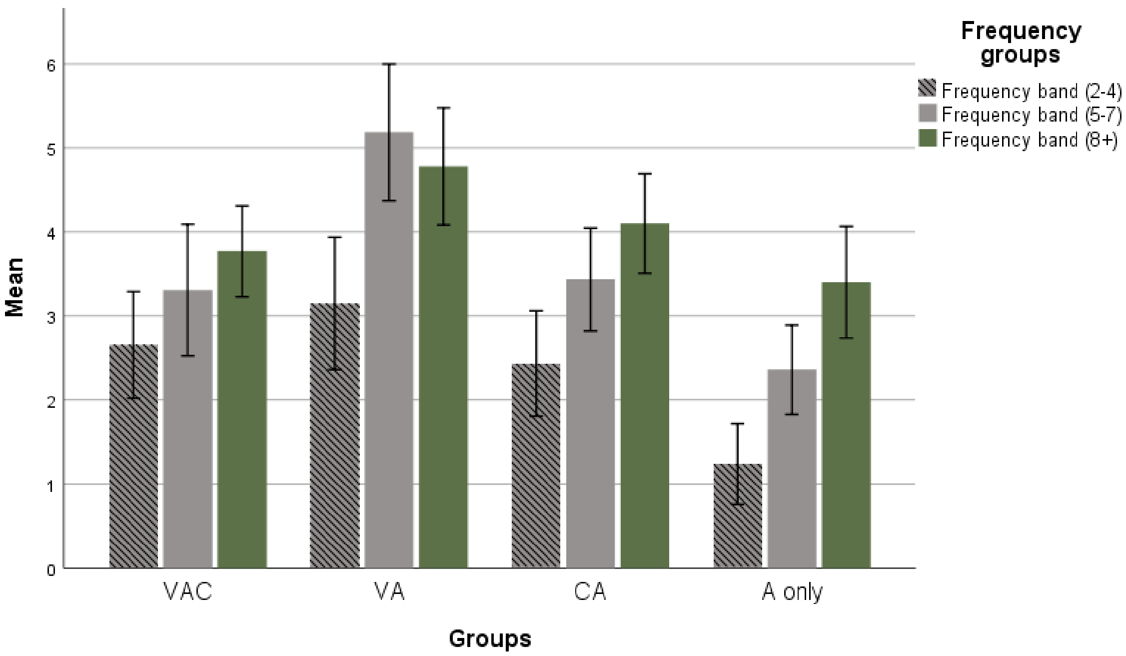


Figure 4.9 The mean scores of words of the three frequency groups answered correctly on the spoken meaning recognition immediate post-test by the four experimental groups

Table 4.27. The mean scores of words of the three frequency groups answered correctly on the spoken meaning recognition immediate post-test by the four experimental groups

Groups	Frequency bands	(2-4)	(5-7)	(8+)	H	Sig. (2-sided test)
VAC	Mean	2.65	3.31	3.77	5.52	.063
	Std. Deviation	1.573	1.934	1.336		
VA	Mean	3.15	5.19	4.78	14.529	.001
	Std. Deviation	1.994	2.058	1.761		
CA	Mean	2.43	3.43	4.1	13.295	.001
	Std. Deviation	1.675	1.633	1.583		
A only	Mean	1.24	2.36	3.4	22.535	.000
	Std. Deviation	1.165	1.287	1.607		

As found in the meaning recall test, there was a considerable increase in the mean scores from the first frequency group (2-4) to the higher two frequency groups (5-4 and 8+). The mean numbers of the third frequency group (8+) were the highest in comparison to the mean numbers of the first and second frequency groups for the VAC, CA, and A only groups. The Kruskal-Wallis H test indicated that the scores of the three frequency groups were significantly different for the VA group, $H(2) = 14.529$, $p < .001$; for the CA group, $H(2) = 13.295$, $p < .001$; and for the A only group, $H(2) = 22.535$, $p < .000$, but not for the VAC group, $H(2) = 5.52$, $p < .063$. In order to see where the differences existed, a number of Mann-Whitney post-hoc tests were conducted.

Table 4.28. Mann-Whitney U tests for frequency of occurrence in the spoken meaning recognition immediate post-test

Frequency bands	Mann-Whitney U	Z	Sig. (2-tailed)	r
VAC				
(2-4) - (5-7)	269.5	-1.278	.201	-0.1772
(2-4) - (8+)	201.5	-2.562	.010	-0.3553
(5-7) - (8+)	309.5	-0.536	.592	-0.0743
VA				
(2-4) - (5-7)	169	-3.431	.001	-0.4669
(2-4) - (8+)	187.5	-3.102	.002	-0.4221
(5-7) - (8+)	331.5	-0.58	.562	-0.0789
CA				
(2-4) - (5-7)	303	-2.215	.027	-0.286
(2-4) - (8+)	216	-3.509	.000	-0.453
(5-7) - (8+)	341.5	-1.633	.102	-0.2108
A only				
(2-4) - (5-7)	170.5	-2.829	.005	-0.4001
(2-4) - (8+)	89	-4.445	.000	-0.6286
(5-7) - (8+)	185	-2.533	.011	-0.3582

The results obtained from the Mann-Whitney U tests for the three frequency groups are reported in Table 4.28. Similar to the findings of the meaning recall test, the scores of the first frequency group (2 – 4) were significantly lower than those of the third frequency group (8 +) for the VAC group, U = 201.5, z = -2.562, p <.010, r = -0.35; for the VA group, U = 187.5, z = -3.102, p <.002, r = -0.42; for the CA group, U = 216, z = -3.509, p <.000, r = -0.45; and for the A only group, U = 89, z =

-4.445, $p < .000$, $r = -0.62$. The effect sizes for these comparisons ranged between moderate and large. Furthermore, it was found that differences between the scores of the first frequency group and the second frequency group were statistically significantly different for the VA group, $U = 169$, $z = -3.431$, $p < .001$, $r = -0.46$ and for the A only group, $U = 170.5$, $z = -2.829$, $p < .005$, $r = -0.40$. The strength of the effect in the two cases was moderate. Considering the differences between the second (5-7) and third frequency (8+) groups, the data revealed that differences were only significant for the participants under the A only condition, $U = 185$, $z = -2.533$, $p < .011$, $r = -0.35$, with a medium effect. These results indicate that frequency of occurrence influences the learning of meaning recognition, however its effect varies across the four input conditions.

4.4.4 Summary of key findings:

This last section of this chapter presented the findings of the fourth research question, which was concerned with the effect of frequency of occurrence on incidental vocabulary learning of spoken form recognition, meaning recall and meaning recognition across the four of audio-visual input conditions. The most interesting results found were:

- There was a general effect of frequency on the learning of the three knowledge types, however the effect was modulated by the modality of input;
- The effect of repetition was more evident on meaning than form.

4.5 Chapter summary

This chapter presented the findings of the first four research questions pertaining to vocabulary learning and retention. Different statistical analyses were carried out to examine those four research questions. The following are the most interesting findings found:

- The four experimental groups had similar level of pre-existing knowledge of the target words;
- Incidental vocabulary learning of the three vocabulary knowledge types has occurred in the four different modalities of audio-visual input;
- The multimodal input conditions (VAC, VA, and CA) were more effective for promoting incidental vocabulary learning of the three knowledge aspects than the single input condition, A only;
- The presence of captions was found useful for enhancing the spoken form recognition knowledge facet;

- The visual images of the videos proved effective for fostering meaning recall and meaning recognition knowledge;
- Large attrition rates were found across the four treatment conditions of the three vocabulary knowledge aspects;
- The effect of repetition on incidental vocabulary learning varied based on the modalities of audio-visual input and the target word knowledge aspects.

The next chapter presents the findings of the WM data.

Chapter 5 Working memory findings

The principle aim of this chapter is to explore the relationship between working memory (WM) and L2 incidental vocabulary learning and retention. In particular, it aims to examine whether the WM systems, verbal and visuospatial, contributed differently to the relationship. The chapter also aims to examine whether the involvement of the verbal and visuospatial subsystems of WM in incidental vocabulary learning and retention can be better expressed through simple tasks (forward digit recall, dot matrix) or complex tasks (backward digit recall, the odd one out). Furthermore, it intends to explore the predictive role of WM in moderating the effects of the different modalities of audio-visual input on incidental vocabulary learning and retention.

As described in the Methodology Chapter, the study utilized four WM tests: two verbal (forward digit recall and backward digit recall) and two visual (dot matrix and the odd one out). These meant to measure both the verbal system and the nonverbal system of WM. As a reminder, 12 randomly selected participants from each experimental group, who had completed all the vocabulary-learning phases, were invited to complete the four WM tests. Scores on the four WM tests were considered as independent, predictor variables that may potentially overshadow incidental vocabulary learning and retention. The vocabulary score was used as a dependent variable. The study made use of a variety of statistical techniques for the analysis of the WM data. To simplify the presentation of the findings, the chapter is divided into three main sections that cover the research questions pertaining to WM in this study. In each section the analysis of the WM data with incidental vocabulary short-term learning is reported first; then the analysis of WM and incidental vocabulary long-term retention.

5.1 Section one: research question 5

R-Q 5: Is there a relationship between WM and incidental vocabulary short-term learning and long-term retention? If so, which subset(s) of WM, phonological loop or visuospatial sketchpad contribute(s) to this relationship?

The aim of this section is to explore the relationship between WM and both incidental vocabulary short-term learning and incidental vocabulary long-term retention. The relationship between WM and both incidental vocabulary short-term learning and long-term retention was examined using correlational analysis tests. That is, the extent to which the four independent variables, forward digit recall, backward digit recall, dot matrix, and the odd-one-out tests, correlated with the scores on the three vocabulary immediate and delayed post-tests was examined. The aim was to examine the contribution of the two subsystems of WM - verbal and visual - to determine whether the two subsystems contributed differently to this relationship. The preliminary testing of the important assumptions of parametric correlation procedures (i.e., linearity, normality, and outliers) was carried out utilising scatterplots and normality tests. The findings of this preliminary testing showed that the data violated these important assumptions of the parametric correlation (Pearson). The non-parametric Spearman's correlation analysis was thus run.

Prior to exploring the relationship between WM and incidental vocabulary learning and retention, a one way ANOVA test (normally distributed data) was carried out to explore whether the four groups of participants differed significantly in their scores on the four WM tests. This is to get a general impression of the differences of the participants' scores on the four WM tests. The results of the one way ANOVA test (presented in Table 5.1) indicated that the four treatment groups were not significantly different from each other with regards to their scores on the forward digit recall test, $F (3, 44) = .228$, $P >. 876$; on the backward digit recall test, $F (3, 44) = 1.408$, $p > .253$; on the dot matrix test, $F (3, 44) = 2.372$, $p > .083$; and on the odd one out test, $F (3, 44) = 1.604$, $p >. 202$. These results could be taken to argue that the four groups had similar WM spans.

Table 5.1. A one way ANOVA test for the WM scores across the four experimental groups

WM measures		Sum of Squares	df	Mean Square	F	Sig.
Forward digit recall	Between Groups	111.229	3	37.076	.228	.876
	Within Groups	7140.25	44	162.278		
	Total	7251.48	47			
Backward digit recall	Between Groups	795.417	3	265.139	1.408	.253
	Within Groups	8285.83	44	188.314		
	Total	9081.25	47			
Dot matrix	Between Groups	909.729	3	303.243	2.372	.083

	Within Groups	5625.08	44	127.843		
	Total	6534.81	47			
The odd-one-out	Between Groups	766.417	3	255.472	1.604	.202
	Within Groups	7007.5	44	159.261		
	Total	7773.92	47			

5.1.1 WM and incidental vocabulary short-term learning

In order to explore the relationship between WM and incidental vocabulary learning, a Spearman's correlation test was employed. The results of the Spearman's test concerning the relationship between WM and incidental vocabulary learning are shown in Table 5.2.

Table 5.2. Correlation analysis of the four WM measures and the three vocabulary immediate post-tests

Measure	1	2	3	4	5	6	7
1. Forward digit recall	-						
2. Backward digit recall	.653**	-					
3. Dot matrix	.358*	.174	-				
4. The odd-one-out	.616**	.584**	.306*	-			
5. Spoken form recognition immediate post-test	-.026	-.083	-.174	.189	-		
6. Spoken meaning recall immediate post-test	.593**	.390*	-.094	-.094	-.298*	-	
7. Spoken meaning recognition immediate post-test	.433**	-.034	-.132	-.132	.344*	.697**	-

Note: ** Correlation is significant at the .01 level (2-tailed). Note: * Correlation is significant at the .05 level (2-tailed).

As expected, the results displayed Table 5.2 showed that the participants' scores on the spoken form recognition immediate post-test correlated with their scores on both the spoken meaning recall immediate post-test, $r_s = -.298$, $p <.040$ and with scores on the spoken meaning recognition immediate post-test, $r_s = .344$, $p < .017$. The two meaning tests, spoken meaning recall and spoken meaning recognition, correlated highly with one another, $r_s = .697$, $p <.000$. These results seem to suggest that the three vocabulary tests are assessing multiple aspects of the same construct (vocabulary knowledge). Regarding the main effects of WM, it was found that there was a positive relationship between the verbal forward digit recall WM measure and the spoken meaning recall immediate post-test, $r_s = .593$, $p <.000$ and the spoken meaning recognition immediate post-test, $r_s = .433$, $p <.001$. The size of the effect was found to be between moderate and large, as per Cohen's guidelines (1992). A positive correlation was also observed between the other verbal WM measure, backward digit recall and the scores on the spoken meaning recall immediate post-test, $r_s = .390$, $p < .021$, yet with somewhat smaller strength of correlation (moderate effect size). The significant findings seem to suggest that the verbal/phonological WM system, the phonological loop, is closely related to L2 incidental vocabulary short-term learning. Nonetheless, the Spearman's correlation test revealed no significant correlations between either the dot matrix visual WM measure and the three vocabulary immediate post-tests or the odd one out visual WM test and three vocabulary immediate post-tests.

Up to this point, the data pertaining to the relationship between WM and incidental vocabulary short-term learning was presented. In the following section, the relationship between WM and vocabulary long-term retention is examined.

5.1.2 WM and incidental vocabulary long-term retention

The correlations between the four WM measures and the three vocabulary delayed post-tests are shown in Table 5.3.

Table 5.3. Correlation analysis of the four WM measures and the three vocabulary delayed post-tests

measure	1	2	3	4	5	6	7
1. Forward digit recall	-						
2. Backward digit recall	.653**	-					
3. Dot matrix	.358*	.174	-				
4. The odd-one-out	.616**	.584**	.306*	-			
5. Spoken form recognition delayed post-test	-.082	-.113	.017	.173	-		
6. Spoken meaning recall delayed post-test	-.279	-.137	.258	.154	.484**	-	
7. Spoken meaning recognition delayed post-test	-.074	.044	.131	.076	.447**	.681**	-

Note: ** Correlation is significant at the .01 level (2-tailed). Note: * Correlation is significant at the .05 level (2-tailed).

Similar to the findings of the Spearman's correlation test shown in Table 5.2, the Spearman's correlation test (shown in Table 5.3) revealed that the participants' scores on the spoken form recognition delayed post-test moderately correlated with their scores on both the spoken meaning recall delayed post-test, r_s =, .484, p <.000, as well as with their scores on the spoken meaning recognition delayed post-test, r_s = .447, p <.001. A strong correlation was also found between the spoken meaning recall delayed post-test and the spoken meaning recognition delayed post-test, $r_s = .681$, $p <.000$. These significant may imply that the three vocabulary tests are measuring different dimensions of the same construct (vocabulary knowledge). The Spearman's test showed a moderate correlation nonsignificant correlations between the four WM measures, forward digit recall test, backward digit recall test, dot matrix test, the odd one out test and the three vocabulary delayed post-tests. These results may suggest non-significant links between WM and long-term retention of vocabulary items.

5.1.3 Summary of key findings:

These above results provide evidence that WM is related to incidental vocabulary learning short-term learning. In particular, the correlational analysis showed that the verbal/phonological WM system, contributed significantly to this relationship. The key findings revealed in this section were:

- The four experimental groups were not statistically significantly different in their scores on the four WM tests;
- WM had a significant relationship with incidental vocabulary short-term learning; the verbal/phonological WM system significantly contributed to this relationship;
- The visual WM system, the visuospatial sketchpad, did not contribute to incidental vocabulary short-term learning;
- WM was not related to long-term retention of vocabulary items.

Since, the scores of the three vocabulary tests, spoken form recognition test, spoken meaning recall test, and spoken meaning recognition test are correlated, it can be argued that they should be combined to form a single composite score. This composite score should offer an accurate measure of vocabulary construct. Therefore, in the subsequent analyses composite vocabulary scores will be used to represent the three vocabulary tests.

5.2 Section two: research question 6

R-Q 6: Which of the WM task types, simple or complex, best predicts incidental vocabulary short-term learning and long-term retention?

This part addresses the sixth research question which concerns the type of WM tasks, simple or complex that best predicted incidental vocabulary short-term learning and long-term retention. The purpose was to find out which WM measures independently predicted the composite vocabulary scores for incidental vocabulary learning and retention. For this purpose, a standard multiple regression analysis was conducted. It is most appropriate when there is no theoretical basis for entry of the predictor/independent variables in the model (Field, 2009). It examines "how much unique variance in the dependent variable each of the independent variables explained" (Pallant, 2016, p. 150).

Pre-analysis screening procedures were run to ensure that the data met the key assumptions of normality, linearity, homoscedasticity, and multicollinearity (Field, 2009; Pallant, 2016). Linearity and homoscedasticity were checked by plotting scatter-plots. Multicollinearity was inspected through correlation coefficients and Tolerance/VIF values. Normality was checked through histogram with normal curve and p-p plots. The four-predictor variables (forward digit recall test, backward digit recall test, dot matrix test, and the odd-one-out test) entered into the predication model simultaneously, as no variable was considered to be logically prior to any other.

5.2.1 WM and incidental vocabulary short-term learning

Table 5.4 presents the results of the standard multiple regression analysis. The model, which accounted for 41.9 % of the unique variance ($R = .647$, $R^2 = .419$) in the composite vocabulary scores, significantly predicted incidental vocabulary learning, $F(4, 43) = 7.752$, $p < .000$. As indicated in Table 5.4, incidental vocabulary learning was only predicted by the simple verbal forward digit recall test with a high beta value ($\beta = .690$, $p < .05$). On the other hand, the other three-predictor variables did not significantly contribute to incidental vocabulary learning (p values >.05). These findings demonstrate that the simple verbal forward digit recall WM test made a significant influence on incidental vocabulary learning.

Table 5.4. Standard multiple regression analysis predicting the composite vocabulary score for the three immediate post-tests

	B	Std. Error	Beta
(Constant)	7.913	8.148	
Forward digit recall test	.349	.091	.690*
Backward digit recall test	.05	.082	.11
Dot matrix	.016	.067	.03
The odd-one-out	-.141	.075	-.289

Note. * p < .05

So far, the data related to the relationship between WM and incidental vocabulary learning, represented by the composite scores for the three vocabulary immediate post-test has been presented. The relationship between WM and incidental long-term vocabulary retention is examined below.

5.2.2 WM and incidental vocabulary long-term retention

The findings of the standard multiple regression analysis are shown in Table 5.5. The prediction model contained four predictor variables (forward digit recall test, backward digit recall test, dot matrix test, and the odd-one-out test) that were simultaneously entered into the model. In contrast to the findings reported in Table 5.4, the prediction model was not statistically significant, $F (4, 43) = .129$, $p > .971$ and did not account for any significant variability in the retention scores ($R = .109$, $R^2 = .012$). The four WM predictor variables did not significantly add to the prediction model (p values >.05). Overall, the results may suggest that the simple and complex WM measures have no significantly independent influences on vocabulary long-term retention.

Table 5.5. Standard multiple regression analysis predicting the composite vocabulary score for the three delayed post-tests

	B	Std. Error	Beta
(Constant)	20.535	12.630	
Forward digit recall test	.028	.142	.046
Backward digit recall test	-.070	.127	-.130
Dot matrix	-.036	.104	-.057
The odd-one-out	.007	.117	.013

5.2.3 Summary of key findings:

This section presented the analysis of the sixth research question, which was concerned with the independent influences of the simple and complex WM visual and verbal tasks on both incidental vocabulary short-term learning and long-term retention. The key findings revealed in this section were:

- The simple verbal forward digit recall test significantly influenced incidental vocabulary short-term learning;

- The other three WM measures did not contribute significantly to the explained variance in the composite vocabulary scores for incidental vocabulary learning;
- The four WM measures did not have significant influences on incidental vocabulary long-term retention.

5.3 Section three: research question 7

R-Q 7: Does WM mediate the effect of the four modalities of audio-visual input on incidental vocabulary short-term learning and long-term retention?

This section serves to examine the effect of individual differences in WM capacity on the effects of the different modalities of audio-visual input on incidental vocabulary short-term learning and long-term retention. This is to explore whether WM mediated the effects of the multiple modalities of audio-visual input. Following the procedure adopted by Lusk et al. (2009), Varol and Erçetin (2016), and Kozan et al. (2015), the subjects' scores on each of the WM measures were divided into two subgroups, resulting in two levels of WM capacity: low and high for each treatment group. This sub-grouping technique was important to clearly define the role of WM capacity in vocabulary learning through different modalities of input. The participants' raw scores on the four WM tests were divided into high and low levels on the basis of a median split. The composite vocabulary scores of these subgroups were compared. Then, a mixed design 2 (high WM capacity, low WM capacity) x 4 (experimental conditions: VAC, VA, CA, A only) ANOVA test was conducted for each of the four WM tests. Partial eta-squared (ηp^2) was used as an effect size measure, which was interpreted in light of Cohen's (1988) guidelines for effect size interpretations (.010 - .039 small effect, .050 - .110, medium effect, and .140 – 200 a large effect size). As highlighted earlier in this chapter, the three vocabulary immediate post-tests (spoken form recognition, spoken meaning recall and spoken meaning recognition) were combined to form a single composite score to represent incidental vocabulary short-term. Composite scores were also computed for the three vocabulary delayed post-tests.

In the following sub-sections, the participants' individual scores on the four WM tests are firstly examined. Following that, the results of WM and incidental vocabulary short-term learning are reported. Next, the outcomes of WM and incidental vocabulary long-term retention are outlined. Both descriptive statistics and inferential statistics were conducted for each of the four WM tests. Before carrying out the inferential analysis, ANOVA's assumption of normality was checked through a Shapiro-Wilk normality test (the sample size < 50 participants). This normality test yielded non-significant results, ($p > .05$), indicating that this assumption was met. The other key assumption of ANOVA, homogeneity of variance, was tested by a Leven's test. The results of the Leven's test suggested that this assumption was satisfied.

5.3.1 Individual performance data

The data presented in the first section of this chapter demonstrated that the differences in the four experimental groups' WM scores were found to be nonsignificant ($p = <0.05$). In this sub-section, the participants' raw scores on the four WM measures, are examined at individual level to give a general impression of the differences of the participants' scores on the four WM tests. This seeks to justify their appointment into different WM groups (high level and low level). Descriptive statistics of the four WM measures for the whole sample is presented in Table 5.6. Participants' scores on each of the four WM tests are also illustrated in Figures 5.1, 5.2, 5.3, and 5.4.

Table 5.6 Descriptive statistics of the WM measures

Experimental group	Student No	Forward digit recall	Backward digit recall	Dot matrix	The odd-one-out
VAC	1	103	109	115	108
	2	92	88	110	104
	3	84	96	96	109
	4	92	85	107	109
	5	82	88	102	89
	6	90	85	102	89
	7	92	106	116	106
	8	92	101	113	109
	9	88	93	110	106
	10	80	75	88	79
	11	99	94	80	104
	12	100	124	118	131

VA	1	105	91	110	97
	2	77	78	107	87
	3	84	83	99	92
	4	97	98	110	119
	5	94	88	105	104
	6	94	96	113	104
	7	77	85	110	104
	8	90	101	110	124
	9	90	85	110	104
	10	107	85	113	114
	11	92	88	113	104
	12	69	78	99	99
CA	1	80	78	82	89
	2	118	132	99	131
	3	82	80	102	92
	4	82	78	93	72
	5	92	101	99	104
	6	100	124	116	104
	7	80	88	96	94
	8	92	85	116	104
	9	94	85	82	114

	10	90	101	91	104
	11	88	91	93	92
	12	94	114	93	106
A only	1	92	93	113	124
	2	75	75	118	101
	3	84	88	107	111
	4	109	88	113	126
	5	86	85	80	104
	6	94	101	130	111
	7	86	96	107	106
	8	86	93	85	94
	9	92	93	121	99
	10	118	109	113	111
	11	86	91	102	116
	12	105	117	88	134
All four experimental groups	Mean	92.27	94.38	104.06	104.96
	Minimum and maximum	69-128	75-132	80-130	72-134
	Skewness	1.102	0.924	-0.43	0.083

Note: maximum score: 150; minimum score: 65

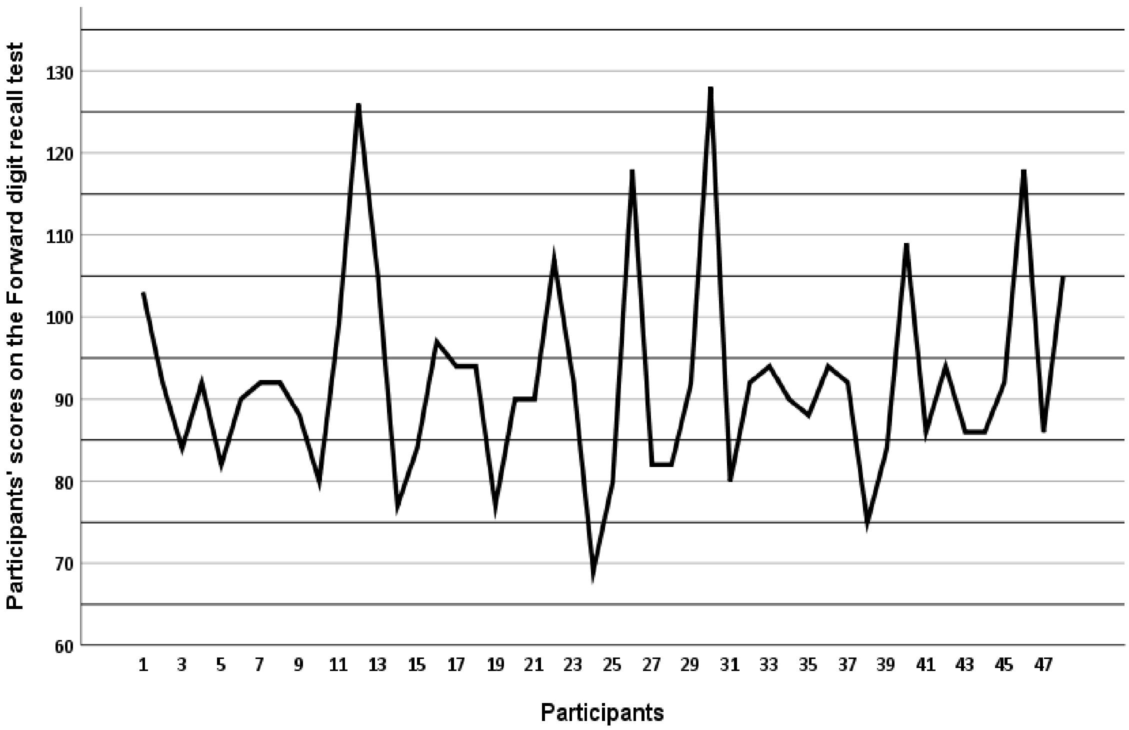


Figure 5.1 Participants' scores on the forward digit recall test

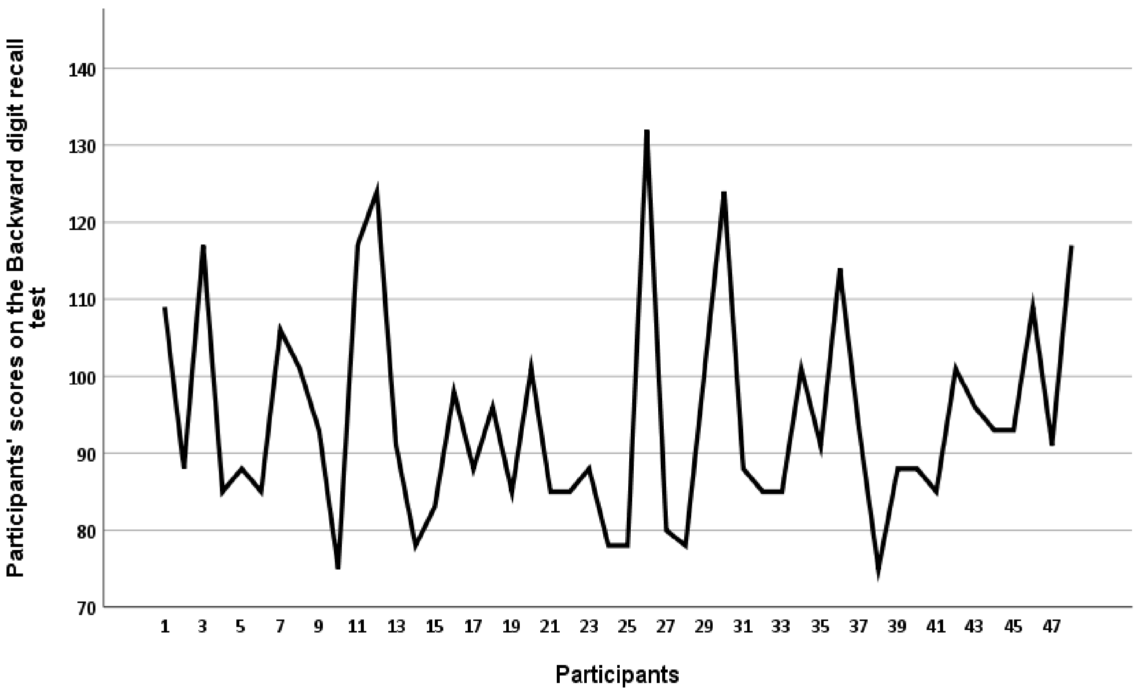


Figure 5.2 Participants' scores on the backward digit recall test

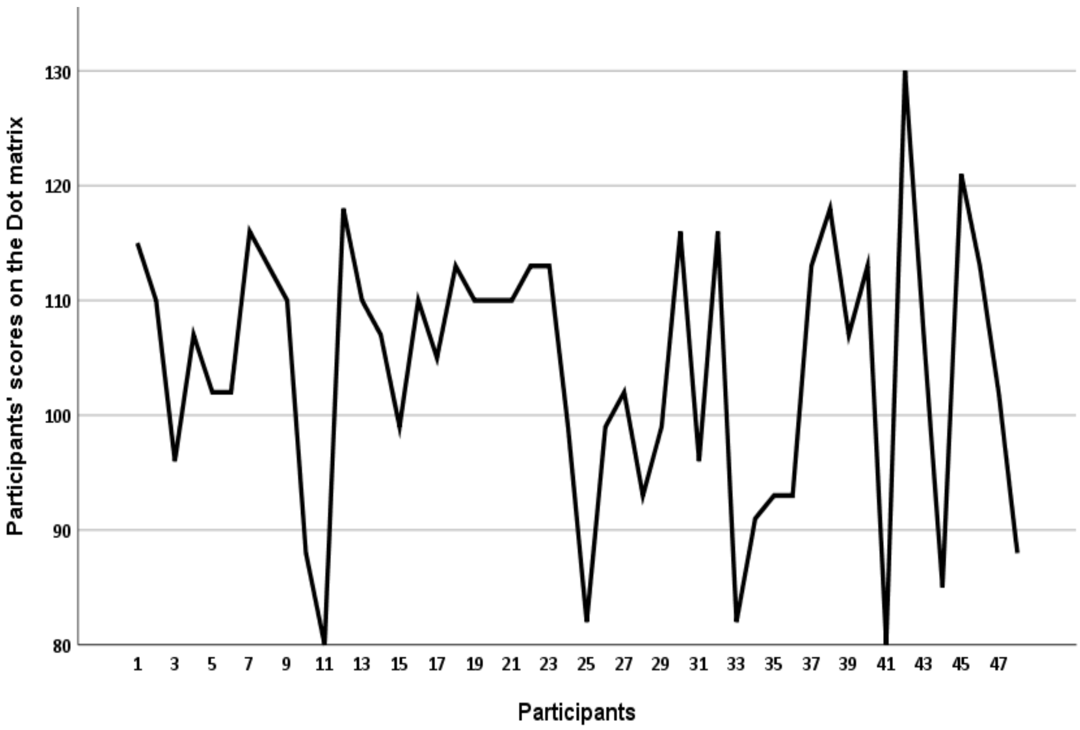


Figure 5.3 Participants' scores on the dot matrix test

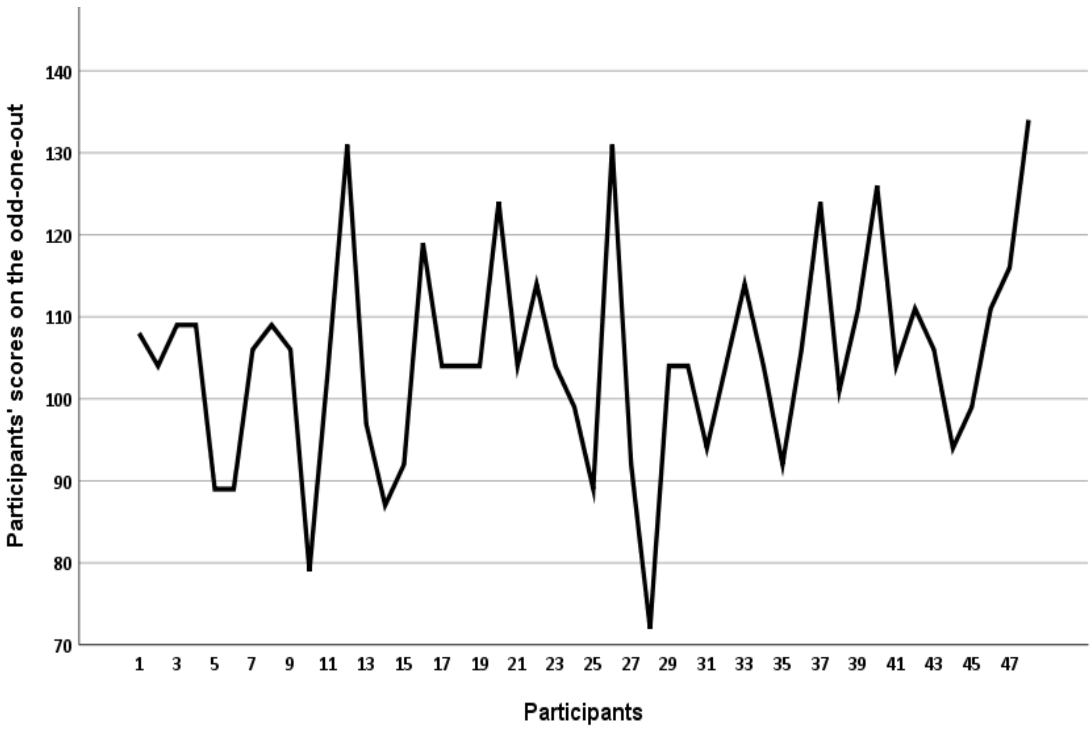


Figure 5.4 Participants' scores on the odd one out test

From the table and figures above, it can be seen that the participants' scores on the simple verbal forward digit recall test, which ranged between 69 and 128, were the lowest compared to their scores on the other three WM measures. On the other hand, Figure 5.4 shows that the greatest scores, which ranged between 72 and 134 were gained on the complex visual odd-one-out test. As the data in the table above demonstrates, the learners within each experimental group differed in their gained scores on each of the four WM tests. By examining the results in the table and figures above very closely, it can be noted that 16 participants scored higher than the mean, i.e. M= 92.27 and 23 students gained smaller scores than the mean on the simple verbal forward digit recall test. On the backward digit recall test, the data shows that the participants who achieved smaller scores than the mean (M= 94.38) were 30, while their counterparts who gained larger scores than the mean were 18. Participants who gained smaller scores on the dot matrix test than the mean (M=104.06) were 21, whilst 27 participants managed to obtained higher scores than the mean. With regards to the odd-one-out test, the results showed that 15 participants scored lower than the mean (M=104.96), whereas 21 participants gained greater scores than the mean score.

Participants across the four experimental groups gained the highest mean scores on the visual odd-one-out task (M = 104.96), while the lowest mean scores were found in the forward digit recall task (M= 92.27). The learners' mean scores for the simple verbal forward digit recall test (M= 92.27) were lower than their mean scores for the complex verbal backward digit recall test (M= 94.38). The mean scores for the two visual WM tests (dot matrix and the odd-one-out) were quite close to each other. As for skewness, the table above shows that the participants' scores on the verbal WM measures as well as the visual odd-one-out measure were positively skewed. This indicates that the tests were rather difficult. However, for the visual dot matrix measure the results are negatively skewed, suggesting that the test was easy, i.e. the high scores were relatively higher than the low scores.

5.3.2 WM and incidental vocabulary short-term learning

The role of WM in incidental vocabulary short-term learning, represented by the composite scores for the three vocabulary immediate post-tests is examined in this sub-section. The results of the verbal WM measures, forward digit recall and backward digit recall are reported first; followed by the outcomes of the visual WM tests, dot matrix and the odd-one-out.

5.3.2.1 Verbal WM measures

This part aims to explore whether the verbal WM system (the phonological loop), measured by the simple and complex memory tests affected incidental vocabulary learning through the multiple modalities of audio-visual input. In particular, it intends to examine whether participants with high verbal WM capacity better benefited from the multiple modalities of input than those with low verbal WM capacity.

5.3.2.2 Forward digit recall test

Figure 5.5 presents the means for the composite vocabulary scores of participants with high and low WM capacities (measured by the verbal simple forward digit recall test) across the four experimental groups. The outcomes showed that students with high verbal WM capacity in the VAC and A only groups had higher composite vocabulary scores (M = 33.17, M = 25.57) than their counterparts with low verbal spans (M = 31.00, M = 20.33, respectively). In contrast to this, students with high verbal WM capacity involved in the VA and CA conditions gained smaller composite vocabulary scores (M = 40.00, M = 36.00) than their counterparts with low verbal WM capacity (M = 42.17, M = 40.33, respectively). A mixed design 2 x 4 ANOVA test was employed to explore whether these differences were statistically significantly different, and whether there was an interaction between the verbal WM system and the different modalities of audio-visual input.

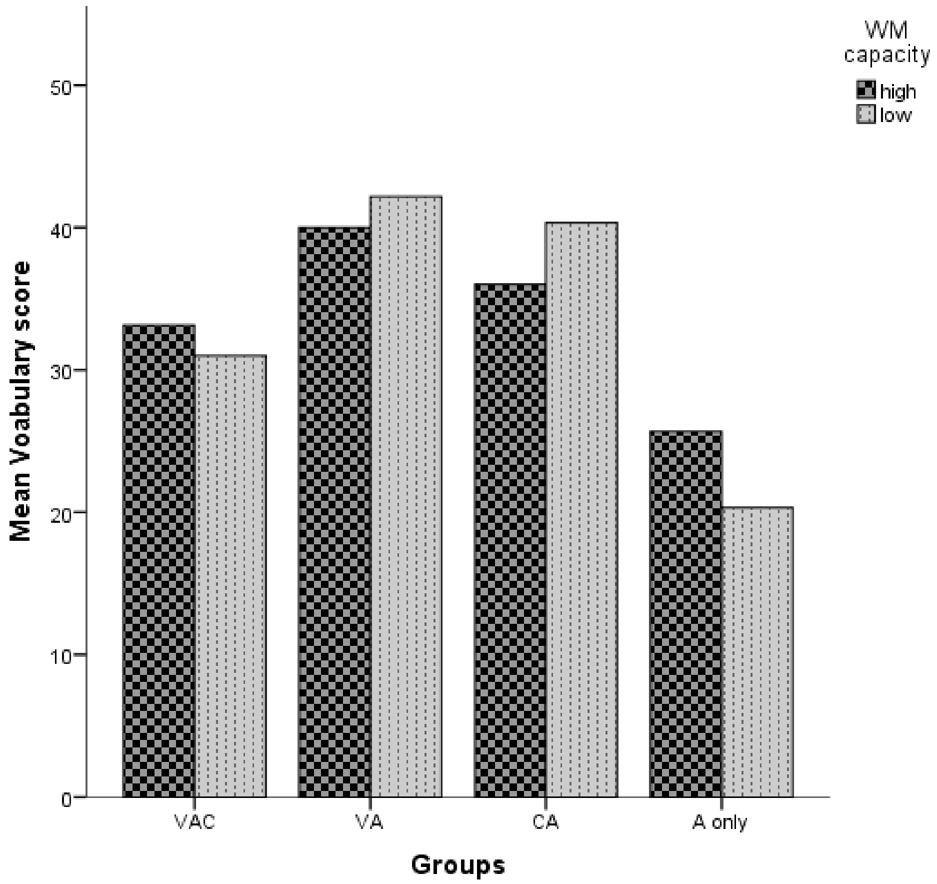


Figure 5.5. Means for the composite vocabulary scores representing the immediate post-tests of students with high and low WM spans based on the forward digit recall test across the four treatment groups

The results of the mixed design ANOVA (given in Appendix L) demonstrated a significant main effect for the modalities of input, $F(3, 40) = 13.185$, $p < .000$, partial $\eta^2 = .497$. However, non-significant effects were found for WM, measured by the forward digit recall test, $F(1, 40) = .012$, $p > .913$, partial $\eta^2 = .000$ as well as for the interaction between the WM measure and the different modalities of input, $F(3, 40) = .967$, $p > .418$, partial $\eta^2 = .068$. These nonsignificant results may suggest that individual differences in WM span (measured by the simple verbal forward digit recall test) did not account for the variations in the composite vocabulary scores across the four audio-visual input conditions. The nonsignificant interaction between WM and experimental conditions may indicate that the verbal WM system did not mediate the effects of the different input modalities.

5.3.2.3 Backward digit recall test

Figure 5.6 displays the means for the composite vocabulary scores of participants with high and low WM capacities (measured by the backward digit recall test) across the four input conditions. The results revealed that across the four treatment groups, participants with high verbal memory spans had higher composite vocabulary scores than low verbal WM span participants. A mixed design 2 x 4 ANOVA test was run to find out whether these differences were statistically significantly different, and whether there was an interaction between the verbal WM system and the different modalities of audio-visual input.

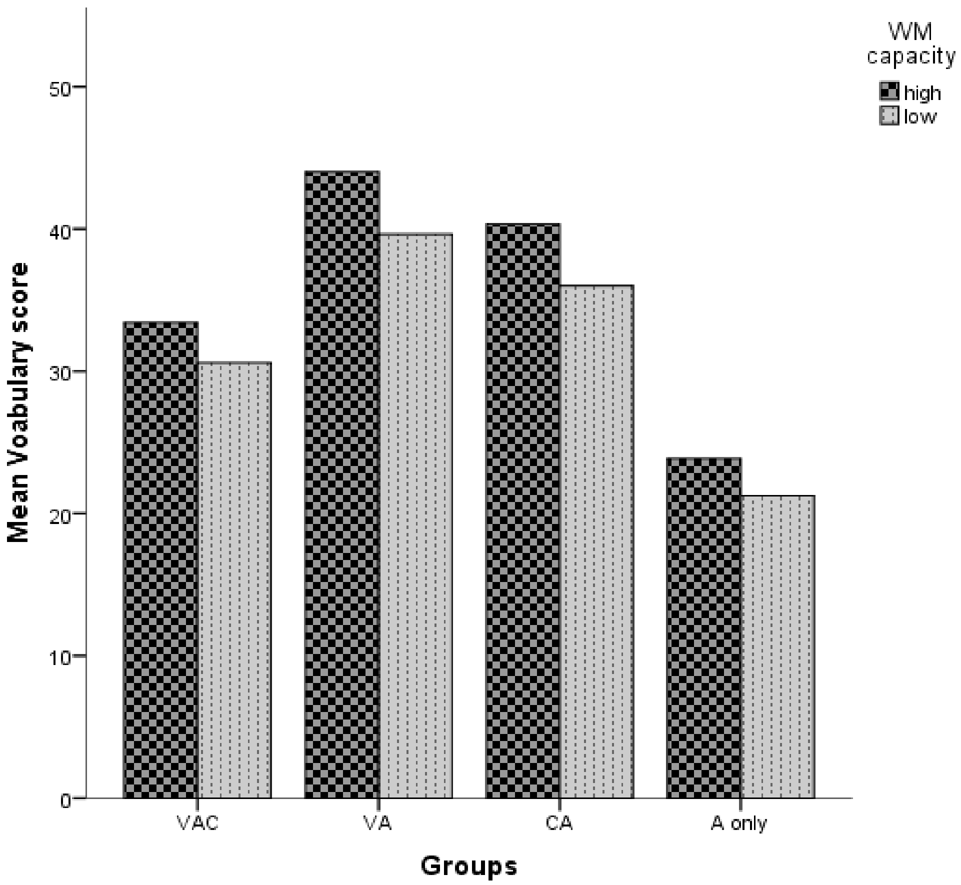


Figure 5.6. Means for the composite vocabulary scores representing the immediate post-tests of students with high and low WM spans based on the backward digit recall test across the four treatment groups

The mixed design ANOVA (see Appendix L) showed a significant main effect for the modalities of input, F (3, 40) = 13.071, $p < .000$, partial $\eta^2 = .495$. However, the main effect for WM, measured by the backward digit recall test, F (1, 40) = 2.400, $p > .129$, partial $\eta^2 = .057$ as well as the interaction between the WM measure and the different modalities of audio-visual input were nonsignificant, F (3, 40) = .042, $p > .988$, partial $\eta^2 = .003$. These nonsignificant outcomes may suggest that individual differences in WM capacity (measured by the complex verbal backward digit recall test) did not account for the variations in the composite vocabulary scores across the four audio-visual input conditions. The nonsignificant interaction between WM and the input modalities may suggest that the verbal WM system did not mediate the effects of the different input modalities.

5.3.2.4 Visual WM measures

This part aims to explore whether the visual WM system (the visuospatial sketchpad), measured by simple and complex tests affected incidental vocabulary learning through the different multiple modalities of audio-visual input. Particularly, this section aims to explore whether participants with high visual WM capacity better benefited from the multiple modalities of input than those with low visual WM capacity.

5.3.2.5 Dot matrix test

The means for the composite vocabulary scores of subjects with high and low visual WM spans (indexed by the dot matrix visual WM test) across the four experimental groups are presented in Figure 5.7. The findings were similar to those found in Figure 5.5 above, in that students with high scores on the dot matrix WM measure under the VAC and the A only input conditions gained slightly higher composite vocabulary scores ($M = 33.29$, $M = 23.13$) than their counterparts with low visual WM spans ($M = 30.80$, $M = 22.75$, respectively). Conversely, students with low visual WM capacity involved in the VA and CA groups gained higher composite vocabulary scores ($M = 42.00$, $M = 40.90$) than those with high WM capacity ($M = 40.78$, $M = 24.50$, respectively). A mixed design 2 x 4 ANOVA test was conducted to find out whether these differences in the composite vocabulary scores between participants with high and low visual memory spans were statistically significantly different, and whether there was an interaction between the visual WM system and the different modalities of audio-visual input.

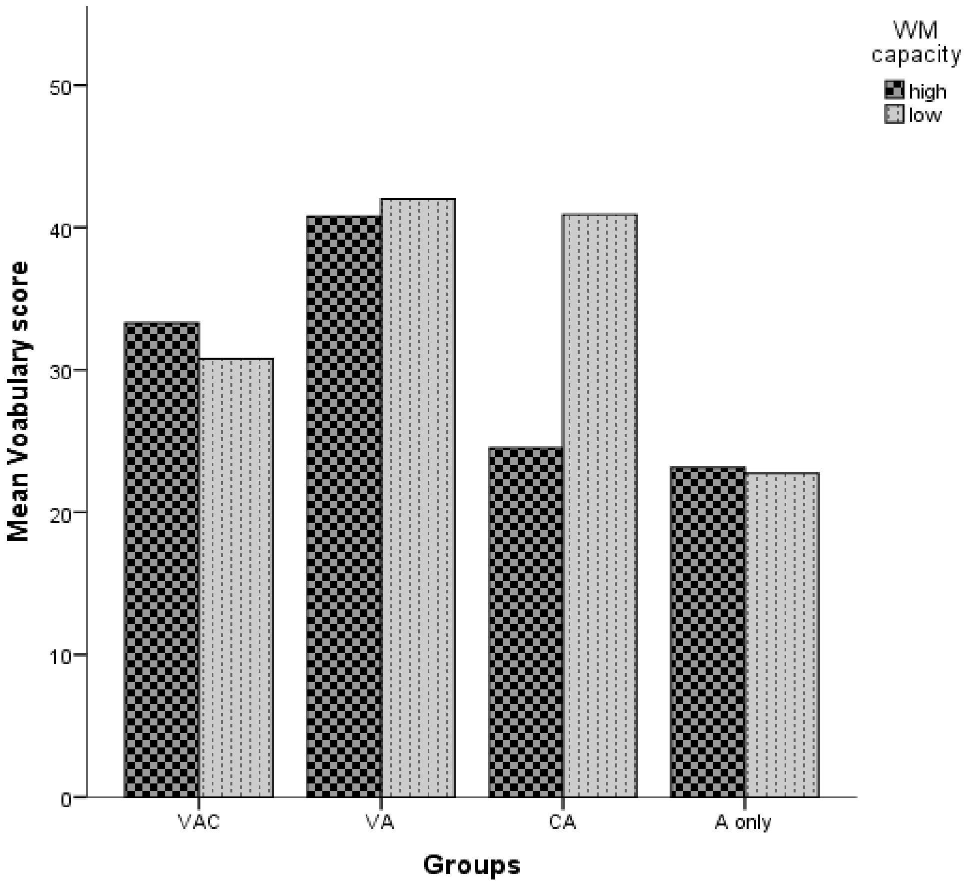


Figure 5.7. Means for the composite vocabulary scores representing the immediate post-tests of students with high and low WM spans based on the dot matrix test across the four treatment groups

The outcomes of the mixed design ANOVA (seen Appendix L) showed that the main effect for the different audio-visual conditions of input was statistically significant, $F(3, 40) = 11.017$, $p < .000$, partial $\eta^2 = .452$. The main effect for WM, measured by the simple visual dot matrix test, $F(1, 40) = 2.445$, $p > .126$, partial $\eta^2 = .058$ as well as the interaction between the visual WM measure and the modalities of input, $F(3, 40) = 2.767$, $p > .054$, $\eta^2 = .172$ were, however, nonsignificant, though this interaction between the two independent variables approached a statistical significance ($p > .054$). These nonsignificant results may be taken to imply that individual differences in WM capacity (indexed by the simple visual dot matrix test) did not account for the variations in the composite vocabulary scores. It also suggests that the visual WM system did not moderate the effects of the different modalities of audio-visual input.

5.3.2.6 The odd one out test

Figure 5.8, presents the means for the composite vocabulary scores of subjects with high and low WM spans (indexed by the odd one out visual WM test) across the four experimental groups. The means for the composite vocabulary scores of students with high visual WM spans across the three treatment groups (VAC, VA, and A only) (M = 32.78, M = 43.75, M = 23.22) were higher than the means of their counterparts with low WM capacity (M = 30.67, M = 35.75, M = 22.33, respectively). On the other hand, the low visual WM capacity group under the CA condition gained slightly higher means (M = 39.40) than those with high visual WM capacity (M = 37.29). To determine whether these differences in the composite vocabulary scores between participants with high and low visual WM capacities were statistically significantly different, and whether there was an interaction between the visual WM system and the four input conditions, a mixed design 2 x 4 ANOVA test was run.

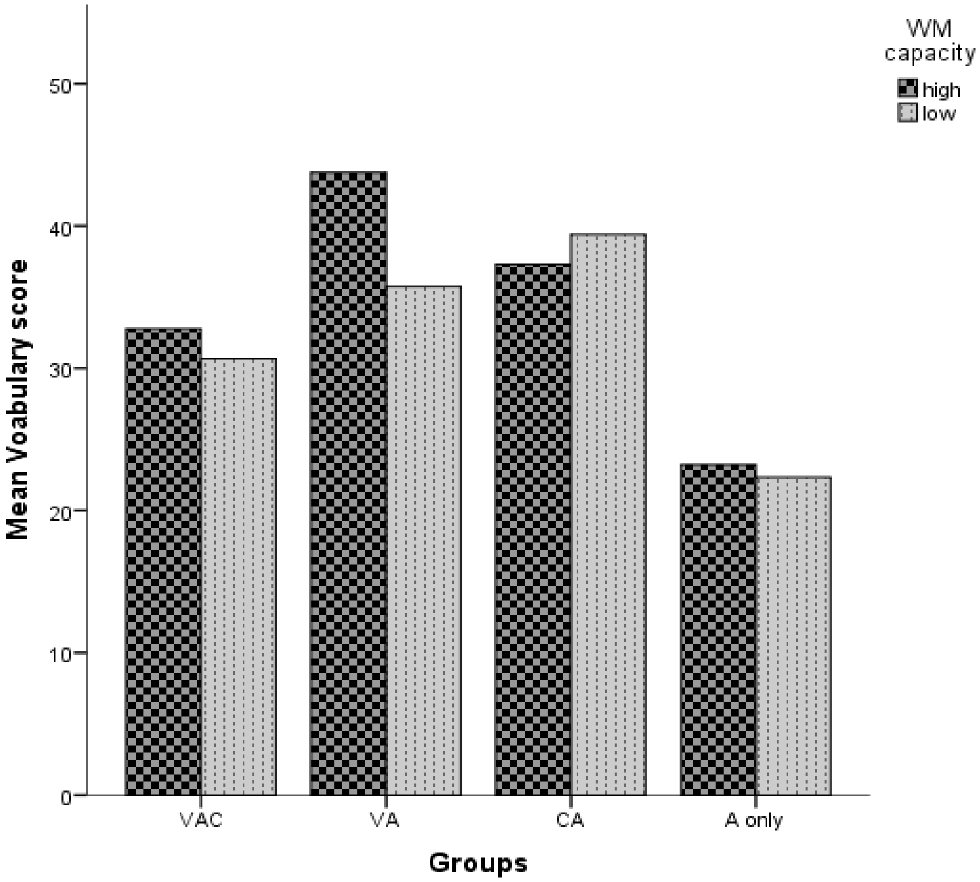


Figure 5.8. Means for the composite vocabulary scores representing the immediate post-tests of students with high and low WM spans based on the odd one out test across the four treatment groups

The findings of the mixed design ANOVA (see Appendix L) demonstrated a significant effect for the different modalities of input, $F(3, 40) = 10.160$, $p < .000$, partial $\eta^2 = .432$. The ANOVA test, however, revealed a nonsignificant effect for WM, measured by the complex visual odd one out test, $F(1, 40) = .856$, $p > .361$, partial $\eta^2 = .021$ as well as for the interaction between the visual WM system and the input conditions, $F(3, 40) = .858$, $p >.471$, partial $\eta^2 = .060$. These non-significant findings may suggest that individual differences in WM capacity (measured by the complex visual odd one out test) did not account for the variations in the composite vocabulary scores. The results also suggest that the visual WM system did not moderate the effects of the different modalities of audio-visual input.

Up to this point, the data concerning the role of WM in incidental vocabulary short-term learning was analysed. In the following sections, the role of WM in incidental vocabulary long-term retention is examined.

5.3.3 WM and incidental vocabulary long-term retention

The role of WM in incidental vocabulary long-term retention, represented by the composite scores for the three vocabulary delayed post-tests is investigated in this section. The results of the verbal WM measures, forward digit recall and backward digit recall are reported first; followed by the outcomes of the visual WM tests, dot matrix and the odd-one-out.

5.3.3.1 Verbal WM measures

This section aims to explore whether the verbal WM system (the phonological loop), measured by simple and complex tests affected vocabulary long-term retention through the multiple modalities of audio-visual input. More importantly, the section aims to examine whether participants with high verbal WM capacity better benefited from the multiple modalities of input than those with low verbal WM capacity.

5.3.3.2 Forward digit recall test

The composite vocabulary scores of participants with high and low WM spans (gauged by the verbal forward digit recall WM test) across the four experimental groups are shown in Figure 5.9. The findings showed that participants with high WM capacity in the VAC treatment group, had slightly better composite vocabulary scores ($M = 12.00$) than their counterparts with low WM spans ($M = 11.40$). In contrast, across the other three groups (VA, CA, and A only) the composite scores of subjects with high WM span ($M = 17.67$, $M = 10.17$, $M = 7.17$) were lower than those with low WM spans ($M = 20.33$, $M = 16.17$, $M = 10.83$, respectively). A mixed design 2 x 4 ANOVA test was conducted to explore whether these differences were statistically significantly different, and whether there was an interaction between the verbal WM system and the four modalities of audio-visual input.

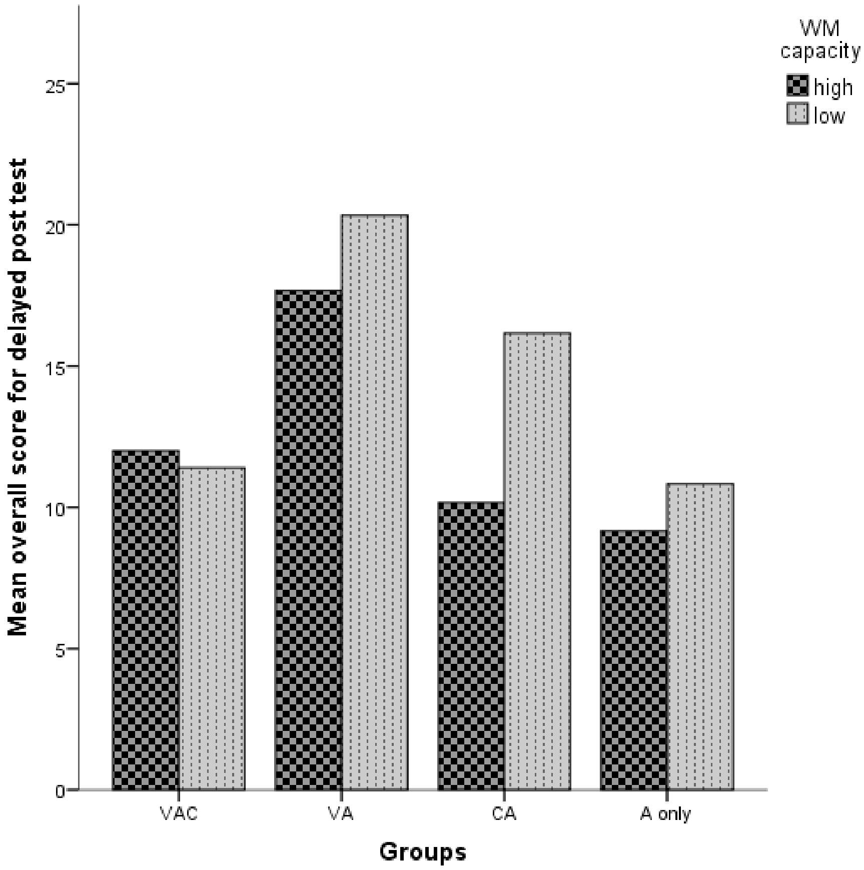


Figure 5.9. Means for the composite vocabulary scores representing the delayed post-tests of students with high and low WM spans based on the forward digit recall test across the four treatment groups

The findings of the mixed design ANOVA (see Appendix M) indicated that the main effect for the first independent variable, different modalities of audio-visual input, was statistically significant, $F (3, 40) = 3.791$, $p < .017$, partial $\eta^2 = .221$. However, the main effect for WM, tested by the simple verbal forward digit recall, was statistically non-significant, $F (1, 40) = 1.461$, $p > .234$, partial $\eta^2 = .035$. Likewise, the effect for the interaction between the independent variables, the WM measure and the different modalities of input was not significant, $F (3, 40) = .462$, $p > .710$, partial $\eta^2 = .033$. These nonsignificant outcomes may suggest that individual differences in WM capacity (measured by the simple verbal forward digit recall test) did not account for the variations in the composite vocabulary scores across the four input modalities. The non-significant interaction between WM and the different modalities of audio-visual input indicates that the verbal WM system did not mediate the effects of the different input modalities on long-term retention of the target words.

5.3.3.3 Backward digit recall test

Figure 5.10 illustrates the means for the composite vocabulary scores of students with high and low verbal WM capacity (gauged by the verbal backward digit recall test) across the four input conditions. As shown in the figure below, students with high WM capacity in the VAC, CA, A only experimental groups had greater composite vocabulary scores (M = 12.43, M = 14.50, M = 10.13) than their counterparts with low WM capacity (M = 10.80, M = 11.83, M = 9.75, respectively). Conversely, the mean scores of low WM span subjects involved in the VA group (M = 19.38) were higher than those of high WM span subjects (M = 18.25). To determine whether these differences were significantly different, and whether there was an interaction between the verbal WM system and the four input conditions, a mixed design 2 x 4 ANOVA test was carried out.

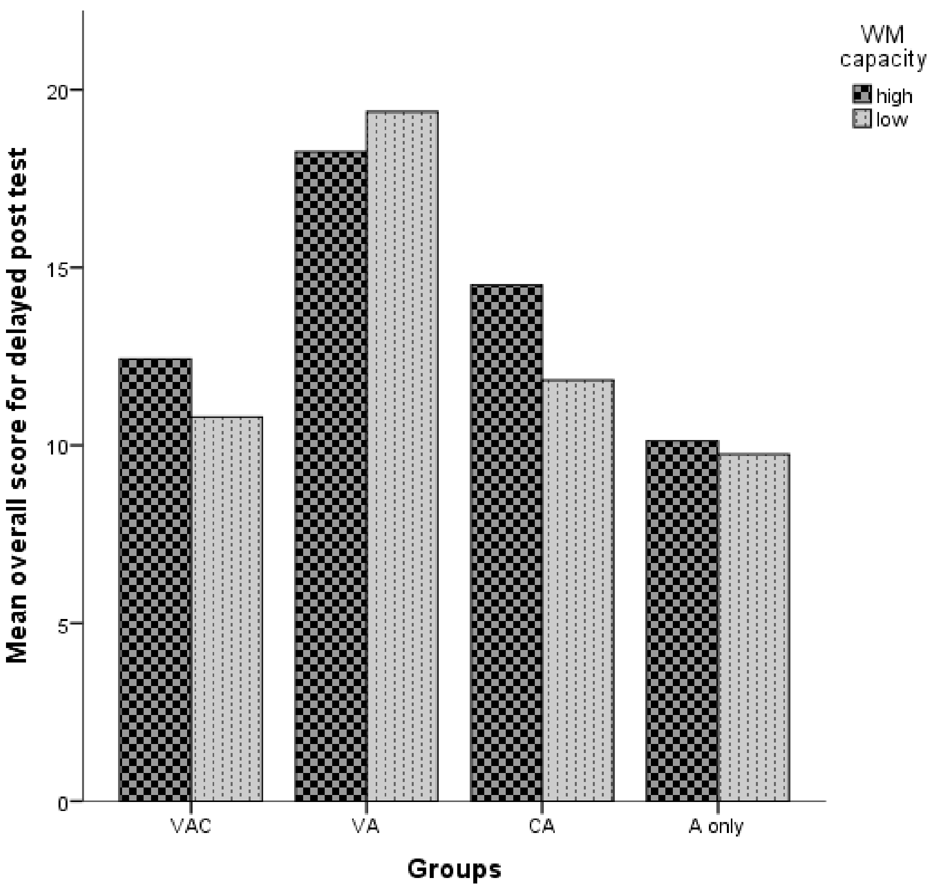


Figure 5.10. Means for the composite vocabulary scores representing the delayed post-tests of students with high and low WM spans based on the backward digit recall test across the four treatment groups

The results of the ANOVA test (in Appendix M) showed a significant effect for the first independent variable, different modalities of input, $F (3, 40) = 3.129$, $p < .036$, partial $\eta^2 = .190$. On the other hand, the effect for WM, indexed by the complex verbal backward digit recall test, $F (1, 40) = .173$, $p > .680$, partial $\eta^2 = .004$ as well as the interaction between the WM measure and the multiple modalities of input, $F (3, 40) = .148$, $p > .931$, partial $\eta^2 = .011$ were not significant. These results demonstrate that individual differences in WM capacity (measured by the complex verbal backward digit recall test) did not account for the variations in the composite vocabulary scores for the delayed post-tests across the four input modalities. The non-significant interaction between WM and the input conditions suggests that the visual WM system did not moderate the effects of the different input modalities on long-term retention of the target words.

5.3.3.4 Visual WM measures

This section aims to explore whether the visual WM system (the visuospatial sketchpad), measured by simple and complex tests affected vocabulary long-term retention through the multiple modalities of audio-visual input. In particular, it aims to examine whether participants with high visual WM capacity better benefited from the multiple modalities of input than those with low visual WM capacity.

5.3.3.5 Dot matrix test

The composite vocabulary scores for the three delayed post-tests of participants with high and low WM span (measured by the visual dot matrix WM test) across the four treatment groups are displayed in Figure 5.11. The figure below displays that participants with high visual WM spans in the VA and A only groups obtained higher composite vocabulary scores ($M = 20.67$, $M = 10.50$) than their counterparts with low visual WM spans ($M = 14.00$, $M = 9.00$, respectively). On the other hand, participants with high visual WM spans under the VAC and CA conditions had smaller composite scores ($M = 11.57$, $M = 5.50$) than their counterparts with low visual WM spans ($M = 12.00$, $M = 14.70$, respectively). A mixed design 2 x 4 ANOVA test was conducted to see whether these differences were significantly different, and whether there was an interaction between the visual WM system and the different modalities of audio-visual input.

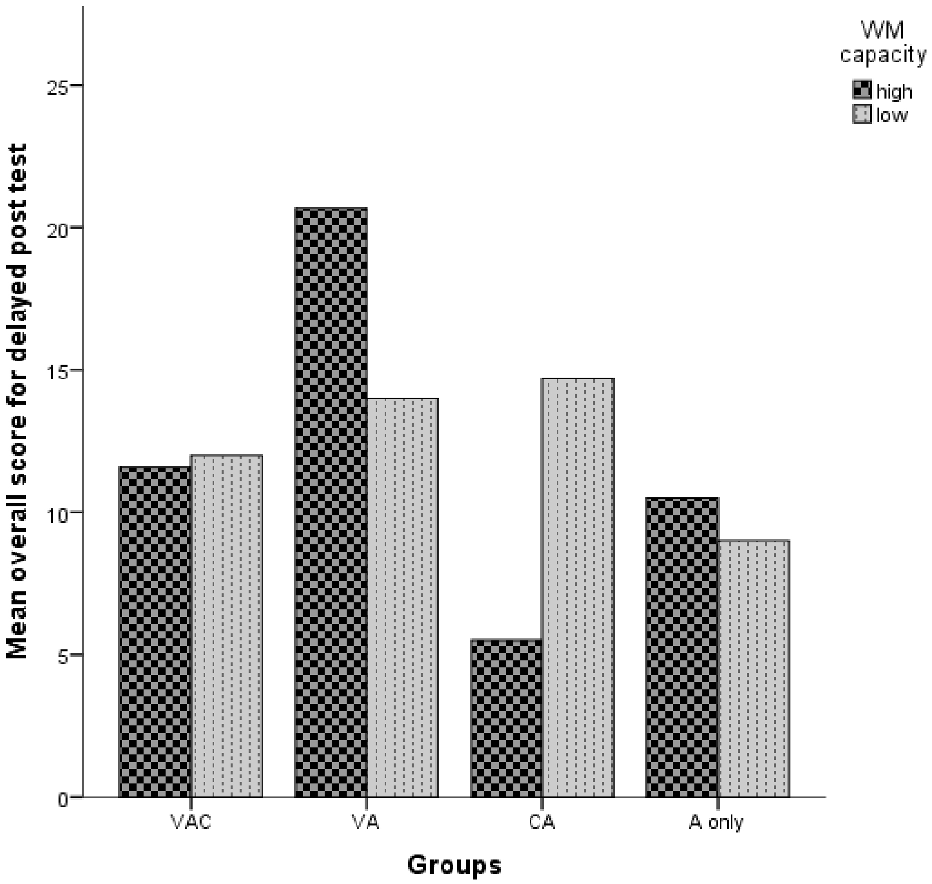


Figure 5.11. Means for the composite vocabulary scores representing the delayed post-tests of students with high and low WM spans based on the dot matrix test across the four treatment groups

The output of the mixed design ANOVA (shown in Appendix M) revealed statistically non-significant results for the main effects of the multiple modalities of input, $F (3, 40) = 2.433$, $p > .079$, partial $\eta^2 = .154$; for WM, measured by the dot matrix test, $F (1, 40) = .027$, $p > .871$, partial $\eta^2 = .001$; and for the interaction between the two independent variables, the WM measure and the multiple modalities of input, $F (3, 40) = 1.807$, $p > .161$, partial $\eta^2 = .119$. These non-significant findings may be taken to indicate that individual differences in WM capacity (measured by the simple visual dot matrix test) did not account for the variations in the composite vocabulary scores for the delayed post-tests across the four input modalities. The non-significant interaction between WM and the experimental groups also indicates that the visual WM subset did not mediate the effects of the different input modalities on long-term retention of the target words.

5.3.3.6 The odd one out test

Figure 5.12 presents the means for the composite vocabulary scores of participants with high WM span and low visual WM capacities (gauged by the visual odd one out WM test) across the four treatment conditions. The findings indicated that the composite vocabulary scores of subjects with high visual WM spans in the VA and A only groups (M = 21.00, M = 10.33) were higher than the scores of subjects with low visual WM spans (M = 15.00, M = 9.00, respectively). In contrast, high visual WM span subjects under the VAC and CA conditions had smaller composite scores (M = 11.44, M = 11.57) than their counterparts with low visual WM spans (M = 12.67, M = 15.40, respectively). To examine whether these differences in scores between participants with high and low WM capacities were significantly different, and whether there was an interaction between the visual WM subset and the different input conditions, a mixed design 2 x 4 ANOVA test was computed.

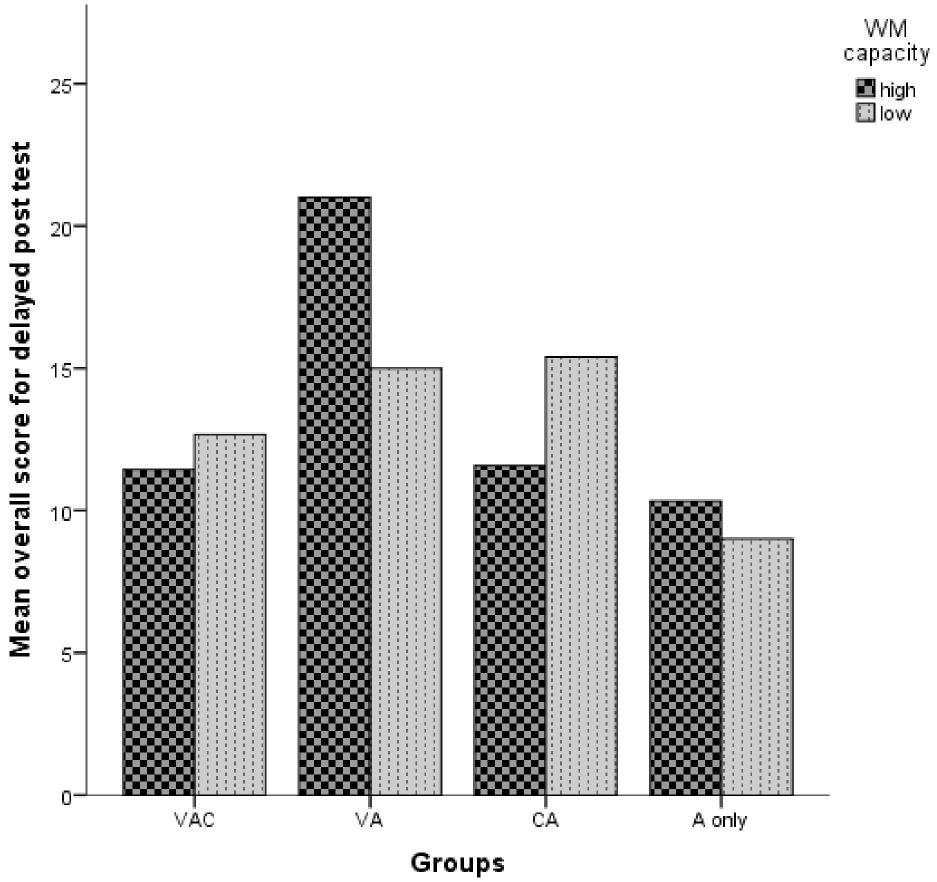


Figure 5.12. Means for the composite vocabulary scores representing the delayed post-tests of students with high and low WM spans based on the odd one out test across the four treatment groups

The outcomes of the mixed design ANOVA (presented in Appendix M) showed that the main effect for the first independent variable, different modalities of input, $F (3, 40) = 2.544$, $p > .070$, partial $\eta^2 = .160$ was not significant. Similarly, the effect for the second independent variable, WM, $F (1, 40) = .067$, $p > .796$, partial $\eta^2 = .002$, as well as the interaction between the WM measure and the multiple modalities of input, $F (3, 40) = .995$, $p > .405$, partial $\eta^2 = .069$ were nonsignificant. These non-significant results may point out that individual differences in WM capacity (tapped by the complex visual odd one out test) did not account for the variations in the composite vocabulary scores for the delayed post-tests across the four input conditions. The non-significant interaction between WM and the experimental conditions may also indicate that the visual WM system did not mediate the effects of the different input modalities on long-term retention of the target words.

5.3.4 Summary of key findings:

This last part of this chapter presented the analyses of the last research question, which examined the role of WM in moderating the effects of the different modalities of audio-visual input on incidental vocabulary short-term learning and long-term retention. The most important findings revealed in this section were:

- Individual differences in WM capacity, measured by the verbal and visual tests did not account for the variations in the composite vocabulary scores for the three vocabulary immediate post-tests and three delayed post-tests across the four input conditions;
- WM capacity did not mediate the effects of the different input modalities on short-term learning and long-term retention of the target words.

5.4 Chapter summary

The chapter presented the analyses of WM data collected through four measures: forward digit recall, backward digit recall, dot matrix, and the odd one out. The most significant findings revealed in this chapter were:

- WM was closely related to incidental vocabulary short-term learning;
- The verbal WM system, the phonological loop, contributed significantly to the relationship between WM and incidental vocabulary short-term learning;
- The simple and complex verbal WM tasks, forward digit and backward digit, contributed differently to the relationship between WM and incidental vocabulary learning;
- The effects of the different modalities of audio-visual input (VAC, VA, CA and A only) on incidental vocabulary learning and retention were not mediated by individual differences in WM capacity;
- Nonsignificant relationship was found between WM and incidental vocabulary long-term retention

The following chapter discusses the key findings found in chapters four and five.

Chapter 6 Discussion

The present research, uses a quasi-experimental design, aimed at examining the effects of four different modalities of audio-visual input: (1) video, audio and caption (VAC); (2) video and audio (VA); (3) caption and audio (CA) and audio only (A only) on L2 incidental vocabulary short-term learning and long-term retention of three distinct aspects of word knowledge, spoken form recognition, meaning recall and meaning recognition. Students' vocabulary knowledge was assessed through a spoken form recognition test, spoken meaning recall test, and spoken meaning recognition test. The three vocabulary tests were administered at three different times as: pre-tests; immediate post-tests; and delayed post-tests. The study also aimed at looking at the effects of both frequency of occurrence of the target lexical items and working memory (WM) on incidental vocabulary learning and retention. The study ran four verbal and visuospatial WM tasks to measure the learners' WM capacity: forward digit recall, backward digit recall, dot matrix, and the odd-one-out.

This penultimate chapter is dedicated to the discussion of the key findings presented in the preceding two chapters in light of the relevant research literature and the theoretical underpinnings guiding this study. Firstly, the chapter presents the interpretations of the findings of vocabulary data and then the discussion of the outcomes of the WM data. The structure of the chapter is based on the study's research questions. Accordingly, the chapter opens with the results regarding the impact of the different input conditions on incidental vocabulary learning. It then moves on to discuss the findings regarding the differential effects of the four audio-visual conditions on vocabulary knowledge of spoken form recognition, meaning recall, and meaning recognition. The third section presents the discussion of the findings of the delayed post-tests. The fourth part addresses the outcomes concerning the effects of frequency of occurrence on incidental vocabulary learning. Finally, the last three sections discuss the findings obtained from the WM data.

6.1 Section one: research question 1

R-Q 1: To what extent do audio-visual conditions, VAC, VA, CA, and A only, enhance/promote L2 incidental vocabulary short-term learning of spoken form recognition, meaning recall, and meaning recognition?

The purpose of this research question was to explore the extent to which the four modalities of audio-visual input (VAC, VA, CA, and A only) contributed to L2 incidental vocabulary short-term learning of the three different aspects of vocabulary knowledge, spoken form recognition, meaning recall and meaning recognition. The present study expanded upon earlier studies in different ways. Firstly, it developed a desing that would to tease apart the effects of different modalities of input (visual input, wirtten input, and spoken input) on the learning and retention of three vocabulary knowledge constructs, spoken form recognition, meaning recall, and meaning recognition. It is also among the few studies that focused on the effect of a key item-related variable (frequency of occurrence) on incidental vocabulary learning through different audio-visual input conditions. Unlike the majority of the earlier studies, the present study also examined the long-term effects of the different modalities of audio-visual input on incidental vocabulary retention of the three vocabulary knowledge dimensions. Additionally, this study looked at the role of an important learner-related factor (WM) in incidental vocabulary learning and retention through different modalities of audio-visual input.

To answer the first research question, the findings revealed that incidental vocabulary learning did take place through the four input conditions, VAC, VA, CA, and A only. That is, the four audio-visual input conditions led to significant learning gains of the three vocabulary knowledge facets, spoken form recognition, meaning recall, and meaning recognition. These findings add further evidence to the large body of research that indicates incidental learning of different aspects of lexical knowledge occurs through different input modalities: written mode (reading) (Al-Homoud & Schmitt, 2009; Brown et al., 2008; Horst, 2005; Laufer & Rozovski-Roitblat, 2015; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Webb, 2008), aural mode (listening) (Ellis, 1995; Van Zeeland & Schmitt, 2013; Vidal, 2011), and audio-visual modes (viewing) (Montero Perez et al., 2014; Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Winke et al., 2010).

In particular, the four experimental groups have significantly improved their knowledge of the 36 target words across the three vocabulary knowledge types after participating in the research intervention, as their mean scores on the immediate post-tests were significantly greater than their scores on the pre-tests (see Figures 4.1, 4.2, 4.3 and Tables, 4.3, 4.4, 4.5). On average, for spoken form recognition knowledge, all the four experimental groups had a total score of M = 2.31 on the pre-test, which then significantly increased to M = 14.18 on the immediate post-test. On the spoken meaning recall pre-test, these four groups had a total score of M = 0.68 which also significantly increased to M = 6.57 on the immediate post-test. Similarly, they had an overall score of M = 1.33 on the meaning recognition pre-test which significantly increased to M = 10.81 on the immediate post-test. Importantly, learners across the four experimental groups were able to pick up between two and five new words from each of the four 15-minute video clips (of the 9 tested items in each video). However, gain rates varied significantly across the four groups, this will be discussed in the next section. Although these are encouraging gains, they may be an underestimation of actual gains that might have taken place from these four input conditions for the following reasons. Firstly, the 36 target words were drawn from the 2.000 (K2) most frequent words list and infrequent words list (off-list), however words from the 1.000 (K1) most common words list and academic words list (AWL) may have been learned but not tested. In addition, words from the K2 list and off-list with one occurrence in the video clips were not included as target words, yet it has been indicated that one encounter with the target item could lead to some sort of learning (Pellicer-Sánchez & Schmitt, 2010). The chances of learning words with one exposure increase, when they are well-supported visually or are particularly salient in the materials (Rodgers, 2013). Thus, it is likely that the participants have increased their knowledge of words with one occurrence, but were not tested.

The present results, which seem to suggest that L2 learners can learn words incidentally after viewing context-rich video clips, align with the findings of several audio-visual studies (e.g., Montero Perez et al., 2014; Montero Perez et al., 2018; Peters et al., 2016; Peters & Webb, 2018). For example, Peters et al. (2016) found that under the captions and subtitles treatment conditions, their learners were able to improve spoken form knowledge of 38 % to 48% of the target lexical items and 19 % to 20% of meaning recall knowledge after watching a short 13-minute documentary video clip once. Similarly, in their study, Montero Perez et al. (2018) indicated that their treatment groups (full captioning; keyword captioning; and glossed keyword captioning) were able to increase their knowledge of different vocabulary aspects after viewing three relatively short video clips. In particular, out of the 18 target words tested, the subjects across the treatment groups were able to learn between 8 and 11 in the form recognition test, 3 and 6 in the meaning recall test, and between 10 and 12 in the meaning recognition test. Altogether, the findings of the present study along with the findings of the previous studies suggest that the different modalities of audio-visual input have the potential to foster incidental learning of several vocabulary knowledge constructs.

Another interesting finding that have emerged was that the range of the increase in the three vocabulary knowledge types varied noticeably, irrespective of the treatment conditions. The results showed that participants across the four experimental groups gained more scores on the receptive recognition tests (i.e., spoken form recognition and meaning recognition) than on the recall productive meaning test. In fact, learners across the four treatment groups showed an improved knowledge of about 26% to 53% of words on the receptive spoken form recognition test, about 18% to 39% of words on the receptive spoken meaning recognition test, and the range of increase reported in the productive vocabulary, spoken meaning recall test was between 11% and 25% lexical items. Thus, the order of acquisition of the three vocabulary knowledge dimensions measured in this study was as follows: spoken form recognition > meaning recognition > meaning recall. These findings, seem to imply that the four different modalities of audio-visual input are more likely to have a bigger effect on receptive knowledge aspects (i.e., form recognition and meaning recognition) than on productive knowledge aspect (i.e., meaning recall). Therefore, regardless of the input modality, it can be concluded that incidental acquisition of the receptive knowledge aspects seems to develop faster than incidental acquisition of the productive knowledge aspects.

These findings are in agreement with the outcomes of several past studies which showed that participants made larger gains on the receptive tests than on the productive tests (Brown et al., 2008; Montero Perez et al., 2014; Montero Perez et al., 2018; Peters et al., 2016; Peters & Webb, 2018; Van Zeeland & Schmitt, 2013; Waring & Takaki, 2003). Even though the present study and Peters et al.'s (2016) study adopted different methodologies, the findings of the current study closely corroborate their findings. Peters et al. (2016) found that the learning rates reported in the spoken form recognition test, (which ranged from 32.4% to 48.2%) were higher than those reported in the meaning recall test (ranging from 19.3% to 20.8%). A similar pattern of results was also found by Montero Perez et al. (2018), who reported that knowledge of form was developed better than knowledge of meaning. That is, of the 18 target words, their participants were able to recognise the forms of 10 words, while they managed to recall the meanings of about 6 words.

There are multiple possible explanations for why participants' scores on the receptive recognition tests were higher than those on the recall/productive test. One of the possible reasons, according to Van Zeeland and Schmitt (2013) and Nation (2001), is that receptive vocabulary knowledge precedes productive knowledge and acquiring a word receptively is far easier and faster than learning it productively. In their study, Van Zeeland and Schmitt (2013) found that acquiring meaning at a productive level was the hardest, while receptive knowledge of form was the easiest to develop.

A further likely explanation is that the use of distinct vocabulary test types may have affected the subjects' performance. As mentioned in section 3.11, the present study employed three different test types to gauge incidental learning and retention of the three vocabulary knowledge dimensions. The first one was the spoken form recognition test, which evaluated the students' abilities to recognise the spoken form of the target words. This required students to indicate whether they had heard the aural forms of the items in the video clips. The test did not require learners to know the meanings of the words. The second vocabulary test used in this study was the spoken meaning recall test. In this test, the test-takers had to provide a synonym, explanation, or L1 translation for the target words without being prompted. According to Brown et al. (2008), this test type (meaning recall, or translation test) is the best indication of whether learners really have learned the meanings of the target items. The third test was the prompted spoken meaning recognition test, which was in a form of multiple-choice test. In this test, learners had to choose the right answer for each item from a number of alternatives. Thus, it could be that learners gained higher scores on the receptive tests than they did on the productive test because it may seem easier for them to choose the right answer in a multiple-choice meaning recognition test or to tick the familiar spoken word form in the Yes/No box than to produce the meanings in the meaning recall test, which they may have found challenging.

To summarise, the findings of the first research question revealed that L2 incidental vocabulary short-term learning is possible through the four different audio-visual input modalities of VAC, VA, CA, and A only. The results of the first research question indicated that participants across the four experimental groups were able to significantly increase their vocabulary knowledge from the pre-tests to the immediate post-tests, which may imply that the treatment had a positive effect on vocabulary development. The effects of these four audio-visual input modalities were more pronounced in the two receptive knowledge aspects (spoken form recognition and meaning recognition) than the productive knowledge aspect (meaning recall). The next section presents the discussion of the differential effects of the four different audio-visual input conditions on incidental vocabulary learning.

6.2 Section two: research question 2

R – Q 2: Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary short-term learning of the spoken form recognition, meaning recall, and meaning recognition?

The central aim of this study was to compare the differential effects of the four audio-visual input conditions (VAC, VA, CA and A only) on incidental learning of three aspects of vocabulary knowledge, spoken form recognition, meaning recall, and meaning recognition. As explained in section 3.8, students involved in the VAC condition watched the videos with audio and full captions. Students in the VA condition viewed the videos with audio without captions. Students in the CA condition listened to the audio of the videos while reading on-screen full captions without watching the videos. Students in the A only condition listened only to the soundtracks of the videos without viewing the videos nor reading the captions. The aim of this section is to present the discussion of the principle findings related to the second research question.

The findings of the first research question have established that the four treatment conditions (VAC, VA, CA, and A only) significantly contributed to incidental vocabulary growth of the three vocabulary knowledge types. To compare the differential effects of the four modalities on the learning of the three vocabulary knowledge aspects, a series of one-way ANOVA tests were conducted. The findings of the ANOVA tests demonstrated that the four conditions had differential effects on the learning of the three knowledge constructs. In particular, the three multimodal input conditions - VAC, VA, and CA - were more effective for incidental vocabulary acquisition of the three types of word knowledge than the single input condition, A only. That is, learning gains reported in the spoken form recognition, spoken meaning recall, and spoken meaning recognition tests of the VAC, VA, and CA groups were significantly higher than the gains of the single input condition (A only). Precisely, the results from the spoken form recognition test demonstrated that the vocabulary mean scores of the three multimodal groups, which ranged from M = 12.56 to M = 19.13, were significantly larger than the mean score of the A only group (M = 9.36). On the spoken meaning recall test, the mean scores of the three multimodal groups, which ranged from M = 6.42 to M = 9.15, were also significantly larger than the mean score of the single input group (M = 4.08). Similar patterns were also found in the spoken meaning recognition test in which the mean (M = 6.56) of the single input group (A only) was significantly smaller than the means of the VAC, VA, and CA groups, M = 11.12; M = 14.19; M = 11.07, respectively.

These expected results, which confirmed the effectiveness of multiple modalities of input for enhancing incidental vocabulary learning, are consistent with the results of a number of earlier studies that compared the effectiveness of various multimodal learning conditions (Akbulut, 2007; Al Seghayer, 2001; Bird & Williams, 2002; Chun & Plass, 1996; Montero Perez et al., 2014; Montero Perez et al., 2018; Sydorenko, 2010). These studies provided unambiguous evidence for the effectiveness of multimodal input modes over a single input mode. For example, Chun and Plass (1996) found that the participants who had access to combined textual and pictorial glosses performed better on the vocabulary tests than their counterparts who had access to textual glosses only. In the same vein, Akbulut (2007) revealed that the dual gloss modes of text and pictures and text and videos were more effective for promoting vocabulary learning of form recognition, meaning recognition and production, than a single condition. Although the evidence regarding the differential impacts of different multimodal input modes is still inconclusive, the present findings and the findings of these previous studies suggest that multimodal input modes are more effective than a single mode for facilitating incidental learning of different vocabulary knowledge aspects.

There is a range of plausible explanations for the potential advantage of the multiple modalities of audio-visual input over the single modality input. Firstly, as for the positive effectiveness of multiple modalities of input, the literature indicates that multiple modalities of input offer a potential for learning by presenting the target materials through different channels (Guichon & McLornan, 2008). The provision of different input modes in these multimodal conditions, VAC, VA, and CA may have helped the participants to develop their knowledge of the target words due to the fact that they had multiple channels of information to learn from. Another merit for the multiple modalities of input that may have contributed to the improvement in vocabulary knowledge can be attributed to the fact the multimodal learning conditions were designed in view of human cognitive abilities (Mayer, 2014). Baddeley and Hitch's (1974) assumed that WM consists of two modality-specific systems, the phonological loop responsible for verbal information, and the visuospatial sketchpad, which deals with visuospatial information. Presenting information verbally and visually allows for a better use of these cognitive resources (i.e., the phonological loop and the visuospatial sketchpad) which in turn optimizes the limited capacity of WM systems. Since the audio-visual conditions (VAC, VA) presented visual and verbal information, learners in these conditions may have benefited from the use of both modality-specific systems of WM to maintain and encode the presented information, which according to Mayer (2014), would enhance the learning and the recollection of information.

Further possible reasons for the significant differences in learning gains between students in the CA condition and the A only condition may be due to the proposal that the provision of on-screen texts (captions) in the CA condition prompted students to notice the unfamiliar words. It is possible that captions may have drawn students' attention to the target words, as was found in some eye-tracking studies (for example, Montero Perez et al., 2015). In addition, although the captions were presented at the same pace as the auditory inputs, they may have enabled learners to segment or chunk the spoken texts. The importance of speech segmentation or chunking rests on the fact that it reliably marks meaning (Ellis, 2003). In addition, speech segmentation may help L2 learners recognise unfamiliar words and better comprehend the text. In contrast, the real-time nature of audio inputs with no explicit boundaries between words makes it difficult for L2 learners in the A only condition to notice the unfamiliar items in the auditory input, as they may not know when each sentence begins and ends (Van Zeeland & Schmitt, 2013).

Nonetheless, though the three multimodal conditions, VAC, VA, and CA were more effective for word learning than the single mode (A only), this does not deny the relative effectiveness of listening (A only condition) on incidental vocabulary learning. The findings of the present study have reported that the A only condition resulted in incidental vocabulary learning. These findings support those of past studies that confirmed that the aural mode of input (listening) could lead to some incidental vocabulary learning, though the learning gains were considerably lower than those reported in many reading and audio-visual studies (Brown et al., 2008; Van Zeeland & Schmitt, 2013; Vidal, 2003, 2011).

In facilitating incidental vocabulary learning, the present findings regarding the superiority of the multimodal conditions over the single input provides experimental evidence for Mayer's (2005; 2014) multimedia principle. The multimedia principle refers to the advantage of presenting information verbally and visually over the single input mode. Mayer (2005) argued that learners who receive a dual presentation of information, verbal and visual, tend to perform better on achievement measures than those who only receive information through a single mode. The present findings regarding the superiority of the multimodal input conditions over the single input condition mirror those of many previous studies, which provided empirical evidence for the multimedia principle (Akbulut, 2007; Al Seghayer, 2001; Chun & Plass, 1996; Montero Perez et al., 2018, to mention a few).

Moreover, the beneficial effects of the multiple modalities of input can be explained in light of the dual coding theory (DCT) of Paivio (1986). The essential feature of DCT is that humans have two interconnected but functionally independent cognitive systems responsible for processing verbal and visual information. Paivio (1991) postulated that the activation of these two cognitive systems in processing incoming verbal and visual information enhances the recollection of the presented information, as the verbal and non-verbal codes act as a complementary form of presenting information. Hence, processing information in both verbal and non-verbal forms is likely to enhance learning (Mayer, 2014). DCT posits that the advantage of multimodal presentation of information is attributed to the fact that information presented visually and verbally is encoded by the two processing cognitive systems (verbal and visual), which thence makes the recollection of information easier. Relating the findings of this study to this theoretical account, it can be argued that the participants under the multimodal conditions benefited from the visual (images of the videos/ captions) and verbal (audio) presentation of information which enabled them to establish direct cognitive connections between the dually-presented information in their WM. Thus paving the way for durable recall of the presented information.

Another interesting finding that deserves attention in this study is the differential effects of the three audio-visual conditions (VAC, VA, and CA) on the three aspects of vocabulary knowledge, spoken form recognition, meaning recall, and meaning recognition. The findings showed that these three multimodal conditions had differential effects on incidental learning of the three vocabulary knowledge dimensions. Concerning the spoken form recognition knowledge, the findings indicated that the highest mean score on the spoken form recognition test was attributed to participants in the CA group ($M = 19.13$), followed by participants in the VAC group ($M = 14.77$). This suggests that the availability of captions in such captioning conditions (CA and VAC) were more effective than the non-captioning conditions for developing form knowledge. In other words, the CA and VAC conditions boosted the learning of form better than the other non-captioning conditions (VA and A only).

These results are in line with the findings of several earlier studies that demonstrated that the captions groups recognised more correct word forms than non-captioning groups (Montero Perez et al., 2014; Montero Perez et al., 2018; Winke et al., 2010). The possible explanation for the potential advantage of captions on the learning of form is attributable to the fact that captions aid students to isolate word forms and pay attention to them (Montero Perez et al., 2014). Additionally, CA and VAC input conditions may have helped participants integrate the written (captions) and spoken input (audio) information in support of receptive recognition of word forms.

Another interesting finding relates to the effect of the VA condition on incidental learning of meanings at the two mastery levels, receptive and productive. The findings showed that the VA group significantly outperformed the CA and VAC groups on the meaning recall and meaning recognition tests. That is, students in the VA group significantly outscored (M = 9.15) students in the VAC and CA groups (M = 6.42; M = 6.47, respectively) on the meaning recall test. Similarly, the VA group gained significantly higher scores (M = 14.19) on the meaning recognition test than the VAC and CA groups (M = 11.12; M = 11.07, respectively). One such speculation for the superiority of the VA condition over the other two audio-visual conditions, VAC, and CA could be related to the fact that images of the videos helped learners to better learn about the meanings of the target words. It appears that learners in the VA group were able to correctly infer the meanings, based on the pictorial input as the target words were well-supported by visual images. Vocabulary items in the documentary video clips occurred with clear links to the imagery in the videos.

Moreover, the finding regarding superiority of the non-captioning visual condition (VA) over the captioning visual condition (VAC) in enhancing the learning of word meanings runs partially counter to Sydorenko (2010), findings, which showed that the VAC condition was more effective for meaning recall than the VA condition. The reported advantage of the VA condition over the VAC condition could be explained using Mayer's (2005; 2014) redundancy principle. The redundancy principle refers to situations where students learn better from "graphics and narration than from graphics, narration, and redundant on-screen text" (Mayer & Johnson, 2008, p. 380). Since the multimodal condition (VAC) involved simultaneous presentation of video, audio and caption (text), this may have resulted in a redundancy effect. More specifically, the redundant information may have caused 'split-attention', as participants' attention was distracted by the redundant information. In other words, since the participants in the VAC condition may have had to divide their attention between captions, audio, and images of the videos when watching captioned videos, this may have slowed down information processing and learning. These findings corroborate Sydorenko's (2010) findings, which showed that her subjects paid most attention to captions and then video, while audio received the least of their attention.

To sum up, the analyses of the second research question demonstrated that the vocabulary gains of the participants involved in the three audio-visual conditions (VAC, VA, and CA) were statistically significantly higher than those of the participants in the A only group. The three multimodal conditions had differential effects on the three vocabulary knowledge constructs. The captioning conditions were more effective for the learning of form than the non-captioning conditions. The visual condition (VA) was the most effective condition for the fostering of meaning learning.

6.3 Section three: research question 3

R – Q 3: Which of the four audio-visual conditions, VAC, VA, CA, or A only, lead to better incidental vocabulary long-term retention of the three vocabulary knowledge aspects, spoken form recognition, meaning recall, and meaning recognition?

This section serves to discuss the findings of the third research question concerning incidental long-term retention of the three learned knowledge dimensions of the 36 target words. As a reminder, the present study administered three vocabulary-delayed post-tests, spoken form recognition, spoken meaning recall, and spoken meaning recognition one-month after the end of the four-week treatment. The purpose was to examine the long-term effects of the different audio-visual input conditions (VAC, VA, CA and A only) on the participants' incidental vocabulary gained knowledge.

The results indicated that the facilitative effects of the four input modes were found in the immediate post-tests, however the effects had somewhat disappeared at the time of the delayed post-tests. Particularly, large attrition across the three vocabulary knowledge aspects over a one-month period was observed, irrespective of the input modality. Approximately more than half of the initial gains reported in the immediate post-tests were forgotten over the course of four weeks. Across the four experimental groups, there were significant differences between their mean scores for the three immediate post-tests and the scores of the three corresponding delayed post-tests. On average, across the four treatment groups, the attrition rates ranged between 43% and 72 % in the spoken form recognition test, 40% and 51% in the spoken meaning recall test, and 50% and 60% in the spoken meaning recognition test. These findings, which indicate participants in all four experimental groups only retained small amounts of the initial learning gains after one month, support the argument of memory research. This suggests that forgetting is likely to occur quickly after initial learning (Baddeley, 1997). The attrition rates are rather disappointing, considering the effort and time involved in the design and execution of the current research intervention. However, a considerable drop over time is not an uncommon phenomenon, as earlier studies affirmed that forgetting is a natural fact of learning (Schmitt, 2010; Waring & Takaki, 2003).

The present outcomes appear to concur with the findings of several past reading studies that reported small to moderate retention rates after a time interval had passed. At the one week delayed post-tests, Waring and Takaki (2003), for instance, found a steady drop in the participants' scores on the three vocabulary tests, word-form recognition, meaning recognition, and meaning recall. Their participants have forgotten nearly 28 % of forms, 25% of meaning recognition, and over 50% of meaning recall knowledge. However, in the three-month delayed post-tests, participants' attrition rates nearly doubled those of the first delayed post-tests, indicating that vocabulary knowledge would decay over time.

However, the results of the present study do not replicate those found in (Akbulut, 2007). Akbulut reported smaller attrition rates than those revealed in the present study. On average, his participants have lost about 13 % to 18 % of form recognition knowledge, 7 % to 14 % of meaning recognition knowledge, and 13 % to 20 % of meaning recall knowledge. One likely explanation for the contrast between the findings of Akbulut's and those of the present study could be related to the effect of other sources of language learning on incidental vocabulary learning. It is possible that Akbulut has used target words that were likely to be met in the formal language learning classes or in subsequent activities. Meeting the target words outside the treatment would have given the students the chance to consolidate the learned items and would reinforce their learning. In other words, the selected words might have been drawn from the most frequent word lists (i.e. K1 or AWL), which were likely to be encountered in the ongoing or concurrent classroom activities or language textbooks during the interim period. In the same way, Peters and Webb (2018) indicated that their participants' scores on the delayed post-tests were considerably higher than those found on the immediate post-tests. Peters and Webb speculated that the learners' better performance on the delayed post-tests was due to the possibility of deliberate learning taking place between the immediate and delayed post-tests administrations. However, unlike Akbulut's and Peters and Webb's studies, the present study has selected the vast majority of the target words from the off-list words that were highly unlikely to be met during the ongoing classroom instruction. There was no evidence that participants had any formal exposure to the target words in their daily classes between the two test time points, immediate post-tests and delayed post-tests. Additionally, the scores on the delayed post-tests are significantly smaller than the scores on the immediate post-tests, suggesting that participants had not met the target words outside the treatment.

One of the possible interpretations for the significant attrition rates reported in the present study lies in the fact that there were limited repeated encounters with the target words during the intervention. In fact, students encountered the vast majority of target words (about 77 % of the 36 target words) between two to seven times only, which could be the reason for the significant loss of the learned words. A number of researchers (for example, Hulstijn, 2003; Waring & Takaki, 2003) asserted that in order for students to retain long-term knowledge of the acquired words, they need to meet the target words more frequently in different contexts. In their study, Waring and Takaki (2003) found that vocabulary items encountered less than eight times were completely lost after three months. In this regard, Hulstijn (2003) and Vidal (2011) emphasized the need and importance for frequent repetitions for the target words in order to install durable memory traces of the learned words, because according to Schmitt (2008), if repetition was neglected, a substantial attrition of the recently partially acquired words will take place. In addition, since the participants across the four groups viewed/listened to the four video clips only once, this may have eliminated the chances of consolidating the learned items. Coady (1993) remarked that the single exposure to the materials could lead to only a 5-15 % probability of a long-term retention of the target words.

One of the concerns about the administration of delayed post-tests is related to the lack of control over deliberate learning that may take place in the time interval between immediate and delayed post-tests, which may render the results of delayed-post-tests unreliable, as Nation and Webb (2011) pointed out. However, the present findings provided no evidence that learning occurred between the two test time points, as a steady decrease in initial acquired knowledge at the immediate post-tests was detected at the delayed post-tests, suggesting that learners may have not been able to reinforce their acquired knowledge outside the research intervention. These findings are somewhat reassuring, as they would eliminate some of the doubt raised about the validity of the delayed post-tests in studies adopting pre-immediate and-delayed-post-test designs.

A further important finding revealed in this study was the differential effects of the four experimental groups on the long-term retention of the three vocabulary knowledge types. Despite the fact that the four experimental groups experienced a large decay, some of them showed greater attrition rates compared to the others. For example, though learners in the CA condition gained the greatest mean score (M= 19.13) of the 36 target items in the spoken form recognition immediate post-test, they showed the highest rate of decrease in the delayed post-test (M = 5.27). This surprising result seems to suggest that though this input condition (CA) was more effective for short-term learning of the spoken form than other input modes, it could not maintain its advantage for long-term retention. Furthermore, the retention rates of learners in the VA group in the spoken meaning recall test (M = 5.48) and in the spoken meaning recognition test (M = 7.19) of the 36 learned target words were significantly higher than the scores of the other three groups (VAC, CA, and A only). This audio-visual input condition (VA), which was the most effective condition for fostering short-term learning of meaning, managed to somewhat maintain this advantage in sustaining the long-term retention of meaning knowledge. These findings may imply that the VA input condition provided the best opportunities for the short-term incidental learning as well as long-term retention of meanings at the two mastery levels. The finding of the superiority of the VA condition in retaining the largest rates of the target words is in agreement with Akbulut (2007) which demonstrated that the visual groups retained more meanings than the non-visual groups.

Apart from their high learning gains on the immediate post-tests, the reason for why the participants in the VA condition managed to retain the greatest numbers of words in the delayed post-tests can be explained in terms of the DCT of Paivio (1986). Paivio postulated that using verbal and visual information would lead to the formation of inferential connections between the two types of presented information in the two cognitive processing systems (verbal and non-verbal). Such inferential connections would help form complex networks in human memory, which are said to positively influence the recall of presented information (Mayer & Anderson, 1991) and facilitate retention of information (Akbulut, 2007). Based on this theoretical perspective, it can be implied that the dual presentation of information in the VA condition, enabled students to make referential connections between the two modes of input (video and audio), which in turn led to better performance on the delayed post-tests. Chun and Plass (1996, p. 193) offered another explanation for this result, arguing that the better performance of the participants involved in the visual group was due to what is known as "hypermnesia effect, which predicts a better recall of pictures over time". Chun and Plass speculated that unlike words that tend be lost over time; pictures allow learners to form a mental model of the presented information resulting in a long-term learning of the presented information. Thus, it can be argued that since students in the VA condition had already formed and stored mental images of the target words, it is probable that the mental images of the target words were activated and retrieved during the delayed post-tests, which thereby, led to the better retention of words.

In general terms, the analyses of the third research questions demonstrated that the four audio-visual groups lost considerable amounts of their initial vocabulary learning gains after one month, regardless of the modality of the audio-visual input. However, the different modalities of audio-visual input had a differential effect on the retention on the three vocabulary knowledge aspects. The attrition rates in meaning recall and recognition knowledge dimensions of the VA condition were significantly smaller than those of the other experimental conditions, VAC, CA, and A only.

6.4 Section four: research question 4

R – Q 4: To what extent does the frequency of occurrence affect L2 incidental vocabulary learning from the four experimental conditions?

This part discusses the key results of the fourth research question pertaining to the effects of frequency of repetitions on incidental vocabulary learning of spoken form recognition, meaning recall, and meaning recognition across the four experimental groups (VAC, VA, CA, and A only). It is recalled that the 36 target words were appointed to three frequency groups based on their numbers of repetitions in the video clips: frequency group 1 (2-4 repetitions) frequency group 2 (5-7 repetitions) frequency group 3 (8+ repetitions).

The data analyses of the effect of frequency of occurrence on incidental vocabulary learning revealed several interesting results. Firstly, the results demonstrated that there was a general effect of frequency of occurrence on incidental learning of the three vocabulary knowledge dimensions, spoken form recognition, meaning recall, and meaning recognition. That is to say, as frequency of encounters increases, the likelihood of incidental vocabulary learning increases as well. These outcomes reinforce the findings of numerous earlier studies which confirmed that frequency of repetition plays a key role in facilitating incidental vocabulary learning (Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010; Peters et al., 2016; Peters & Webb, 2018; Vidal, 2011; Waring & Takaki, 2003; Webb, 2007, 2008). The vast majority of evidence establishing the relationship between incidental vocabulary learning and frequency of occurrence comes from reading studies. To the best of the researcher's knowledge, only three audio-visual studies have looked at the effect of frequency of occurrence (Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013). The study by Rodgers (2013) showed a medium-sized correlation between frequency of occurrence and incidental vocabulary learning in one of the two vocabulary tests used (i.e., the tough multiple-choice test). Similarly, Peters and Webb (2018) detected a positive relationship between the numbers of encounters and incidental vocabulary learning of meaning recall and meaning recognition. However, direct comparisons of the results of the present study with those of the two audio-visual studies are difficult, since they were designed and executed differently.

Despite observing a general effect of frequency on incidental vocabulary learning, the present findings indicated that the effect of exposure frequency to unknown words was not straightforward, as it varied across the three different word knowledge aspects, spoken form recognition, meaning recall, and meaning recognition. In particular, the results revealed that there was no consistent pattern for the effect of repeated encounters on the learning of the three vocabulary knowledge types. Concerning the spoken form, descriptive statistics (see Table 4.23) showed that the scores of participants across the four audio-visual groups noticeably increased from the first frequency group (2-4) to the second frequency group (5-7), yet moving to 8+ occurrences, the effect of frequency started to diminish. This suggests that the learning of spoken form requires few exposures to the target words, precisely between two and seven occurrences, however after eight or more repetitions, repetitions do not continue to facilitate the acquisition of form. This finding is closely consistent with the results of Van Zeeland and Schmitt's (2013) listening study, which showed that the acquisition of forms requires between 3 and 7 repetitions.

The case of receptive and productive knowledge of meaning (meaning recall and meaning recognition) was, however, different. The findings of the study (reported in Table 4.25 and Table 4.27) revealed a clearly gradual increase in the learning gains along the frequency scale. That is, receptive and productive knowledge of meaning continued to develop each time the frequency of occurrence increased, regardless of input conditions. However, the real increase in learning gains of these two knowledge aspects started with 5-7 encounters, and accelerated with eight and more encounters. This implies that learners need a greater number of encounters with target words in order to acquire word meanings receptively and productively. These findings seem to add to the growing body of evidence indicating that 8+ repetitions are necessary for incidental learning to take place (Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003).

These conflicting outcomes regarding the differential effects of repetitions on the learning of the different word knowledge facets can be taken to suggest that the learning of some knowledge aspects, such as spoken form recognition, needs fewer repetitions than the other two aspects (i.e. meaning recall and recognition). These findings are congruent with the findings of some earlier studies (e.g., Pellicer-Sánchez & Schmitt, 2010; Van Zeeland & Schmitt, 2013) which suggested that the acquisition of different word knowledge dimensions requires different amounts of encounters. For example, Van Zeeland and Schmitt (2013) found that the learning of form and meaning required different numbers of repetitions. In particular, they indicated that the effect of repetition on form was more evident than on meaning.

There are several plausible explanations for the inconsistent results reported in the current research regarding the effect of repetition on the learning of the three aspects of word knowledge. Firstly, Zahar et al. (2001) noted that some aspects of word knowledge may be facilitated with one repetition, while other aspects, such as associations, collocations, and meanings may require a greater number of encounters in order to develop. The bigger impact of frequency of occurrence on meaning than on form may indicate that incidental learning of meaning is more difficult than learning of form. Another possible interpretation is the context in which the target words occurred. Pellicer-Sánchez and Schmitt (2010) argued that the context of a word may mediate the effects of the frequency of occurrence on learning of unknown word meanings. Webb (2008) also elucidated that the context of unknown words may make target word meanings either transparent or opaque, which thus results in a variation in the number of repetitions needed to learn word meanings. There could be, however, other possible factors that moderated the effects of frequency of repetitions on the learning of word forms. Webb and Chang (2015) indicated that different factors like including learners' proficiency, background knowledge, and illustrations in texts make the relationship between frequency of encounters and incidental vocabulary learning complicated. In his study, Vidal (2011) found factors such as word elaboration in the text (e.g., definitions, descriptions, synonyms, and paraphrasing) and the type of word were more important than repetitions in enhancing incidental vocabulary learning. The current study, however, could not control for such factors that may have had an impact on the learning of the different word knowledge types, as it used authentic materials, which were difficult to manipulate.

The most compelling finding to emerge from the analysis of the fourth research question was that the effect of frequency of occurrence seems to be modulated by the type of input condition, that is, the effects of frequency was not constant across the four audio-visual groups. On the spoken form recognition test, significant differences in gains between the frequency groups were found for learners in the VA and A only groups. While in the VA group there was a significant increase only in scores from the first frequency group ($M = 3.81$) to the second frequency group ($M = 5.30$), in the A only group the differences in gains ($M = 1.44$; $M = 3.88$; $M = 3.72$) between the three frequency groups (2-4; 5-7; 8+, respectively) were statistically significant. However, differences in vocabulary gains from the three frequency groups were not statistically significantly different for learners in the VAC and CA groups. These findings could be taken to imply that the effect of repetition of the unknown words may be moderated by the modality of the input condition. These findings partially reflect those of the comparative study by Vidal (2011) who found that the effect of frequency of repetitions varied across different input modes. That is, repetitions influenced incidental vocabulary learning of participants in the reading mode more than those in the listening mode.

The reason for the lack of effects of frequency of occurrence on the performance of students in the VAC and CA conditions in the present study might be attributed to the provision of captions in these conditions, which may have drawn students' attention to the forms of the unknown words.

Thus, they might have acquired word forms easily, without the need for multiple exposures. In other words, it is likely that the captions in the VAC and CA input conditions allowed learners to easily discern phonological forms of the unfamiliar words from the first few encounters. Danan (2004) indicated that captions helped his participants to differentiate easily between known and unknown words, which in turn may have increased the likelihood of acquiring the forms of the words. Since the VA and A only input modes did not involve captions, it is likely that students in these groups needed several meetings with the target words to notice the forms in order to acquire them, which may explain why they needed more encounters with the target words than their counterparts in the VAC and CA conditions.

Moreover, in the spoken meaning recall test and spoken meaning recognition test, there was a significant increase in vocabulary gains from the first frequency band (2-4) to the third frequency band (8+) for all experimental groups, meaning that participants' mean scores of the third frequency group were significantly higher than those of the first frequency group. However, in the VA and A only groups, the differences in gains between the first and second frequency bands were also significant. It appears that learners in the VA and A only conditions were more sensitive to the effect of repetitions than their counterparts in the VAC and CA conditions. In other words, learners in the captioning groups (VAC and CA) seem unable to take as much advantage of repeated encounters with the target words.

Generally, the present findings supported the claim that frequency of occurrence is one of the factors that affects incidental vocabulary learning, but the more complex task is to define a set number of repetitions necessary to ensure incidental vocabulary learning of different vocabulary knowledge aspects across the different input modes. This is because there are several factors that may lessen the effects of repetitions (Nation, 2001).

6.5 Section five: research question 5

R- Q 5: Is there a relationship between WM and incidental vocabulary short-term learning and long-term retention? If so, which subset(s) of WM, phonological loop or visuospatial sketchpad contribute(s) to this relationship?

The second half of this PhD thesis explored the relationship between WM and incidental vocabulary learning and retention from the four audio-visual conditions, VAC, VA, CA, and A only. This section presents the discussion of the fifth research question, which explored the relationship between WM capacity and incidental vocabulary learning and retention. The purpose was to determine whether WM has a link with incidental vocabulary short-term learning and long-term retention and if so, which component(s) (phonological loop or visuospatial sketchpad of WM) contribute(s) to the relationship.

Results of the analysis of this research question revealed that there was direct evidence for the relationship between WM and incidental vocabulary short-term learning. Most importantly, it was found that the participants' scores on the spoken meaning recall and recognition immediate post-tests correlated significantly with their scores on the verbal WM measures. These results provide experimental support for the large body of evidence that suggests that WM is one of the most significant predictors of vocabulary learning (Gathercole & Baddeley, 1989; Gathercole & Baddeley, 1990; Gathercole, Service, Hitch, Adams, & Martin, 1999; Gupta, 2003; Malone, 2018; Martin & Ellis, 2012; Masoura & Gathercole, 1999, 2005; Speciale et al., 2004; Verhagen & Leseman, 2016). However, the study failed to establish a link between WM and incidental long-term vocabulary retention. The correlation between the three vocabulary delayed post-tests and WM scores was not statistically significant. This result is contrary to the results of some previous studies, which have suggested a direct link between WM abilities and long-term retention of vocabulary items (Gathercole & Baddeley, 1990; Vallar & Baddeley, 1984). Taken together, the present findings may assume that WM is crucially involved in incidental short-term vocabulary learning, but not in long-term retention, regardless of the type of input conditions. These findings can be explained by the fact that vocabulary learning is a cognitive ability underpinned by certain cognitive mechanisms, including WM capacity. WM thus provides a resource that allows learners to process the incoming inputs (vocabulary items) and integrate that with relevant knowledge retrieved from long-term memory (Gathercole & Baddeley, 1990). Furthermore, Gathercole, Alloway, Willis, and Adams (2006, p. 277) argued that WM "acts as a bottleneck for learning in many of the individual learning episodes required to increment the acquisition of knowledge".

The present findings also showed that the verbal WM component (the phonological loop) was the system that significantly contributed to the attested relationship between WM and incidental vocabulary short-term learning. In particular, the correlational analysis revealed that the scores on the verbal WM measures, forward digit recall test and backward digit recall test correlated significantly with the scores on the vocabulary spoken meaning recall immediate post-tests (r_s = .593, $p < .000$; r_s = .390, $p < .021$), respectively. The results also showed that there was a significant correlation between the forward digit recall WM test and the spoken meaning recognition vocabulary immediate post-test, r_s = .433, $p <.001$. The results, however, found non-significant associations between the scores on the visuospatial WM measures (dot matrix and the odd-one-out) and the three vocabulary immediate and delayed post-tests. These findings replicate those of Malone (2018) which demonstrated a strong link between the scores of learners on the verbal WM tasks, nonword repetition and operation span and incidental vocabulary learning, but no significant correlation between incidental vocabulary learning and the scores on the visual WM task (Shapebuilder).

Taken together, the present results of the differential role of the verbal/phonological WM system and visual WM system in incidental vocabulary learning, which lend further weight to the perspective of the domain-specific of WM systems, are readily accommodated by the theoretical accounts of the domain-specific WM model of Baddeley and Hitch (1974). Baddeley and Hitch postulated the existence of two different cognitive systems (the phonological loop for verbal information and visuospatial for visuospatial information), each of them is domain-specific, governed by a domain-general system (the central executive). In the same vein, Alloway et al. (2006) contended that different performance on verbal and visual WM tasks is underpinned by different WM systems, so that explains why there were differential correlations between the two the WM systems (the phonological loop and visuospatial sketchpad) and vocabulary learning. Put it simply, the findings may suggest that the verbal WM subsystem, phonological loop is significantly involved in novel L2 vocabulary learning. In addition, the possible explanation for why the visuospatial WM system did not exhibit a significant relationship with the composite vocabulary scores may be related to the prediction that the visuospatial WM component plays a marginal role in vocabulary learning, as hypothesized by Baddeley (2003). Baddeley argued that the relevance of the visuospatial WM system to language learning is far less clear than the verbal WM system.

In the present study, the unique contribution of the verbal WM system found in vocabulary learning is explained in light of Baddeley's (2003) theoretical argument which states that the phonological loop has conceivably evolved to support the acquisition of language, in general and vocabulary, in particular. The reason why the verbal WM system is important for vocabulary learning is attributed to its significant role in establishing long-term mental representations of the newly presented vocabulary items. Gathercole and Baddeley (1990) explained that the process of acquiring new vocabulary items involves two main steps: firstly, establishing a stable long-term representation of the newly presented sequence of sounds or underlying concept, and secondly, linking such sounds with relevant existing representations. Achieving a stable long-term representation requires a temporary representation that is constructed by the verbal WM system (Gathercole, 2006). By this logic, the larger the verbal WM capacity, the better the temporary representation and thus the better the learning and vice versa. In addition, Alloway et al. (2014) argued that the underlying cognitive mechanism, the phonological loop, underpins both vocabulary learning and performance on verbal WM tasks, such as digit recall and nonword repetition. On the basis of this theoretical analysis, it is possible to assume that the learners' ability to repeat a sequence of digits in the forward digit and backward digit tests and to learn new vocabulary items are constrained/supported by their phonological loop capacity.

The significant finding regarding the effects of WM on L2 vocabulary learning lends strong support to Wen's (2016b, 2019) essential constituents of his WM model. Wen posits that one of the signature features of WM is the limited capacity of its different components. The limited capacity of the two systems (the phonological WM and the executive WM) manifests itself with regards to the limited amount of information that can be maintained and processed during the process of L2 learning. More relevantly to vocabulary learning in the present study, it is possible that the limitation of the learners' WM capacity has constrained their acquisition and development of the target vocabulary items. Learners with larger WM capacity were better able to acquire the target lexical items. In other words, due to the learners' WM capacity limitation, it was difficult to maintain and process more than a limited amount of information (vocabulary words) at a given time.

In particular, the role of the phonological WM system in the acquisition of vocabulary lends support to the theoretical conceptualizations of the WM construct postulated by Wen (2016b, 2019). Specifically, Wen posits that the phonological WM system plays an instrumental role in L2 lexis acquisition and it is a driving force behind L2 vocabulary acquisition. That is, during vocabulary learning, L2 learners rely on their phonological WM system to maintain the vocabulary items. According to Wen (2018, p. 35), the obvious explanation for this significant relationship between the phonological WM system and vocabulary learning is related to the fact that the mechanisms (i.e. the phonological short-term store and the articulatory rehearsal) associated with the phonological WM system "play a critical role in the *chunking* procedure and subsequently the *consolidating* process of linguistic sequences into long-term knowledge".

The present findings which suggest that the different components of WM have distinctive effects on different SLA tasks and domains, corroborate Wen's theoretical argument regarding the distinctive cognitive systems and functions of WM. Wen posited that WM consists of different cognitive components (i.e. phonological WM, executive WM) that have separate functions and roles in SLA. From the Wen's (2016) framework of the integrated model of WM, the differential contribution of the simple verbal forward digit recall task and complex verbal backward digit recall task recorded in this study can be explained as suggesting differential importance for the different systems of WM. Particularly, the responsibility of the verbal/phonological WM system (e.g., phonological short-term store and articulatory rehearsal) is potentially more important to vocabulary learning than the responsibility of the other WM system, the executive WM system. Unlike the phonological WM system which is a modality-specific system (i.e., verbal-based information), the executive WN system is a domain-general system that is responsible for updating, task switching, and inhibitory control. Wen (2019) argues that in vocabulary learning both WM systems are involved differently. The phonological WM component is involved in coding and consolidating the phonological forms of the target lexical items into long-term representations, whilst, the executive WM system plays a key role in coordinating attentional resources for interpreting the semantic characteristics of these newly acquired words.

The present findings of the significant association between the verbal/phonological WM system and vocabulary learning are broadly consistent with a multitude of earlier studies that established a close relationship between vocabulary learning processes and the verbal/phonological WM abilities in L1 and L2 domains. For example, Engle de Abreu et al. (2012) established a clear and reliable correlation between the children's verbal WM abilities, measured by both simple digit span and nonword repetition tasks, and vocabulary learning. A further study on bilingual and monolingual children by Verhagen and Leseman (2016) recorded a direct relationship between the verbal WM system and word acquisition. The researchers also found that simple and complex memory measures had differential associations with vocabulary learning. Taken together, these positive associations reported in previous studies and in the present study add further support to the theoretical proposal of Baddeley and Hitch's (1974) multicomponent model of WM and Wen's (2016b) integrated model of WM that considers the verbal/phonological WM component as a 'language learning device'.

To conclude, the study established a close relationship between WM and incidental vocabulary short-term learning, suggesting a pivotal importance for the verbal/phonological WM system in L2 vocabulary acquisition. The verbal WM subsystem (measured by a forward digit test and a backward digit test) made significantly contributed this relationship. Therefore, it is safe to conclude that this WM system has a fundamental role to play in L2 vocabulary acquisition

6.6 Section six: research question 6

R – Q 6: Which of the WM task types, simple or complex, best predicts incidental vocabulary short-term learning and long-term retention?

This section presents the discussion of the findings of the sixth research question, which was concerned with the type of WM tasks, simple or complex that best predicted incidental vocabulary learning and retention. The purpose was to determine whether the simple and complex WM tasks played independent roles in incidental vocabulary learning and retention (represented by the composite vocabulary scores). As a reminder, the study ran two simple memory tests, forward digit recall and dot matrix and two complex memory tests, backward digit recall and the odd one out.

Of great interest is the intriguing finding revealed by the correlational analysis that indicated the composite vocabulary scores for incidental vocabulary short-term learning were more strongly related to one particular verbal WM measure (i.e., simple forward digit) than the other verbal WM measure (i.e., complex backward digit) and both the non-verbal WM measures (the odd one out and dot matrix). To put it differently, the simple and complex WM measures played independent roles in incidental vocabulary learning. Whilst, performance on the verbal simple WM measure (forward digit recall) showed strong associations with incidental vocabulary learning, the correlation between the verbal complex backward digit recall task and incidental word learning was moderate. The regression analysis further demonstrated that the verbal simple forward digit recall test was the only strong predictor of the composite scores for the three vocabulary immediate post-tests. These results indicate that the simple forward digit recall task, which is associated with a domain-specific system (the phonological loop) in Baddeley's (1986) WM model, was a stronger predictor of incidental vocabulary learning than was the complex backward digit recall task. These findings build upon previous evidence of the relationship between simple span tasks such as forward digit and nonword repetition and vocabulary learning (e.g., Gathercole et al., 1999; Martin & Ellis, 2012).

The theoretical basis of the test battery adopted in this study (Alloway, 2007) is Baddeley's (1986) multicomponent model of WM. According to the author of test battery, simple and complex span tasks are distinguished from each other based on the demands imposed by the task. Simple tasks (e.g., forward digit recall) engage storage capacity of a domain-specific component of WM (the phonological loop). In this type of memory task, a test-taker has to recall a sequence or string of spoken digits in the correct order of presentation. In contrast, complex span tasks (for example, backward digit recall), in which the subject is required to report a series of digits in the reverse order of the presentation, involve both storage and processing aspects of the WM system (Alloway et al., 2010). Thus, it can be argued that the different and independent roles played by the simple and complex WM measures found in the present study are due to the different cognitive activities the task imposes.

Another possible alternative explanation for the differential correlations between the simple and complex WM tasks appears to be attributable to the perspective that some WM tasks are more predictive than others in certain cognitive tasks. A range of studies has reported that complex span tasks are more closely linked with complex cognitive activities than simple span tasks. For example, Conway et al. (2005) reported that tasks that impose storage and processing demands (complex spans) are better predictors than those that require storage activity only (simple tasks) of some complex cognitive behaviour, such as reading comprehension, reasoning, and problem solving. Similar results of Daneman and Carpenter (1980) indicated that the complex span task, reading span was a stronger predictor of reading comprehension than the simple span tasks, as the simple tasks only assess the storage aspect of the phonological loop.

The strong correlation found in this study between the simple forward digit recall task with incidental vocabulary learning may be related to the fact that simple verbal tasks such as forward digit and nonword recall are more closely linked to vocabulary learning, as evidenced in many studies. The present findings echo those of Engle de Abreu et al. (2012) who indicated that the verbal simple memory tasks (forward digit recall and nonword repetition) were strongly related to vocabulary and exhibited a weaker link with reading, while complex memory tasks (counting recall and backward digit) correlated closely with syntax and reading. Likewise, Martin and Ellis (2012) found that the simple verbal nonword repetition and nonword recognition tasks associated more closely with vocabulary measures than complex verbal tasks (i.e. listening span) did.

6.7 Section seven: research question 7

R – Q 7: Does WM mediate the effect of the four modalities of audio-visual input on incidental vocabulary short-term learning and long-term retention?

This last section is dedicated to the discussion of the last research question which examined whether WM has a role to play in moderating the effects of different modalities of audio-visual input on incidental vocabulary learning and retention (represented by global scores for the three vocabulary immediate and delayed post-tests). Specifically, it looked at whether the role of WM in incidental vocabulary learning interacted with the four input modes (VAC, VA, CA, and A only). As has been highlighted before, the continuous variable (i.e., scores on the four WM tasks) was converted into a dichotomous variable (high span vs low span) by splitting the quantitative scale at one point and assigning learners above that point into a 'high' span group and learners below that point into a 'low' span group. Therefore, across the four experimental groups, participants' raw scores on each of the four WM tests, forward digit recall, backward digit recall, dot matrix, and the odd one out were classified as high or low span groups based on the median split.

The most striking findings that have emerged was that concerning incidental vocabulary short-term learning (represented by composite scores for the three vocabulary immediate post-tests) across the four audio-visual input conditions, there was no significant impact of individual differences in verbal WM capacity (measured by the forward digit recall and backward digit recall tests) and visual WM capacity (measured by the dot matrix and odd one out tests). A similar pattern was found for the long-term retention (represented by the composite score for the three delayed post-tests). That is, composite vocabulary scores for the immediate and delayed post-tests of high and low WM span groups were not significantly different across the four treatment groups. The findings of the one-way ANOVA tests showed that subjects with high WM visual and verbal WM capacities did not perform significantly better than their counterparts with low verbal and visual spans on the vocabulary tests.

In addition to this finding, the two-way interaction between WM capacity and the four treatment conditions was not significant, suggesting that WM capacities did not influence the effects of the different modalities of audio-visual input on incidental vocabulary learning and retention. What is curious about these results is that it was expected that learners with high verbal and visual WM capacities would benefit more from the combination of visual and verbal information in the different input conditions than learners with low WM capacities, given the predictive role of WM in learning through multiple input conditions. Literature concerning individual variations in WM capacities states that individuals with high WM levels are adept at a wide range of learning tasks (Gyselinck et al., 2008). It has been stated that learning from different input modes is a higher-order cognitive process that draws upon several cognitive sub-processes (such as, WM) to be successful (Mayer, 2014). The findings of the present study, however, seem to indicate that WM does not moderate the effects of multiple modalities of audio-visual input.

These surprising and counterintuitive outcomes run counter to a growing body of evidence indicating a facilitative effect of WM capacity on input modality (Gyselinck et al., 2008; Lusk et al., 2009; Malone, 2018). These studies confirmed that individual differences in WM capacity, which are theoretically related to the capacity to simultaneously maintain and process verbal and visual information, could affect performance on a variety of learning tasks through different input modes. Malone (2018), for instance, found a differential role of WM in incidental vocabulary learning based on input modality. The effects of WM on incidental vocabulary learning was more evident in the dual presentation modes (listening while reading) than the single presentation mode (reading) and the relationship between WM and single input mode was weaker than that of the dual presentation modes. Malone (2018) argued that the dual modality of input increased the memory loads and the inclusion of aural (listening) mode to the visual (reading) placed greater demands on participants' WM. However, since Malone's study did not have a listening-only group, it is insensible to claim that the demand placed on WM was just due to the inclusion of the audio.

One way to interpret the present surprising findings is that it is likely that the effects of multiple input modalities may have overridden the role of WM in vocabulary learning through multimodal input conditions. Input modality has been found to influence incidental vocabulary learning as highlighted in the discussion of the vocabulary results (see section 4.2). Presenting information in visual and verbal modes may have eliminated the effects of WM capacity as multiple input modalities might benefit low WM capacity individuals to the point where high and low WM capacity individuals learn the same amount of information from the multimodal conditions. In other words, it seems plausible to suggest that presenting information in different modalities (dynamic visual, on-screen text, and verbal, as is the case in the present study) has a positive impact on incidental vocabulary learning regardless of the learners' WM capacity. It can be thus deduced that multiple input modalities could be effective for both low and high WM spans. In this case, mean differences in vocabulary tests of the four experimental groups could be safely attributed to the type of input modality rather than to individual variations in WM spans. This explanation echoes the findings of Bayraktar and Altun (2014) who reported that their participants' performance was affected positively by the multimedia design types/modes irrespective of their WM capacities.

Another possible reason for these nonsignificant findings could be due to the division of the participants into high and low spans based on the median split. There is a concern that it is likely that the cut-off point to differentiate low and high spans was not well-defined. In particular, it is possible that dividing participants into high and low spans may have underestimated the individual variations in WM capacity, that is, students with scores close to the cut-off point but on the opposite side are regarded as very distinct rather than very similar. As an example, in the forward digit recall test, participants who were above the median (which was 92) were appointed to the 'high' span group even if they were only one point above the median (e.g., 93). Similarly, participants with scores slightly below the median (e.g., 91) were categorized into the 'low' span group. This division likely altered the true nature of individual differences in WM capacity, as the participant who had a score of 93 would be considered distinct from his peer who scored 91 and thus they would be assigned into two different span groups. Furthermore, it was observed that most of the participants' scores on the four WM tests were clustered around the median. A potentially problematic implication of underestimating the actual individual variations in WM is that it may increase the possibility of type II errors by reducing experimental power (McClelland, Lynch Jr, Irwin, Spiller, & Fitzsimons, 2015). One way to lessen the risk of reducing statistical power is by recruiting a greater number of participants (Iacobucci, Posavac, Kardes, Schneider, & Popovich, 2015). However, given the relatively small number of subjects (n = 12 in each treatment group) who completed the four WM tests in this study; this may have prevented the statistical significance from surfacing.

Overall, the findings of this research question suggest individual differences in WM capacity, measured by the verbal and visual, did not account for the variations in the composite vocabulary scores in each of the four audio-visual input conditions. The non-significant interactions between WM and the input conditions indicate that WM did not mediate the effects of the different input modalities.

The following chapter presents the conclusions drawn from the aforementioned results.

Chapter 7 Conclusion

This concluding chapter provides a brief summary of the present study by firstly providing a summary of the key findings described in chapters four and five, before highlighting the original contributions the study made to knowledge in the field of L2 vocabulary learning. The chapter then considers the implications of the research findings. Finally, it discusses the limitations of the study and offers some suggestions for future research to help further extend the knowledge gained.

7.1 Summary of the findings

Chapters four and five revealed several interesting results with regards to the effects of the four audio-visual input modalities (VAC, VA, CA, and A only) and WM on L2 incidental vocabulary learning and retention. More specifically, Chapter four looked at vocabulary learning gains from the four input modalities and compared those gains across the four treatment groups. The chapter also presented the results of the analyses of the delayed post-tests before providing the findings of the role of frequency of occurrence in incidental vocabulary learning. Chapter five presented the WM data regarding its relationship with incidental vocabulary learning and retention and looked at the nature of the relationship whether it operated through verbal WM measures or through the nonverbal WM measures. The chapter also considered the role of WM in mediating the effects of input modalities on incidental vocabulary learning and retention.

7.1.1 The differential effects of the four audio-visual input conditions, VAC, VA, CA, and A only, on L2 incidental vocabulary learning of spoken form, meaning recall, and meaning recognition (research questions, 1 & 2)

The aim of the first research question was to examine whether the four audio-visual input conditions (VAC, VA, CA, and A only) contributed to incidental vocabulary learning and, if so, which input better enhanced the three vocabulary knowledge aspects. The findings showed that the four input conditions did promote L2 incidental vocabulary development of spoken form recognition, meaning recall, and meaning recognition. The data showed that participants' scores across the four treatment groups significantly increased between the pre-tests and the immediate post-tests. Particularly, after viewing/listening to four relatively short video clips, i.e. one hour, learners across the four groups were able to learn between 2 and 5 of each of the 9 tested words. These results stress the pedagogical value of the different modalities of audio-visual input in enhancing learners' vocabulary knowledge. The study also found that the effect of the four different modalities of input on the learning of receptive recognition knowledge aspects, spoken form and meaning recognition was greater than that on the learning of meaning recall. The learning of the three vocabulary knowledge facets was similar to the pattern found in several incidental vocabulary studies, i.e. form > meaning (Brown et al., 2008; Montero Perez et al., 2014; Montero Perez et al., 2018; Peters et al., 2016; Peters & Webb, 2018; Van Zeeland & Schmitt, 2013; Waring & Takaki, 2003). This suggests that the learning of receptive knowledge aspects is usually easier than the learning of productive knowledge aspect, regardless of input modality.

The analyses of the second research question demonstrated that the effectiveness of the input condition differed based on its modality type (VAC, VA, CA, A only) and the aspect of vocabulary knowledge assessed (i.e., spoken form recognition, meaning recall, meaning recognition). The three multimodal input groups, VAC, VA, and CA significantly outperformed the single input group (A only) on the three vocabulary immediate post-tests. These results appear to confirm the positive effect of the multimodal presentation of information reported in the literature (Akbulut, 2007; Al Seghayer, 2001; Bird & Williams, 2002; Chun & Plass, 1996; Montero Perez et al., 2014; Montero Perez et al., 2018; Sydorenko, 2010). The findings also indicated that the CA input group significantly outperformed the other three groups on the spoken form recognition immediate post-test, followed by the VAC group. This may suggest that the captioning conditions (CA and VAC) were more effective for the learning of form than the non-captioning conditions (VA and A only). The results have also shown that the performance of students involved in the VA condition on the meaning recall and meaning recognition immediate post-tests was significantly superior to that of the other three groups (VAC, CA, and A only). These may imply that this audio-visual input condition (VA) has an advantage for the short-term acquisition of meaning recall and meaning recognition knowledge facets. These findings lend support to the advantage of 'dual coding' of information.

7.1.2 The differential effects of the four audio-visual input conditions, VAC, VA, CA, and A only, on incidental vocabulary long-term retention of spoken form, meaning recall, and meaning recognition (research question 3)

While the data of the first and second research questions have demonstrated that incidental vocabulary learning is possible through the four input modalities, the outcomes of the third research question showed that the gains were prone to a large decay over time. That is, it was indicated that learners across the four treatment groups were unable to remember large amounts of the initial learning gains one-month after the end of the research intervention. Precisely, more than half of the initial learning gains were lost after one month, irrespective of the input modality. In addition, the analysis of this research question showed that the four input modalities had differential effects on the long-term retention of the three vocabulary knowledge aspects. The largest attrition rate found in the spoken form recognition delayed post-tests was for learners in the CA condition. The attrition rates of learners in the VA condition of the meaning recall and meaning recognition were significantly smaller than those of the other experimental groups. These findings seem to suggest that the long-term effect of the four input modalities varies across the three vocabulary knowledge types.

7.1.3 The effect of frequency of occurrence on incidental vocabulary learning (research question 4)

The effect of frequency of word encounters on incidental vocabulary learning of spoken form recognition, meaning recall, and meaning recognition has been established in the present study, suggesting that the more a word is encountered in the input, the higher the chances of learning that word. These findings stress the importance of repetition of the target words in enhancing incidental vocabulary learning. However, the effect of frequent encounters varied across the three vocabulary knowledge aspects. That is, the effect of repetition on the learning of meaning at the two mastery levels (receptive and productive) was bigger than on the learning of form. The findings also showed that the effect of word repetitions was modulated by the input modality. In the spoken form recognition test, significant differences in gains from the first and second frequency groups (2-4; 5-7) were found for learners in the VA and A only groups, but not for learners in the VAC and CA groups. In the meaning recall and recognition tests, significant differences in learning gains from the first and third frequency bands (2-4; 8+) were found for the four experimental groups.

7.1.4 The relationship between WM and incidental vocabulary learning and retention and the interaction between WM and the four input modalities (research questions, 5-7)

The second part of this thesis was designed to explore the relationship between WM and incidental vocabulary short-term learning and long-term retention and to examine the interaction between WM and the four input modalities. The results revealed a direct relationship between WM and incidental vocabulary learning, irrespective of the input modalities. WM was found to be an important predictor of incidental vocabulary learning, as it accounted for 41.9% of the unique variance in incidental vocabulary learning. However, no relationship was found between vocabulary long-term retention and WM. The phonological loop was the WM component that contributed significantly to the relationship between WM and incidental vocabulary learning. In particular, the results showed that scores on the verbal WM tests (forward digit recall and backward digit recall) and the composite vocabulary scores for the three immediate post-tests correlated significantly. In addition, the results of the multiple regression also showed that the verbal simple forward digit recall measure was a stronger predictor of incidental vocabulary learning than the verbal complex backward digit recall measure. Concerning the role of WM in mediating the effects of input modalities, it was found that the influence of individual differences in WM capacity on incidental vocabulary learning and retention from the four input modalities was nonsignificant. Vocabulary scores of learners with high and low WM spans measured by forward digit recall, backward digit recall, dot matrix, and the odd-one-out were not significantly different. In addition, WM was not found to mediate the effects of the different input modalities on incidental vocabulary learning and retention.

7.2 Overall contribution of the study

The literature review in Chapter two pointed to a range of issues that needed to be examined to better understand L2 incidental vocabulary learning and retention through different modalities of audio-visual input. This research study is significant and unique, as it has provided crucial insights to incidental word learning and retention. It offered several valuable contributions to the field of L2 vocabulary learning by examining the effects of various input modalities (VAC, VA, CA, and A only) on L2 incidental vocabulary short-term learning and long-term retention as well as by looking at the role of an item-related factor (i.e., frequency of occurrence) and a learner-related factor (i.e. WM). In particular, despite the limitations of this study, it managed to fill the gaps in current knowledge in the following ways. Firstly, the bulk of past audio-visual studies have focused on the differential effects of L1 and L2 subtitles and captions and there was a neglect of the role of imagery in those audio-visual studies. The work presented in this thesis sought to make a valuable contribution to current knowledge by examining the effects of various input modalities (VAC, VA, CA, and A only) on L2 incidental vocabulary short-term learning and long-term retention. The four audio-visual treatment conditions were carefully designed to tease apart the differential effects of different combinations of input modalities on incidental vocabulary learning and retention. More importantly, through these experimental conditions I was able to find out which of the modalities of audio-visual input were useful for facilitating the three different aspects of word knowledge tested in this study.

Also, there was little attention paid to the effect of frequency of occurrence of the target words on short-term word learning in past audio-visual studies. Since, frequency of occurrence (an item-related factor) plays a key role in facilitating incidental word acquisition, this study thus contributed to the field of incidental vocabulary learning by looking at the role played by the frequency of occurrence of the target words in incidental vocabulary learning. Specifically, it attempted to determine the number of repetitions necessary for different aspects of word knowledge through the different audio-visual input conditions.

In addition, most of the former studies have used different types of pre-and-post-tests. That is, those studies pretested the participants' prior knowledge of the target words using test types that were different from the immediate post-test types. Since, using different types of vocabulary tests across the different phases of the study may render the results invalid, the present study filled a knowledge gap by using arguably better-suited vocabulary measurements with regards to applying similar vocabulary acquisition measures across the different phases of the study (pre-tests; immediate post-tests; and delayed post-tests). A further gap in our knowledge is the paucity of studies on long-term effects of the different modalities of audio-visual input. Most of previous audio-visual studies have focused on short-term effects of the different modalities of audio-visual materials measured by immediate post-tests. However, those studies did not administer delayed post-tests to gauge the long-term effects of the different modalities. The present study therefore added to knowledge by examining the long-term effects of the different modalities of audio-visual input. This was done by administering delayed post-tests one month after the end of the four treatment sessions.

Finally, and most importantly, to the best of the researcher's knowledge, there has been no research examining the role of WM (an individual-related factor) in incidental vocabulary learning and retention through multiple modalities of audio-visual input. Exploring the relationship of WM with incidental vocabulary learning and retention through audio-visual materials will add invaluable knowledge to the field of L2 vocabulary acquisition, given the key role of WM in vocabulary learning. This study thus explored the relationship between WM and incidental vocabulary learning and retention, more specifically, the different roles played by the verbal and visuospatial WM systems in incidental vocabulary learning and retention. It also investigated the role played by WM in moderating the effects of the different audio-visual conditions on incidental vocabulary learning and retention.

7.3 Implications

7.3.1 Theoretical implications

The current study was situated within Paivio's (1986) DCT and Mayer's (2005; 2014) CTML. The findings were by and large, adequately explainable through these two theoretical accounts. Importantly, the findings showed how different theories (i.e., DCT and CTML) could complement one another to get an in-depth understanding of the phenomenon under investigation (i.e., incidental vocabulary learning and retention through different input modalities). For example, the study found that learners involved in the multimodal input conditions, VAC, VA, and CA were able to increase their vocabulary knowledge of spoken form recognition, meaning recall, and meaning recognition better than their counterparts in the A only single input condition. These results imply that the combination of multimodal presentation of information had a positive effect on incidental word learning. The theoretical views of Paivio's DCT and the multimedia principle of Mayer (2005; 2014) which postulate that recollection and learning of information is enhanced when information is presented visually and verbally, explain the significant superiority of the input conditions (VAC, VA, and CA) over the single presentation mode (A only condition) on incidental vocabulary learning. In other words, both the DCT and the multimedia principle assume that the use of the verbal system and the imagery system to encode the information received through multimodal input conditions, contributes to recall and learning which is superior to coding in only one modality.

Furthermore, the study provides evidence for Mayer (2005) 'redundancy principle'. The findings demonstrated that the multimodal input condition (VAC) was not as effective as the VA condition for promoting the learning of meaning. The findings, which indicated that the combination of videos, audios, and captions in the VAC condition negatively affected the learning of word meaning, as captions distracted learners from paying attention to the images of the videos and the audios, provided support for the redundancy principle regarding the deleterious effects of simultaneous presentation of the three presentation modes of video, audio and caption (VAC) on learning and recall of the presented information.

In addition, the findings of this study that pointed to the differential associations between the two WM components, the phonological loop and visuospatial sketchpad and incidental vocabulary learning are accommodated by the theoretical views of the multicomponent WM model of Baddeley and Hitch (1974). Baddeley and Hitch posited that WM is composed of two domain-specific systems, the phonological loop and visuospatial sketchpad and a domain-general system, the central executive. The findings provided evidence for the distinction between the verbal and visual WM systems in vocabulary learning.

7.3.2 Pedagogical implications

Besides the main contributions of the empirical findings to knowledge, the present study identified a multitude of issues related to the effects of different input modalities on L2 incidental vocabulary learning and retention, which language teachers and learners should be aware of. Therefore, it puts forward a number of pedagogical implications.

Firstly, the results of this study, which pointed out the potential of audio-visual input conditions as valuable sources of L2 incidental vocabulary learning, will stimulate changes in vocabulary teaching. Since it is impossible to cover the tremendous numbers of words that L2 learners need to master in order to use English proficiently through direct teaching methods alone, teachers are thus encouraged to develop such input modalities and incorporate them into their language materials to foster incidental vocabulary learning. If the proposed treatment sessions (15 minutes each) were employed in 3 of the 5 teaching sessions per week for 15 weeks per academic term (as in the case of the EFL context that hosted the present study), students are likely to be able to incidentally acquire between 90 and 225 lexical items from these interventions only.

In addition, the findings of this study have proven viewing authentic videos to be useful in facilitating incidental vocabulary learning. Authentic videos, which are easily available with a wide variety of topics, may be a particularly valuable source of input for learners in most EFL environments, where there is limited exposure to the target language outside the classroom. It is thus suggested that teachers develop students' awareness of the importance of viewing authentic videos for improving their vocabulary. To best optimize the benefits of authentic videos for promoting vocabulary learning, teachers need to consider the following issues. Firstly, the chosen videos should appeal to the students' interests; the more interesting the videos are to the learners, the better chances to fuel vocabulary acquisition (Rodgers, 2013). The selected videos also need to be suitable for the proficiency level of the learners. Previous research found that if the materials best match the learners' proficiency level, learners are likely to pick up larger vocabulary gains (van Zeeland & Schmitt, 2012). In addition, videos featuring images that clearly portray the target words should be selected. Videos with content well-supported by visual images serve important purposes, such as helping learners to learn about the meanings of the unknown words and elucidating unfamiliar concepts (Mayer, 2014).

The results also showed that the availability of the on-screen captions in audio-visual input conditions might have a bearing on incidental learning of form. The most outstanding effect of the presence of captions was found on the recognition of new word forms. Consequently, the use of captions should be encouraged when the aim is to enhance the learning of form. The captions ought to be clear and at a pace that corresponds to the auditory inputs. Notwithstanding, in using captions with videos, as is the case in VAC condition, special attention should be paid to the potential distraction that the captions may cause.

Another implication concerns the natural fact of vocabulary attrition, the findings showed that the proposed values of the multiple input modalities in fostering incidental vocabulary short-term acquisition largely diminished at the delayed post-tests time point. Admittedly, a large amount of the learning gains that were lost could have been reduced if output tasks were given to the learners to create a durable memory for the acquired word knowledge. Therefore, teachers are advised to create subsequent teaching/learning exercises or output tasks to enable learners to consolidate the vocabulary items learned.

7.3.3 Implications for Saudi EFL context

One of the prime implications of the study is to inform vocabulary learning and teaching in Saudi Arabia. It will be recalled (section 1.2) that despite spending several years of learning English, Saudi EFL learners still show critical deficiency in vocabulary knowledge. This highlights the importance of exploring other potential sources for vocabulary uptake. A range of studies (e.g. Rodgers, 2013) indicated that viewing audio-visual materials regularly may facilitate vocabulary uptake. The findings of the present study supported previous studies findings regarding the positive effect of regular viewing of audio-visual materials on vocabulary knowledge. These findings might have applicability in an extensive viewing programme. Regular viewing forms the basis of extensive viewing. It is suggested that the extensive viewing programmes be introduced in intensive English language programmes in Saudi universities. In order to make the most of this programme, learners should be encouraged to choose and view the videos or episodes of TV that appeal to their interests. Since the extensive viewing is a non-existent approach in Saudi Arabia, a range of steps of preparation need to be considered in order to set up a successful programme. Firstly, teachers should be made aware of the importance of regular viewing of audio-visual materials for enhancing language skills in general, and vocabulary knowledge, in particular. Additionally, teachers of English language programmes at Saudi institutions are recommended to make a large and wide range of supply of audio-visual materials to meet differing tastes and interests. Also, it would be useful to develop learners' awareness that watching audio-visual materials does not only bring joy but also leads to increasing vocabulary knowledge. Furthermore, the findings regarding the effects of the presence of captions when watching videos may have a bearing on setting up an extensive viewing programme. It was found that captions distracted the learners from paying attention to the images of the videos or audio. Consequently, utilizing videos without captions would be useful for enhancing vocabulary knowledge.

In recent years, greater and easier accessibility to audio-visual materials have created opportunities for language learners and teachers to foster vocabulary learning outside classrooms as well. Since watching audio-visual materials (i.e., TV, video, film) is a popular out-of-school activity for young Saudi learners of English, language teachers in Saudi Arabia are encouraged to exploit this activity to foster vocabulary learning. The out-of-school exposure to authentic input can help compensate for the limited time dedicated to vocabulary instruction in Saudi EFL setting.

Besides their entertainment value, audio-visual materials could be a rich source of L2 vocabulary, as the findings of this study revealed. Therefore, a minimal time investment on such activity may help make it more productive. Teachers may encourage learners to view special audio-visual materials that are more conducive to vocabulary learning and train them to apply certain strategies to learn about new word meanings.

In addition to the above, the study provides some insights for language teachers in developing the most suitable combinations of modalities for enhancing different aspects of word knowledge. The study revealed that combining audio with text (as in CA condition) was useful for promoting form knowledge. The study also showed that combining picture and audio facilitates meaning. In developing audio-visual conditions, these results could be taken into consideration as they demonstrate the most effective combinations of audio-visual modalities for different word knowledge types. In developing vocabulary materials, teachers are recommended to consider individual differences in WM capacity, as the present findings provided evidence on the importance of WM in incidental vocabulary learning. WM limitations could compromise the learners' success in vocabulary learning process (Alloway & Alloway, 2010). Even though, the study was unable to reveal an involvement of WM in mediating the effects of the input modalities on L2 incidental vocabulary learning, it should be considered when developing multimodal learning conditions. This is because in learning from multimodal situations, students must maintain and process information from more than one source, which is said to be underpinned by WM capacity of the learners (Mayer, 2005).

However, it is necessary that language teachers understand that the incidental vocabulary learning approach is not a substitution for the other ways of vocabulary learning and instruction. Incidental and intentional vocabulary learning should be considered complementary approaches, as each one is important in the incremental process of vocabulary acquisition (Nation, 2001). A good vocabulary plan should incorporate both approaches to ensure promoting as much vocabulary knowledge facets as possible. Teachers should balance opportunities for learners to learn from the two approaches. It has been suggested that the first (K1) and second (K2) thousands most frequent words need to be learned through intentional tasks such as flashcards or mnemonic techniques, as they are of utmost importance in English language and the off-list words or infrequent words can be taught via incidental learning activities (Al-Homoud, 2007). So, the proposed incidental learning conditions in this study are to be viewed as complementary learning situations that could stimulate vocabulary acquisition.

7.4 Limitations and future directions

Despite the efforts exerted and every care taken when developing and executing the present study, the study inevitably has a number of methodological limitations, which should be acknowledged and addressed. However, these limitations do not negate the validity and reliability of the study's outcome. One of the limitations of this study is the small sample size, in particular the number of learners who completed the four WM tests. Though the original plan was to recruit the whole target sample to undertake the four WM tests, due to the time constraints, it was not possible to recruit more than 12 participants from each experimental group to complete the four WM tests. This relatively small sample size (n = 12) might have prevented the statistical significance differences between the performance of high and low WM span groups to surface. The recruitment of a larger sample size may have allowed statistical differences in vocabulary gains between low and high capacity groups to emerge. Therefore, future research should aim to involve sufficient numbers of subjects, as that could potentially address this methodological flaw of this study.

Furthermore, the target words of this study were limited to concrete nouns and verbs. This limitation leaves an open question about whether the effectiveness of different input conditions proves useful for other word classes and for abstract words. This investigation is particularly important, as it would show whether the visual stimuli in conditions (VAC and VA) would positively influence learning of other word classes such as adjectives, adverbs, and abstract nouns. To address this issue, further investigations need to consider whether multiple modalities of audio-visual input can also be effective for promoting knowledge of other parts of speech, such as adjectives, abstract nouns, and adverbs. In addition, this study accounted for the role of the item-level factor (frequency of repetition), however previous research has proposed a range of potential variables (for example, the informativeness of the context of the target words, word salience; topic familiarity) that could impact the participants' ability to learn about the meanings of the target words, which were not controlled for in this study (Vidal, 2011). Thus, further research needs to take into account the potential effects of these factors on incidental vocabulary learning and retention from different input modalities to expand the literature about the roles of these factors in incidental vocabulary learning from different input modalities.

Furthermore, the present study was limited in regards to the use of the three vocabulary tasks to assess short-term learning and long-term retention, which only measured meaning and form. Nonetheless, as outlined in the Literature Review, word knowledge entails several dimensions, such as collocation, grammatical characteristics, and register (see section 2.1.1) which should be considered when examining the effect of different input modalities on incidental vocabulary growth. Although it is sometimes impossible to measure all vocabulary knowledge aspects in one study due to several factors, (i.e. complexity of designing the tests, time limit, cross-test effects (Schmitt, 2010), future research needs to develop additional vocabulary tests to measure a wider spectrum of incidental vocabulary knowledge uptake from different input modes.

Another suggestion for future research is to employ the eye-tracking method for data collection. Montero Perez et al. (2015) stated that the eye-tracking method enabled them to explore whether their learners focused on the video's images or the captions. This could add valuable explanations to the quantitative measures as which input modality students attend to, which in turn would explain which modality affected vocabulary gains better.

A further limitation that have emerged in this study is that learners were categorised into high and low WM span groups based on their scores on the respective WM tests. The resulting group variable was then incorporated in a two-way ANOVA model as a categorical variable. This division may have led to rather small group sizes associated with a loss of statistical power. Therefore, this point warrants future research to consider using a more robust procedure of dividing participants into capacity groups, in order to gain a better understanding of the involvement of WM in vocabulary learning from the various input conditions.

7.5 Concluding remarks

The main goals for this study were to gain more precise insights into the effectiveness of different audio-visual input modalities on L2 incidental vocabulary learning and retention as well as the roles of both factors, frequency of occurrence and WM, in incidental vocabulary learning and retention. The study was successful in yielding several interesting findings regarding the factors investigated, which serve to strengthen our knowledge about incidental vocabulary learning and retention through different input modalities. I hope that this study lays a solid ground for further research to build on. Though the investigations conducted here have explored some important issues in the field of L2 incidental vocabulary acquisition, many issues still remain to be investigated. Research in incidental vocabulary learning and retention should thus continue to examine these unexplored issues.

List of References

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40*(1), 23-31.

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-Homoud, F., & Schmitt, N. (2009). Extensive reading in a challenging environment: A comparison of extensive and intensive reading approaches in Saudi Arabia. *Language Teaching Research, 13*(4), 383-401.

Al-Homoud, F. A. (2007). *Vocabulary acquisition via extensive input.* (Unpublished PhD thesis), University of Nottingham, Nottingham, UK.

Al-Qahtani, A. (2016). Why do Saudi EFL readers exhibit poor reading abilities? *English Language and Literature Studies, 6*(1), 1-15.

Al-Roomy, M. (2013). *An action research study of collaborative strategic reading in English with Saudi medical students.* (PhD thesis), University of Sussex, Sussex, UK.

Al Qahtani, M. (2005). *The use of vocabulary learning strategies by EFL learners at three different educational levels.* (PhD thesis), University of Essex, Essex, UK.

Al Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning and Technology, 5*(1), 202-232.

Alamri, K., & Rogers, V. (2018). The effectiveness of different explicit vocabulary-teaching strategies on learners' retention of technical and academic words. *The Language Learning Journal, 46*(5), 622-633.

Alhazmi, K., & Milton, J. (2015). Phonological vocabulary size, orthographic vocabulary size, and EFL reading ability among native Arabic speakers. *Journal of Applied Linguistics, 30*, 26-43.

Alloway, T. P. (2006). How does working memory work in the classroom? *Educational Research and Reviews, 1*(4), 134-139.

Alloway, T. P. (2007). *Automated working memory assessment (AWMA)*. London: Pearson.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20-29.

Alloway , T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the automated working memory assessment. *Educational Psychology, 28*(7), 725-734.

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., Williams, S., Jones, B., & Cochrane, F. (2014). Exploring the impact of television watching on vocabulary skills in toddlers. *Early Childhood Education Journal, 42*(5), 343-349.

AlSaif, A. (2011). *Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia.* (PhD thesis), Swansea University, Swansea, UK.

Alsaif, A., & Masrai, A. (2019). Extensive Reading and Incidental Vocabulary Acquisition: The Case of a Predominant Language Classroom Input. *International Journal of Education and Literacy Studies, 7*(2), 39-45.

Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal, 40*(1), 21-33.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (Ed.), *Advances in reading/language research: A research annual* (pp. 231-265). Greenwich, CT: JAI Press.

Assalahi, H. M. (2013). Why is the grammar-translation method still alive in the Arab world? Teachers' beliefs and its implications for EFL teacher education. *Theory & Practice in Language Studies, 3*(4), 589-599.

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*. Oxford, UK: Oxford University Press.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology: Section A, 49*(1), 5-28.

Baddeley, A. (1997). *Human memory: Theory and practice*. Hove, UK: Psychology Press.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189-208.

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford, UK: Oxford University Press

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-90). New York, NY: Academic Press.

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Lnguage, 27*(5), 586-595.

Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7*(2), 212-230.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589.

Baltova, I. (2000). *The effect of subtitled and staged video input on the learning and retention of content and vocabulary in a second language.* (PhD thesis), University of Toronto, Toronto, Canada

Bayliss, D. M., Jarrold, C., Baddeley, A. D., & Gunn, D. M. (2005). The relationship between short-term memory and working memory: Complex span made simple? *Memory, 13*(3-4), 414-421.

Bayraktar, D. M., & Altun, A. (2014). The effect of multimedia design types on learners' recall performances with varying short term memory spans. *Multimedia Tools and Applications, 71*(3), 1201-1213.

Beltrán, R. C., Abello-Contesse, C., & del Mar Torreblanca-López, M. (2010). *Insights into non-native vocabulary teaching and learning*. Bristol, UK: Multilingual Matters.

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading, 7*(1), 15-32.

Berninger, V. W., Raskind, W., Richards, T., Abbott, R., & Stock, P. (2008). A multidisciplinary approach to understanding developmental dyslexia within working-memory architecture: Genotypes, phenotypes, brain, and instruction. *Developmental Neuropsychology, 33*(6), 707-744.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509-533.

Boers, F., Warren, P., He, L., & Deconinck, J. (2017). Does adding pictures to glosses enhance vocabulary uptake from reading? *System, 66*, 113-129.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *British Journal of Sociology, 35*(1), 75-92.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97-113.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford: Oxford university press.

Chen, I., & Yen, J. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers & Education, 63*, 416-423.

Chun, D. M., & Payne, J. S. (2004). What makes students click: Working memory and look-up behavior. *System, 32*(4), 481-503.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*(2), 183-198.

Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Coady, J. (1993). Research on ESL/EFL vocabulary acquisition: Putting it in context. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 3-23). Norwood, NJ: Ablex Publishing

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12*(4), 345-360.

Cockcroft, K. (2016). A comparison between verbal working memory and vocabulary in bilingual and monolingual South African school beginners: Implications for bilingual language assessment. *International Journal of Bilingual Education and Bilingualism, 19*(1), 74-88.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: NJ: Erlbaum.

Cohen, J. (1992). A power prime. *Psychological Bulletin, 112*(1), 153-159.

Cohen, L., Lawrence, M., & Keith, M. (2007). *Research methods in education* (6th ed.). New York: Routledge.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*(4), 429-432.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 62–101). Cambridge, UK: Cambridge University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323-338.

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.

Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson Education Limited.

Crotty, M. (2003). *The foundation of social research: Meaning and perspective in the research process.* Thousand Oaks, CA: Sage.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67-77.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Dehn, M. (2008). *Working memory and academic learning: Assessment and intervention*. New Jersey: 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.

Denscombe, M. (2007). *The good research guide for small-scale social research projects* (3rd ed.). Maidenhead: Open University Press.

Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159-165.

Doolittle, P. E., & Altstaedter, L. L. (2009). The effect of working memory capacity on multimedia learning: Does attentional control result in improved performance? *Journal of Research in Innovative Teaching, 2*(1), 7-23.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Ellis, N. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *Consciousness in Second Language Learning, 11*, 37-57.

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16*(4), 409-441.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Engle de Abreu, P. M. J., Gathercole, S. E., & Martin, R. (2012). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences, 21*(5), 569-574.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation, 44*, 145-200.

Feng, Y. (2017). *How does mode of input affect incidental vocabulary learning.* (Master dissertation), Western University, Canada.

Field, A. (2009). *Discovering statistics using SPSS statistics*. London: SAGE Publications Ltd.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.

Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition, 21*(2), 319-333.

Gathercole, S., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. London: Sage Publications Lt.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology, 93*(3), 265-281.

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). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336-360.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 13*(1), 65-77.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*(4), 341-363.

Grix, J. (2010). *The foundations of research* (2nd Ed.). Basingstoke, England: Palgrave Macmillan.

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85-93.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology: Section A, 56*(7), 1213-1236.

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.

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 22*(3), 353-374.

Hamzah, M. S. G., Kafipour, R., & Abdullah, S. K. (2009). Vocabulary learning strategies of Iranian undergraduate EFL students and its relation to their vocabulary size. *European Journal of Social Sciences, 11*(1), 39-50.

Harrington, M., & Carey, M. (2009). The on-line yes/no test as a placement tool. *System, 37*(4), 614-626.

Hennebry, M., Rogers, V., Macaro, E., & Murphy, V. (2017). Direct teaching of vocabulary after listening: Is it worth the effort and what method is best? *The Language Learning Journal, 45*(3), 282-300.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review, 61*(3), 355-382.

Horst, M., Cobb, T., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

Hu, M. (2013). The effects of word frequency and contextual types on vocabulary acquisition from extensive reading: A case study. *Journal of Language Teaching and Research, 4*(3), 487-496.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition, 21*(2), 181-193.

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing, 19*(3), 227-245.

Hulstijn, J. H. (2003). Incidental and intentional learning. In J. C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). London: Blackwell Publishing Ltd.

Hulstijn, J. H., & Schmidt, R. (1994). *Consciousness in second language learning*: AILA 11.

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology, 25*(4), 652-665.

Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology, 20*(3), 123-140.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(02), 137-166.

Kalajahi, S. A. R., & Pourshahian, B. (2012). Vocabulary learning strategies and vocabulary size of ELT students at EMU in northern Cyprus. *English Language Teaching, 5*(4), 138-149.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2011). The relationship between vocabulary and short-term memory measures in monolingual and bilingual speakers. *International Journal of Bilingualism, 15*(4), 408-425.

King, J. (2002). Using DVD feature films in the EFL classroom. *Computer Assisted Language Learning, 15*(5), 509-523.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(2), 261-271.

Kozan, K., Erçetin, G., & Richardson, J. C. (2015). Input modality and working memory: Effects on second language text comprehension in a multimedia learning environment. *System, 55*, 63-73.

Kweon, S., & Kim, H. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language, 20*(2), 191-215.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Béjoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: MacMillan.

Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of'involvement'. *Selected Papers from AILA'99 Tokyo*, 47-62.

Laufer, B. (2010). Form-focused instruction in second language vocabulary learning. In R. Chacón-Beltrán, C. Abello-Contesse, & M. del Mar Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 15-27). Bristol: Multilingual Matters.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research, 15*(4), 391-411.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research, 19*(6), 687-711.

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning, 46*(1), 101-135.

Liu, Y. T., & Todd, A. G. (2014). Dual-modality input in repeated reading for foreign language learners with different learning styles. *Foreign Language Annals, 47*(4), 684-706.

Lobley, K. J., Baddeley, A. D., & Gathercole, S. E. (2005). Phonological similarity effects in verbal complex span. *The Quarterly Journal of Experimental Psychology Section A, 58*(8), 1462-1478.

Logie, R. (1995). *Visuo-spatial working memory*. Hove, England: Erlbaum.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75*(1), 55-74.

Lowie, W., & Seton, B. (2012). *Essential statistics for applied linguistics*. Basingstoke, UK: Macmillan International Higher Education.

Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology, 40*(4), 636-651.

Mackey, A., & Gass, S. M. (2005). *Second language research, methodology and design*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition, 40*, 651-675.

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.

Marzban, A. (2011). Investigating the role of multimedia annotations in EFL reading comprehension. *Procedia-Social and Behavioral Sciences, 28*, 72-77.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34*(5-6), 383-388.

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(3-4), 422-429.

Masrai, A., & Milton, J. (2012). The vocabulary knowledge of university students in Saudi Arabia. *TESOL Arabia perspectives, 19*, 13-20.

Masrai, A., & Milton, J. (2018). The role of informal learning activities in improving L2 lexical access and acquisition in L1 Arabic speakers learning EFL. *The Language Learning Journal, 46*(5), 594-604.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*(2), 125-139.

Mayer, R. E. (2005). *The cambridge handbook of multimedia learning*. Cambridge, UK: Cambridge university press.

Mayer, R. E. (2014). *The cambridge handbook of multimedia Learning*. New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*(4), 484-490.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*(2), 390-397.

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*(4), 806-812.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*(2), 380-386.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107-119.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mayers, A. (2013). *Introduction to statistics and SPSS in psychology*. Harlow, England: Pearson.

McClelland, G. H., Lynch Jr, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, type II errors, and false–positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology, 25*(4), 679-689.

Meara, P. (1997). Toward a new approach to modeling vocabulary acquisition. In N. Schmitt & M. Mccarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge, UK: Cambridge University Press.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. Mccarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 84-102). Cambridge, UK: Cambridge University Press.

Mertens, D. M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. California: SAGE Publications

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review, 4*(1), 151-172.

Mochida, k., & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing, 23*(1), 73-98.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*(2), 308-328.

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1-2), 1-26.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309-326.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94*(3), 598-610.

Nagy, W., & Anderson, R. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304-330.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. Mckeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-36). Hillsdale, NJ: Erlbaum.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253.

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82.

Nation, I. S. P., & Heatley, A. (1994). VocabProfile, word, and range: Programs for processing text.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

Nation, P. (2004). Vocabulary learning and intensive reading. *English Australia Journal, 21*(2), 20-29.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.

Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.

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* (pp. 1-18). New York: Plenum.

Oxford, R. L., & Scarcella, R. C. (1994). Second language vocabulary learning among adults: State of the art in vocabulary instruction. *System, 22*(2), 231-243.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255-287.

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al.(2011). *Journal of Experimental Psychology: General, 142*(1), 282–287.

Pallant, J. (2016). *SPSS survival manual* (6th Ed.). England: Open University Press.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology, 48*(1), 98-107.

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.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 97-130.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*(1), 31-55.

Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring Yes–No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing, 29*(4), 489-509.

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System, 63*, 134-148.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 1-27.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28.

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.

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 282-308.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Ramachandra, V., Hewitt, L. E., & Brackenbury, T. (2011). The relationship between phonological memory, phonological sensitivity, and incidental word learning. *Journal of psycholinguistic research, 40*(2), 93-109.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Reed, S. K. (2006). Cognitive architectures for multimedia learning. *Educational Psychologist, 41*(2), 87-98.

Richards, J. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77-89.

Richards, K. (2009). Interviews. In J. Heigham & R. Crocker (Eds.), *Qualitative research in applied linguistics: A practical introduction* (pp. 182-199). Basingstoke: Palgrave.

Rigg, P. (1991). Whole language in TESOL. *TESOL Quarterly, 25*(3), 520–540.

Rodgers, M. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions.* (PhD thesis), Victoria University, Wellington.

Rodgers, M. P. (2018). The images in television programs and the potential for learning unknown words. *ITL-International Journal of Applied Linguistics, 169*(1), 191-211.

Sadoski, M. (2005). A dual coding view of vocabulary learning. *Reading & Writing Quarterly, 21*(3), 221-238.

Sadoski, M., McTigue, E. M., & Paivio, A. (2012). A dual coding theoretical model of decoding in reading: Subsuming the laberge and samuels model. *Reading Psychology, 33*(5), 465-496.

Sadoski, M., & Paivio, A. (2004). A dual coding theoretical model of reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1329–1362). Newark, DE: International Reading Association.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in Second Language Learning, 11*, 237-326.

Schmitt, N. (1995). The word on words: An interview with Paul Nation. *The Language Teacher, 19*, 5-7.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48*(2), 281-317.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge Univiversity Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2015). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching, 50*(2), 212-226.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26-43.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition, 19*(1), 17-36.

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.

Schweppe, J., & Rummer, R. (2014). Attention, working memory, and long-term memory in multimedia learning: An integrated perspective based on process models of working memory. *Educational Psychology Review, 26*(2), 285-306.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology, 45*(1), 21-50.

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *ELT journal, 64*(3), 253-260.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*(1), 121-159.

Sorden, S. D. (2012). The cognitive theory of multimedia learning. In B. J. Irby, G. Brown, & R. Lara-Alecio (Eds.), *Handbook of educational theories* (pp. 1-31). Charlotte, NC: Information Age Publishing.

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.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577-607.

Swanson, L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence, 35*(2), 151-168.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14*(2), 50-73.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54*(1), 68-80.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6), 1038-1066.

Vallar, G., & Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology, 1*(2), 121-141.

van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609-624.

Varol, B., & Erçetin, G. (2016). Effects of working memory and gloss type on L2 text comprehension and incidental vocabulary learning in computer-based reading. *Procedia-Social and Behavioral Sciences, 232*, 759-768.

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141*, 65-82. doi:10.1016/j.jecp.2015.06.015

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*(2), 217-234.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56-89.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219-258.

Vogt, W. P. (2007). *Quantitative research methods for professionals*. Boston: Pearson Education.

Waring, R. (1997). Graded and extensive reading: Questions and answers. *The Language Teacher, 21*, 7-13.

Waring, R. (1999). *Tasks for assessing second language receptive and productive vocabulary.* (PhD thesis), University of Wales, Swansea, UK.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(1), 33-52.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232-245.

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667-686.

Wen, Z. (2016a). Phonological and executive working memory in L2 task-based speech planning and performance. *The Language Learning Journal, 44*(4), 418-435.

Wen, Z. (2018). Using formulaic sequences to measure task performance. The role of working memory. In M. J. Ahmadian & M. P. G. Mayo (Eds.), *Recent Perspectives on Task-based Language Learning and Teaching* (pp. 29-52). Boston, MA/Berlin: De Gruyter.

Wen, Z. E. (2016b). *Working memory and second language learning: Towards an integrated approach*. Bristol, UK: Multilingual matters.

Wen, Z. E. (2019). Working memory as language aptitude: The phonological/executive model. In Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. L. Sparks (Eds.), *Language Aptitude: Advancing Theory, Testing, Research and Practice* (pp. 187-214). London: Routledge.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*(1), 13-40.

Wesche, M., & Paribakht, T. S. (1999). Introduction. *Studies in Second Language Acquisition, 21*(2), 175-180.

Wiley, J., Sanchez, C. A., & Jaeger, A. J. (2014). The individual differences in working memory capacity principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 598–619). New York: Cambridge University Press.

Wilkins, D. (1972). *Linguistics in language teaching*. London: Edward Arnold.

Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning and Technology, 14*(1), 65-86.

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist, 11*(2), 87-95.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology, 10*(3), 85–101.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review, 57*(4), 541-572.

Zhong, H. (2009). Vocabulary growth in an English as a foreign language context. *University of Sydney Papers in TESOL, 4*(4), 85-113.

Appendix A Vocabprofile analysis output

Home > VocabProfilers > English (Alt-arrow-left to preserve settings) > **Output**

WEB VP OUTPUT FOR FILE: Untitled (8.85 kb)

Words recategorized by user as 1k items (proper nouns etc): NONE (total 0 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	300	376	1325	81.24%
Function:	...	...	(816)	(50.03%)
Content:	...	...	(509)	(31.21%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	...	...	(329)	(20.17%)
K2 Words (1001-2000):	83	93	138	8.46%
> Anglo-Sax:	...	...	(87)	(5.33%)
1k+2k			...	(89.70%)
AWL Words (academic):	28	31	37	2.27%
> Anglo-Sax:	...	...	(2)	(0.12%)
Off-List Words:	?	96	131	8.03%
	411+?	596	1631	100%

Current profile	
%	Cumul.
81.24	81.24
8.46	89.70
2.27	91.97
8.03	100.00

Words in text (tokens):	1631
Different words (types):	596
Type-token ratio:	0.37
Tokens per type:	2.74
Lex density (content words/total)	0.50
Pertaining to onlist only	
Tokens:	1500
Types:	500
Families:	411
Tokens per family:	3.65
Types per family:	1.22
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

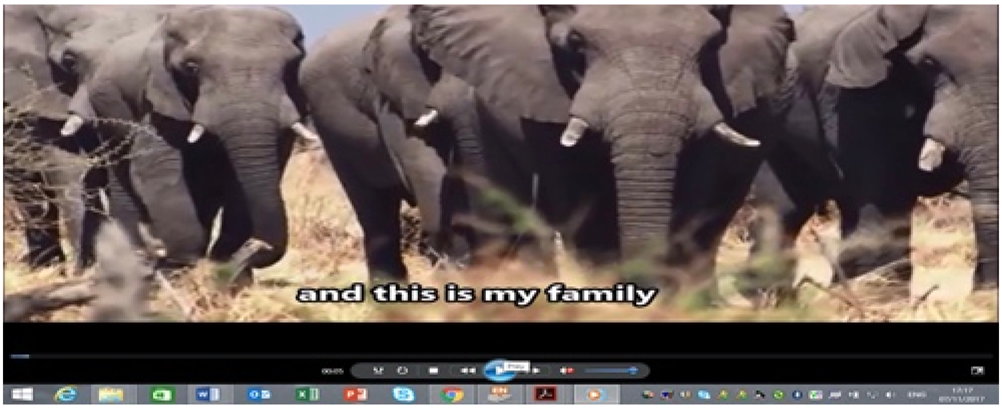
For easy editing as MS Word or Excel table - Edit/print-friendly table

Page menu: Tokens (With shorter texts; includes AWL subs) Types Families (With VP negative + Anglo-Saxons) Colour-Coded Text

Appendix B Examples of treatment condition (VAC) computer screen

The power of lionesses

Elephant



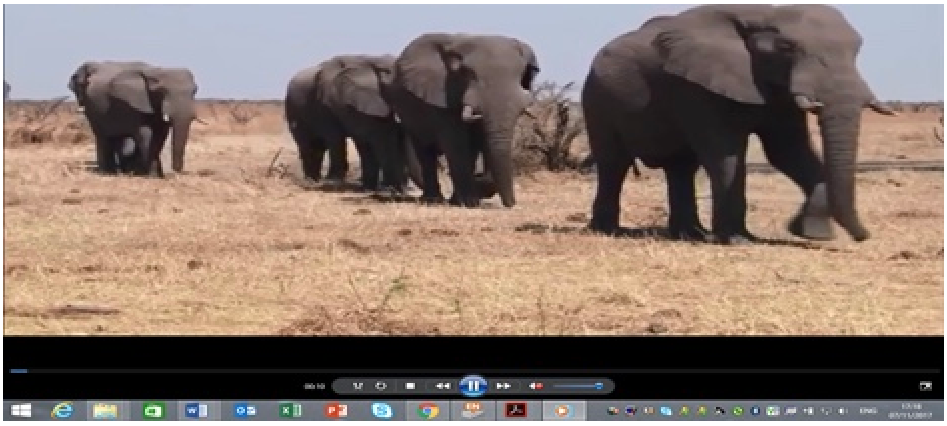
Kangaroo

The wonderful dogs

Appendix C Examples of Treatment condition (VA) computer screen

The power of lionesses

Elephant



Kangaroo

The wonderful dogs

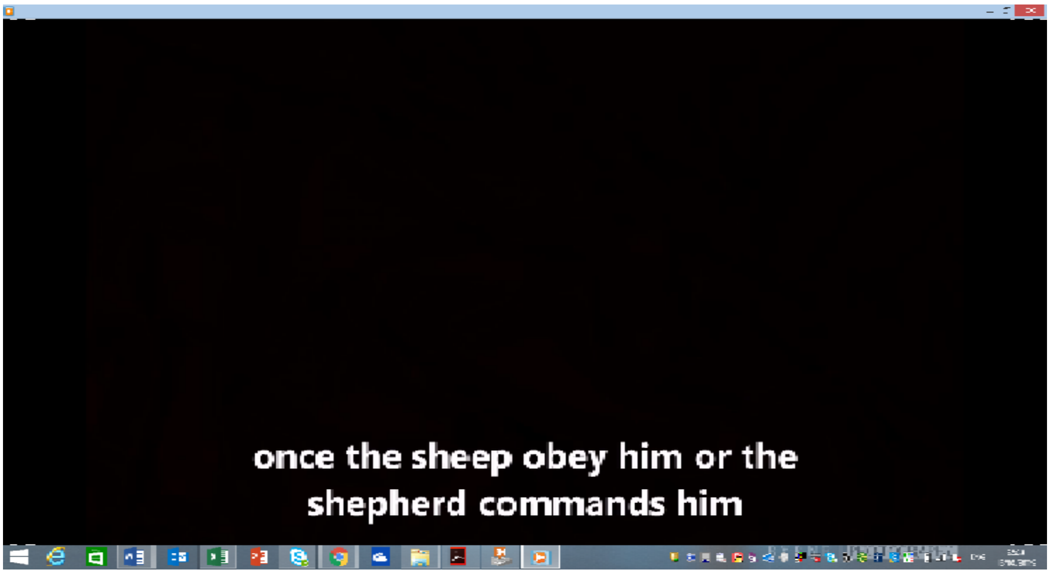
Appendix D **Examples of Treatment condition (CA) computer screen**

The power of lionesses

Elephant

Kangaroo

The wonderful dogs



Appendix E Consent Form and Participant Information Sheet

نموذج موافقة على المشاركة في الدراسة

Study title: L2 Incidental Vocabulary Learning and Retention through Different modalities of Audio-visual Input

عنوان الدراسة: تعلم مفرادات اللغة الإنجليزية من خلال مواد مرئية مسموعة

Researcher: Hassan Alshumrani

اسم الباحث: حسن الشمراني

University of Southampton Ethics approval Number: 25101

Staff/Student number: 226907844

Please initial the boxes if you agree with the statements:

أمام العبارات التالية إذا كنت توافق على محتواهاvمن فضلك قم بوضع علامة

I have read and understood the Participant Information Sheet and have had the opportunity to ask questions about the study. ☐

لقد فهمت المراد من الدراسة ولي الحق في السؤال عنها متى ما أشاء

I agree to take part in this research project and agree for my data to be used for the purpose of this study. ☐

أوافق على المشاركة في هذه الدراسة البحثية ولا أمانع في استخدام بياناتي لغرض الدراسة.

I understand my participation is voluntary and I may withdraw at any time without my legal rights being affected. ☐

أدرك أن مشاركتي في هذه الدراسة تطوعية ويحق لي الانسحاب منها متى ما أريد.

Data Protection

I understand that information collected about me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

كما أني أدرك أن معلوماتي وبياناتي المقدمة اثناء مشاركتي في هذه الدراسة ستحفظ بسرية تامة ولا يحق لأي أحد سوى الباحث ومشرفه الاطلاع عليها. كما أن جميع البيانات ستستخدم فقط لأغراض هذه الدراسة

اسم المشارك / Name of participant (print name)……………………………………………………………..

التوقيع / Signature of participant…………………………………………………………………………….

التاريخ / Date……………………………………………………………………………….

Participant Information Sheet

Study Title: L2 Incidental Vocabulary Learning and Retention through Different modalities of Audio-visual Input

Researcher: Hassan Alshumrani

University of Southampton Ethics approval Number: 25101

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?

I am Hassan Alshumrani, a doctoral student at the University of Southampton, UK. This research is undertaken as a doctoral project, which is part of the requirements needed for PhD degree in Applied Linguistics. I am interested in examining the effects of different modalities of audio-visual input on L2 incidental vocabulary learning and retention. I am also interested in exploring the role of working memory in this process. The following research questions are addressed in this study:

This project is funded by the Saudi Arabia government and sponsored by the University of Southampton.

Why have you been chosen?

As a Saudi English language learner, you have been selected to take part in the present study.

Are there any benefits in my taking part?

My study will add to current knowledge regarding the effects of multimodalities of audio-visual input on L2 incidental vocabulary learning and retention. As a participant in this study, you will be able to find out which input mode (s) suits your vocabulary learning and retention best. Also, you will explore the predicative role of your working memory in vocabulary learning.

Are there any risks involved?

No sensitive topics (such as politics or other issues regarding private life of the participants) will be included in the study materials or in the data collection instruments. There will be no risk or harm involved in this research.

Will my participation be confidential?

Definitely, it will be. The following steps will be exercised to ensure confidentiality. Firstly, the information obtained from or about participants will be kept on a password-protected computer. Secondly, the participants will be given pseudo-names in my research (e.g. in three vocabulary tests and the working memory tests). Thirdly, collected data will not be shown by or displayed to any person, including their class teacher, other than the researcher and his supervisors. Undoubtedly, the data will remain absolutely confidential, and anonymity is also assured.

What happens if I change my mind?

Your participation in this research is voluntary. Once you change your mind and want to withdraw, you can do this without being required to even give a reason. This will not affect your study at all. It would be highly appreciated if you just let the researcher know. In addition, you will have the opportunity to attend other classes, so that their learning is not disadvantaged by your withdrawal.

What happens if something goes wrong?

If you have any concerns or complaints, you can contact Prof. Denis McManus (D.Mcmanus@soton.ac.uk). Or, Research Governance (02380 595058, mad4@soton.ac.uk) is happy to be the named party.

Where can I get more information?

Contact the researcher: Hassan Alshumrani at 00447478667298 or email me at haa1g14@soton.ac.uk.

Appendix F Non-words list adopted from Waring and Takaki (2003) (12 items)

Mand
Mear
Mork
Smorty
Tantic
Jurgs
Molden
Tring
Nase
Palk
Tance
Vack

Appendix G Example of the Spoken Form Recognition Test

Answer sheet (first 7 items)

Listen to a list of words carefully. For each word: choose (Yes) if you heard the word in the treatment sessions, or choose (No) if you did not hear the word.

استمع إلى مجموعة من الكلمات: اختر (نعم) إذا كنت سمعت الكلمة في المادة العلمية، أو أختر (لا) إذا لم تسمعها

	yes	No
1		
2		
3		
4		
5		
6		
7		

Appendix H Example of the Spoken Meaning Recall Test

Answer sheet (first 7 items)

Listen to the following words and write down anything you know about them (definition, synonym, or L1 translation). Tick (√) I don't know, if you do not know the meaning.

استمع إلى الكلمات التالية، اكتب ما تعرفه عن كل كلمة إما (تعريف، مرادف، أو ترجمة بلغتك الأم).

أختر (I don't Know)إذا لم تعرف المعنى،

	Meaning المعنى	I don't know
1		
2		
3		
4		
5		
6		
7		

Appendix I Example of the Spoken Meaning Recognition Test Answer sheet (first 7 items)

Listen to the following words and circle the option with the nearest meaning to the word you hear.

Circle the option (d) “I don’t know”, if you do not know the meaning.

استمع إلى الكلمات التالية وضع دائرة حول الإجابة الصحيحة، أختر (I don’t Know) إذا كنت لا تعرف الإجابة

The participants heard the following words on the recording (herd, swamp, dehydration, trunk, chase, thrive, tendons). They heard each word twice with a 10-second silence between the two presentations.

1.
a) a team of players
b) a group of animals
c) a space for worship
d) I don’t know

2.
a) a place to visit
b) a part of body
c) a wet area of ground
d) I don’t know

3.
a) piece of paper folded in half
b) loss of normal levels of water
c) kind of exercise machine
d) I don’t know

4.
a) a bowl of rice
b) a nose of an elephant
c) a slice of bread
d) I don’t know

5.
a) to follow in order to catch
b) to leave a place
c) to find someone
d) I don’t know

6.

a) to give somebody a present
b) to grow and develop
c) to exchange letters
d) I don't know

7.

a) spaces with nothing in them
b) people working together
c) tissues in the body connecting a muscle to bone
d) I don't know

Appendix J Interview guide

- What do you think about the modality of input you had? (VAC, VA, CA, A only)
- What did you like/not like in this input condition?
- What did you find useful/not useful in this combination of input modes? Why?
- What did you pay attention to during the activity, (video, caption, audio?), Why?
- What modality helped/not helped to learn new words?
- What do you think we should add to/delete from this combination of input modes?
- Have you encountered any difficulties in the experiment?

Appendix K Welch ANOVA tests

Robust Tests of Equality of Means

		Statistica	df1	df2	Sig.
spoken form recognition immediate post-test	Welch	41.216	3	57.391	.000
Spoken meaning recall immediate post-test	Welch	14.493	3	56.867	.000
Spoken meaning recognition immediate post-test	Welch	22.151	3	57.045	.000

a. Asymptotically F distributed.

Appendix L ANOVA results for the four WM measures and incidental vocabulary learning

ANOVA results for the composite scores for the three vocabulary immediate post-tests by treatment condition and forward digit recall WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	2461.560[a]	7	351.651	6.053	.000	.514
Intercept	53742.900	1	53742.900	925.129	.000	.959
Experimental Groups	2297.830	3	765.944	13.185	.000	.497
Forward digit recall test	.710	1	.710	.012	.913	.000
Experimental Groups * forward digit recall test	168.508	3	56.169	.967	.418	.068
Error	2323.690	40	58.092			
Total	59056.000	48				
Corrected Total	4785.250	47				

ANOVA results for the composite scores for the three vocabulary immediate post-tests by treatment condition and backward digit recall WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	2441.502[a]	7	348.786	5.953	.000	.510
Intercept	50779.500	1	50779.500	866.637	.000	.956
Experimental Groups	2297.640	3	765.880	13.071	.000	.495
Backward digit recall test	140.626	1	140.626	2.400	.129	.057
Experimental Groups * backward digit recall test	7.455	3	2.485	.042	.988	.003
Error	2343.750	40	58.594			
Total	59056.000	48				
Corrected Total	4785.250	47				

ANOVA results for the composite scores for the three vocabulary immediate post-tests by treatment condition and dot matrix WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	2762.441[a]	7	394.634	7.804	.000	.577
Intercept	37811.600	1	37811.600	747.705	.000	.949
Experimental Groups	1671.400	3	557.134	11.017	.000	.452
Dot Matrix test	123.646	1	123.646	2.445	.126	.058
Experimental Groups * dot matrix test	419.846	3	139.949	2.767	.054	.172
Error	2022.810	40	50.570			
Total	59056.000	48				
Corrected Total	4785.250	47				

ANOVA results for the composite scores for the three vocabulary immediate post-tests by treatment condition and the odd one out WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	2487.927[a]	7	355.418	6.188	.000	.520
Intercept	43767.600	1	43767.600	762.063	.000	.950
Experimental Groups	1750.610	3	583.535	10.160	.000	.432
The odd one out test	49.140	1	49.140	.856	.361	.021
Experimental Groups * the odd one out test	147.873	3	49.291	.858	.471	.060
Error	2297.320	40	57.433			
Total	59056.000	48				
Corrected Total	4785.250	47				

Appendix M ANOVA results for the four WM measures and incidental vocabulary retention

ANOVA results for the composite scores for the three vocabulary delayed post-tests by treatment condition and forward digit recall WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	686.779[a]	7	98.111	2.032	.075	.262
Intercept	8643.117	1	8643.117	179.021	.000	.817
Experimental Groups	549.112	3	183.037	3.791	.017	.221
Forward digit recall test	70.549	1	70.549	1.461	.234	.035
Experimental Groups * forward digit recall test	66.921	3	22.307	.462	.710	.033
Error	1931.200	40	48.280			
Total	11339.000	48				
Corrected Total	2617.979	47				

ANOVA results for the composite scores for the three vocabulary delayed post-tests by treatment condition and backward digit recall WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	580.882[a]	7	82.983	1.629	.155	.222
Intercept	8036.971	1	8036.971	157.812	.000	.798
Experimental Groups	478.041	3	159.347	3.129	.036	.190
Backward digit recall test	8.813	1	8.813	.173	.680	.004
Experimental Groups * backward digit recall test	22.547	3	7.516	.148	.931	.011
Error	2037.098	40	50.927			
Total	11339.000	48				
Corrected Total	2617.979	47				

ANOVA results for composite scores for the three vocabulary delayed post-tests by treatment condition and dot matrix WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	795.665[a]	7	113.666	2.495	.032	.304
Intercept	5442.809	1	5442.809	119.470	.000	.749
Experimental Groups	332.549	3	110.850	2.433	.079	.154
Dot matrix test	1.213	1	1.213	.027	.871	.001
Experimental Groups * dot matrix test	246.988	3	82.329	1.807	.161	.119
Error	1822.314	40	45.558			
Total	11339.000	48				
Corrected Total	2617.979	47				

ANOVA results for the composite scores for the three vocabulary delayed post-tests by treatment condition and the odd one out WM test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	694.176[a]	7	99.168	2.062	.071	.265
Intercept	7047.995	1	7047.995	146.543	.000	.786
Experimental groups	367.027	3	122.342	2.544	.070	.160
The odd one out test	3.243	1	3.243	.067	.796	.002
Experimental groups * the odd one out test	143.551	3	47.850	.995	.405	.069
Error	1923.803	40	48.095			
Total	11339.000	48				
Corrected Total	2617.979	47				