The Impact of Integrated and Split-attention Online Reading Formats on Building Adult EAP Learners' Vocabulary Knowledge

Hadeel Awad

School of English Literature, Language and Linguistics

Newcastle University

This dissertation is submitted for the degree of

Doctor of Philosophy

October 2022

School of English Literature, Language and Linguistics

Percy Building

Newcastle University

NE1 7RU

United Kingdom

Abstract

Abstract

Cognitive load theory has been used by second language acquisition researchers to explain differences in learners' performance when they are exposed to different instructional formats. The cognitive load related to second language online reading can be affected by many factors. One of the factors is the L2 proficiency of the readers. A second factor is the amount of text displayed on screen. A third factor is the effectiveness of instructional formats. Various instructional formats have been found to affect the learners' cognitive load and working memory differently.

None of the studies which looked at the effect of certain instructional formats on students' cognitive load and their building of vocabulary knowledge have considered the influence of the online split-attention and integrated instructional formats on EAP learners' vocabulary acquisition in particular. This study is designed to investigate whether there is a relation between cognitive load and building vocabulary knowledge; thus, whether word meaning knowledge, word family knowledge, or collocation knowledge among EAP students can be learnt more effectively depending on the different instructional formats of the online reading texts. The two different instructional formats are split-attention and integrated instructional formats. Moreover, the study investigates whether different proficiency levels among learners or their first languages may affect building their vocabulary knowledge in both split-attention or integrated instructional formats.

A quasi-experimental intervention study was conducted to collect quantitative data needed to answer the research questions. The aim of the study was to investigate the effect of split-attention and integrated reading formats on EAP learners' vocabulary knowledge. Specifically, are reading tasks in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word meaning, word families and collocations knowledge? Do adult EAP learners at different proficiency levels, or with different first languages acquire new vocabulary items differently with the different formats?

To answer the research questions, sixty participants (35 Arabic speakers, 25 Chinese speakers) who were a group of adult learners studying English for University Studies at a language centre in a UK university, were assigned into low proficiency level learners and high proficiency level groups according to their IELTS scores. Participants were divided into a group exposed to split-

attention reading instructional format and a group exposed to integrated reading instructional format. Data was collected using a pre-test, treatment, and a post-test.

The results showed that there was an increase in the number of participants who answered the vocabulary questions correctly after the experiment in both *split-attention* and *integrated* groups. This increase was mainly in word meaning and word family questions. A (paired-samples t-test) showed a statistically significant increase in the test scores from pre-test to post-test, and the results showed that the increase was statistically significant in word meaning and collocation knowledge, but not in word family knowledge; however, there was no statistically significant difference between the results of the *split-attention* and the *integrated* group in the post-test although the scores of the *integrated* group were higher than those of the *split-attention* group. While participants of higher proficiency level gave significantly more correct answers than those of lower proficiency level in the *split-attention* group, the difference in the *integrated* group was not statistically significant. L1 did not have any statistically significant effect on acquiring new vocabulary items in either *split-attention* or *integrated* groups.

Acknowledgements

I would like to express my heartfelt gratitude to everyone who supported me throughout my PhD journey.

First and foremost, I am enormously grateful to my supervisors, Prof. Martha Young-Scholten, Dr. William Scott Windeatt, and Prof. Michael Rossington, whose expertise, patience, and support have been invaluable throughout my research and writing process. I also wish to thank my examiners, Dr. Edward Moran and Dr. Emma Nguyen, for their insightful comments and constructive feedback which greatly improved the quality of my work.

Special thanks go to INTO, Newcastle University, for providing the ideal environment for collecting the data needed for this research, and to the talented programmer, Mr. Sami Bader, who wrote my website and gave his precious insights to make it a successful research tool.

This study would not have been possible without the financial sponsorship of Tishreen University in Latakia, Syria, who mainly supported me financially to achieve this degree.

I am thankful to all my friends who supported me throughout this long journey. Special thanks go to two friends who have a very much appreciated academic input to this thesis, Dr. Khansaa Tezeny, and Dr. Alaa Balkees.

I am deeply indebted to my family for their unwavering support and encouragement. To my parents who were also my teachers during school years, Mufeed and Mariette Awad, my siblings, Waseem and Rima, and their families, your love and encouragement have surrounded and motivated me continuously. To my husband, George, without whom this study would not have seen the light, thank you for giving me all the love and support in the world and for your patience throughout the years. To my daughters, Danielle and Gabriella, no words can describe how patient and loving you both were, how hard it was for Danielle to look after a little sister, and how emotional it was to write up with Gabriella sitting on mum's knees. To my extended family members whether they are still with us or left us before seeing my finished work, thank you all for believing in me.

Hadeel Awad

Newcastle October 2022

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Background of the Study

A very large vocabulary knowledge is needed in order for learners to be able to communicate and function in any language. According to Alsahafi (2023)**,** appropriate L2 lexical knowledge is essential for the capability to acquire the second language receptive and productive skills efficiently, and the depth of an EFL students' academic vocabulary knowledge may effectively predict their academic performance and success; thus, over the past forty years or so, second language vocabulary has more and more been acknowledged as an essential constituent of L2 global competence which is highly required during language teaching and learning. McCarthy (1990, viii) states that "no matter how well the student learns grammar, no matter how successfully the sounds of L2 are mastered, without words to express a wider range of meanings, communication in an L2 just cannot happen in any meaningful way". Vocabulary knowledge does not only depend on the number of words a person knows, but also on the person's ability to comprehend the meaning of the words they know. According to Nation (2001), vocabulary knowledge includes the three fields of form, meaning, and use. In addition to that, vocabulary knowledge constitutes of two aspects; breadth and depth. Whereas the depth of vocabulary knowledge refers to the degree to which lexical items are known, focused, strengthened, and explored (Firda et al., 2021), the breadth of vocabulary knowledge refers to the amount of vocabulary known by someone. Laufer (1992) states that a minimum of 3000 words is needed for learners to understand non-fictional reading texts, whereas up to 10000 words are needed by students studying in English at university level (Hazenberg and Hulstun, 1996). Nation (2022) argues that learners should stock over 5000 words in order to be able to listen and speak using the language words. Similarly, the size of vocabulary required for learners to be able to read a basic realistic reading text is around 3000 words, whereas reading at a more advanced stage requires the size of 5000 words (Schmitt et al., 2001). Another large size of more than 8000-9000 word families needed to read English texts which are not adjusted is suggested by Nation (2006). In order for learners to acquire this large size of word knowledge, they should be helped develop their vocabulary knowledge using certain learning strategies which allow them to acquire less frequent or utilised vocabulary in addition to the most frequent or utilised words in the language they are learning. The vocabulary teaching exercises used in this study are taken from Schmitt and Schmitt (2011) where the authors have tried to use the

most efficient research in order to design the most useful exercises for second language vocabulary learners.

One of the most important points to take into consideration is that learning a word implies more than the knowledge of the word meaning, spelling, and pronunciation only (Schmitt, 2000). As a matter of fact, additional criteria to word knowledge involve, amongst other criteria, word collocations, grammatical characteristics, and word family knowledge. Learners should have an understanding of all the forms of word knowledge in order for them to use words confidently. The target vocabulary in this study is introduced within reading texts in addition to vocabulary exercises accompanying the texts. Moreover, the vocabulary exercises in this study follow Schmitt and Schmitt (2011)'s model of concentrating on the three aspects of word meaning, word families (derivative forms of words), and collocations in order to achieve the maximum teaching goal of the exercises. According to Schmitt and Zimmerman (2002), learners do not necessarily know all of the categories of a word family although they may be aware of some of its forms. Nevertheless, learners should know the right form of a word (noun, verb, adjective, or adverb) in a certain context, and hence the focus on derivative forms of words in this study. Learning word collocations is also very essential when learning new vocabulary. Although it is hard to teach all word pairings, integrating the element of collocations knowledge into the exercises in this study aims at helping the learners become more aware of collocations and thus enhancing their intuitive understanding of collocations for individual words.

Another important point to consider during second language vocabulary learning is the importance of context. The rationale behind introducing the target vocabulary within reading texts in addition to exercises in this study is the importance of learning these words within as many various contexts as possible. Various contexts present various types of information about the target vocabulary. Thus, one specific context can teach one meaning of a word whereas another context can teach a totally different meaning of the same word.

Studies which have researched second language vocabulary learning have focused on two major approaches to language learning in general, and vocabulary learning specifically. Barcroft (2009) focused on the effect of *intentional* learning (learning which takes place while there is intention for it); and Schmitt (2010) focused the effect of *incidental* learning (learning which takes place while using language without any special aims of learning any linguistic feature). Comparing the above-mentioned focal points of research, researchers such as Laufer (2003) and Lin and Hirsh (2012) have found that the positive results of intentional learning through explicit teaching have been obvious. Nevertheless, the large size of vocabulary that learners

should know cannot be accounted for by explicit teaching on its own, but by incidental acquisition during exposure to written and spoken language. Huckin and Coady (1999) state that most L2 vocabulary scholars seem to agree that apart from the first few thousand most common words, L2 vocabulary is mainly acquired through extensive reading when the learner guesses the meaning of most of the new words. The same point was discussed in (Carter and McCarthy, 1988, Coady and Huckin, 1997, Huckin et al., 1993, Lewis, 1993, McCarthy, 1990, Nation and Carter, 1989, Schmitt and McCarthy, 1997, Gairns and Redman, 1986, Nation, 1990).

Online reading has been found to be beneficial for text comprehension (Anderson, 2003, Akyel and Erçetin, 2009, Huang et al., 2009). Electronic documents allow for different formats of reading texts and various ways of processing information which can be more challenging for readers when they infer meaning from traditional paper-based reading texts (Coiro, 2003, Protopsaltis, 2008). However, online ESL reading materials have been found to include distractors which may lead to an increased cognitive load , and eventually prevent effective reading (Akbulut, 2007). "Cognitive load" can be defined as the amount of information that can be placed in the working memory at one time. Working memory is the cognitive system which temporarily stores information due to lack of capacity to store information long term, and which is in control of maintaining and manipulating information while the brain has a cognitive task to perform (Baddeley and Hitch, 1974, Daneman and Carpenter, 1980). Both "cognitive load" and "working memory" will be discussed in more detail in this thesis. Research has found that the cognitive load related to online multimedia-based reading in general and L2 online reading specifically can be affected by many factors (Al-Shehri and Gitsaki, 2010). One of these factors is the L2 proficiency of the readers. The second factor is the effectiveness of instructional formats. The third factor is the amount of text displayed on screen.

In order to tackle the effectiveness of instructional formats and the amount of text on screen, this study investigates the impact of two instructional formats; namely, integrated and split-attention online reading formats on building adult EAP learners' vocabulary knowledge. The cognitive load theory (CLT), coined in 1988 by John Sweller, and cognitive theory of multimedia learning, by Richard Mayer (1997), are adopted as the main theoretical framework to this study. Cognitive load theory suggests that various learning tasks affect cognitive load and working memory differently, and this effect is emphasised in a multimedia environment where various interacting elements and tools can result in cognitive overload and a decrease in the learning results. Cierniak et al. (2009) state that "the rationale of cognitive load theory CLT is that the designs of instruction impose cognitive load on learners' limited working memory

and that the cognitive load in turn influences learning outcomes". Studies which use CLT as a theoretical frame have researched various methods of reducing students' cognitive load in order to enhance their learning outcomes. CLT focuses on verbal as well as visual representations of information, and both verbal and visual information representations are integrated into multimedia; thus, researchers have presented the Cognitive Theory of Multimedia Learning (CTML) which is a multimedia-based learning theory with effective theoretical and practical suggestions and implications, and which, according to Moreno and Valdez (2005), can result in successful learning process by helping learners choose from multiple cognitive processes based on multimedia.

It is very important for learners of English as a second language to improve their vocabulary knowledge. According to the cognitive load theory, using the integrated instructional format by dividing the reading text into paragraphs and inserting the related vocabulary questions within the paragraphs, as opposed to using the split-attention format where the whole reading text is followed by the related vocabulary questions, is predicted to reduce the cognitive load on the working short-term memory and lead to better learning of the target vocabulary. To examine this assumption, the researcher has aimed at investigating whether an online reading task in the integrated format could be more effective than in the split-attention format in helping adult EAP learners build their vocabulary knowledge, and whether the learners' English proficiency levels or their first languages could affect their vocabulary knowledge building.

To answer the research questions in this study, a quantitative approach was used. The method used in order to investigate the topic of this study is quasi-experimental. This method was used to collect quantitative data needed for the research. The treatments are integrated and split-attention formats, and the measurements to collect data are the pre- and post-tests. By post-test, it is meant that an immediate post-test was conducted. All experiments were held at a language Centre in a UK university.

1.2 Significance of the Study

Most of the research which focused on the potential of multimedia in building vocabulary knowledge among English Speakers of Other Languages (ESOL), whether they are English for Academic Purposes (EAP) learners or not, even the studies that investigated the impact of certain instructional formats on students' cognitive load and their building of vocabulary knowledge, paid particular attention to multimedia glosses and vocabulary knowledge (Chun, 2001), multimedia annotations and vocabulary acquisition (Akbulut, 2007), or even

multimedia-enhanced dictionaries (Peters, 2007). Nevertheless, none of the studies that I have come across so far have focused particularly on the impact of the online split-attention and integrated instructional formats on EAP learners' vocabulary acquisition in reading.

This study is designed to investigate whether there is a relation between cognitive load and building vocabulary knowledge; thus, whether vocabulary among EAP students can be learnt more effectively depending on the different instructional formats of the online reading texts; namely, split-attention and integrated instructional formats. The split-attention format is assumed to have a higher cognitive load on EAP students; and thus, be led effective in helping them build their vocabulary knowledge.

Although Al-Shehri and Gitsaki (2010) tried to research the impact of split-attention and integrated instructional formats on students' cognitive load and the way they might lead to more efficient reading of second language texts online and easier learning of vocabulary, their main focus was the effect of different instructional formats on reading comprehension and reading speed, while the focus of this study is the effect of the instructional formats used in the study of Al-Shehri and Gitsaki (2010) on building vocabulary knowledge.

Furthermore, there are two variables, according to previous research, which might play an important role in the effect of certain instructional formats on EAP learners' vocabulary acquisition which Al-Shehri and Gitsaki (2010) did not mention in their study. These variables are the proficiency level of the learners and their first language (L1).

Moreover, Although Al-Shehri and Gitsaki (2010) tried to investigate whether a reading task in the integrated format was more effective than in the split-attention format in making L2 reading easier for ESL readers, and whether an online dictionary with either of the two above-mentioned formats could affect students' reading comprehension and speed, they only implemented a post-test which measured the receptive vocabulary of the participants in their study; their study did not include any pre-test to measure the vocabulary knowledge of their learners before the experiment and compare the results with the post-test.

So, it can be said that Al-Shehri and Gitsaki (2010) was preliminary, and although it had some problems with the methodology that was used, part of the methodology used in their study was adopted and built on in this study.

By considering the above features, it can be claimed that this thesis is an original one; the study is the first of its kind since no previous studies investigated the particular effect of the different

reading instructional formats on building second language vocabulary knowledge. When Al-Shehri and Gitsaki (2010) tried to investigate this field, their methodology had some gaps which this study hopes to fill so that it could contribute to the development of new approaches to teaching vocabulary by using different instructional formats which could reduce the cognitive load on working memory. This study aims at demonstrating that using an integrated instructional format leads to higher vocabulary gains compared with a split-attention format.

1.3 The Scope of the Study

As mentioned above, this study examined the impact of integrated and split-attention online reading formats on building the vocabulary knowledge among adult EAP learners on an Academic English course at a language centre in a UK university. In order to provide the essential background information of the research, the literature related to cognitive load theory, cognitive theory of multimedia learning, working memory and its role in second language acquisition, vocabulary acquisition, second language vocabulary teaching and learning, learning vocabulary through reading, and online reading was reviewed. This literature review defined and limited the research problem. It also developed the rationale for the research questions in addition to exploring the suitable research methods which were used to collect the data needed for this research.

Although vocabulary acquisition and cognitive load were the main focus of this study, fields like reading in general, pros and cons of online reading, in addition to working memory and its roles in first and second language acquisition were discussed because of their very strong relationship to the researched topic.

Students on an Academic English course at a language centre in a UK university were chosen to be the participants of this study. This sample was based on the assumption that these students started learning English language in general, and specifically focusing on English grammar, at a young age in their home countries, but they needed to build their vocabulary knowledge to be able to continue their studies at a university level and to be able to communicate in English while living in England.

1.4 Thesis Organisation

In addition to the introduction, the content of this thesis is divided into five chapters: the literature review of the topics related to the study, research methodology, data analysis and results, discussion of the findings and the conclusion.

Chapter two, the literature review, starts with a definition and explanation of what Cognitive Load Theory (CLT) is. A background of Cognitive Load Theory, in addition to Cognitive Theory of Multimedia Learning is presented. Following that is a review of working memory, how it is measured, and its role in first as well as second language acquisition. The next section gives an overview of vocabulary acquisition, a background of the teaching and learning of vocabulary, and then explains the nature of vocabulary learning and some of the aspects involved in learning a word. The discussion of word meaning knowledge gives an explanation of the nature and criteria of this knowledge; namely, receptive vs. productive knowledge, vocabulary breadth vs. vocabulary depth. After that, some useful ways of learning second language vocabulary are reviewed and discussed before providing an explanation of academic vocabulary, its importance to EAP learners, and its role within EAP learning programs. The following section highlights the topic of vocabulary and reading. The final section of the literature review is a discussion of online reading, why it is considered beneficial, its negative sides, and the factors affecting L2 online reading; such as the individual differences, the different instructional formats (integrated and split-attention formats), and the amount of text displayed on screen. A summary of the literature reviews at the end of the chapter explains the gap in research and the focus of this experiment.

The research methodology occupies chapter three. It highlights the research focus and research questions. It presents and discusses the methodological approach and research design best suited to answer the research questions set out in chapter 2.

Chapter four provides a presentation of the description, analysis and results of the quantitative data collected by the use of experimentation method. The chapter starts with an overview of the research topic and the research questions presented in Ch.2 before addressing these questions by stating and explaining the quantitative data findings. The findings and results of the study are stated and illustrated by tables and/or graphs.

Chapter five discusses the relationship between the research findings and published theory and research. This discussion includes the findings relating to the effect of online reading tasks on vocabulary knowledge, the effect of the different instructional formats of integrated and split-

attention on vocabulary knowledge, the effect of the proficiency level of the participants on their vocabulary knowledge, and the effect of the first language of the participants on their vocabulary knowledge.

Chapter six highlights the final conclusion of the study It includes the limitations observed during the research process, and discusses the implications of the study in relation to the research findings. To conclude, it makes suggestions for further study.

Chapter 2. Literature Review

2.1 Introduction

The following is the literature review in my thesis. It starts with a definition and explanation of what Cognitive Load Theory (CLT) is. A background of Cognitive Load Theory, in addition to Cognitive Theory of Multimedia Learning is presented. Following that is a review of working memory, how it is measured, and its role in first as well as second language acquisition. The next section gives an overview of vocabulary acquisition, a background of the teaching and learning of vocabulary, and then explains the nature of vocabulary learning and some of the aspects involved in learning a word; such as word meaning, word families and collocations. After that, some useful ways of learning second language vocabulary are reviewed and discussed before providing an explanation of academic vocabulary, its importance to EAP learners, and its role within EAP learning programs. The following section highlights the topic of vocabulary and reading. The final section of this literature review is a discussion of online reading and the factors affecting L2 online reading; such as the individual differences, the different instructional formats (integrated and split-attention formats), and the amount of text displayed on screen. A summary of the literature reviews at the end of the chapter explains the gap in research and the focus of this study.

2.2 Cognitive Load Theory (CLT)

2.2.1 Background about cognitive load theory

Cognitive load theory (CLT) was developed in the late 1980s out of a study of problem solving by John Sweller who argued that instructional design may be used to reduce cognitive load among learners. Sweller (1988) considers cognitive load theory to be an educational theory rooted in the domain of cognitive science. Similarly, it is considered by Sweller et al. (2019) and Sweller et al. (2011) as an educational psychological theory established on empirical information which can affect and support learning by providing the guidance, prescriptions and implications needed in any field. Pillay (1994) associates CLT with the mental effort needed for completing a task successfully; thus, higher cognitive load and mental effort are expected to lead to less successful completion of tasks.

Cognitive load theory has been a powerful and realistic interpretation of the human cognitive system and its construction. It therefore has been used by second language acquisition (SLA) researchers to explain any differences in learners' performance when they are exposed to different instructional formats. Various learning tasks have been found to have different effect on cognitive load and working memory, and this effect was shown to be emphasised in a multimedia environment where various interacting elements and tools can result in cognitive overload and decrease the learning results. For example, Cierniak et al. (2009) state that "the rationale of cognitive load theory is that the designs of instruction impose cognitive load on learners' limited working memory and that the cognitive load in turn influences learning outcomes". This presumes that when the cognitive load on the short-term memory is at its minimum level, learning can be most advantageous and the information learned can be remembered for longer time; however, if the difficulty of the instructional design is not taken into consideration properly, a cognitive overload occurs. Because cognitive overload is highly associated with lower performance on tasks, many researchers like Yeung et al. (1998), Al-Shehri and Gitsaki (2010), Hung (2009), Zumbach and Mohraz (2008) and DeStefano and LeFevre (2007) have tried to design instructional formats which can lead to less cognitive load.

Likewise, Cooper (1998) argues that CLT describes learning as a process which happens in connection with an information processing system encompassing long-term memory and short-term memory or working memory. The latter is responsible of organizing, contrasting and, comparing information during the processing procedure. According to Sweller et al. (1998), learners can only process two or three pieces of information simultaneously due to the very limited working memory that human beings have. This limitation, in addition to more information than can be handled , leads to less effective learning results (Paas et al., 2003) and (Sweller et al., 1998). Limited working memory (short-term memory) can only operate up to four elements simultaneously and only for a few seconds (van Merriënboer and Sweller, 2005). Consequently, when instructional materials are being designed, the restriction of working memory should be considered (Leahy and Sweller, 2004). CLT basically states that insufficient knowledge of the cognitive structure in human beings can lead to unsuccessful instructional design because the restrictions of the human cognitive structure and its effect on short-term memory is not thought about effectively by the designers of traditional instructional techniques (Schnotz and Kürschner, 2007).

Major information acquired from the unlimited long-term memory is processed by the short-term working memory according to the latter's capacity and limitations, and is stored in the long-term memory in different models called schemas (Kalyuga et al., 1997, Cooper, 1998,

Kalyuga et al., 1998, Leahy and Sweller, 2004). According to Hung (2009), language learners keep new vocabulary and structures in their long-term memory depending on the way they are going to be used. These models or schemas are moved from long-term memory to working memory as essential components of language which are going to be processed (Kalyuga et al., 1997) and (Paas et al., 2003). In addition, short-term working memory can hold more than one piece of information simultaneously by turning various elements of knowledge into one element, which determines the innate cognitive load (Gerjets et al., 2004).

A main learning process postulated by CLT is automation. Operation of information can occur either consciously in limited working memory or automatically with no influence from the working memory. Automation happens by practice, and it minimises the cognitive load on working memory. In other words, the automatic retrieval of models stored in long memory, also called automation of schemas, is believed to lead to reduced cognitive load. According to Leahy and Sweller (2004) and Sweller et al. (1998), automation can be noticed , for example, among advanced readers who do not need to consciously pay attention to the letters while reading because the letters are automated in childhood, whereas beginner readers or young children who are learning how to read need to notice each letter consciously in order to read the text. Input should be processed in working memory so that information can be built in the form of models or schemas in long-term memory, and this whole mass of mental activity forced on working memory is called cognitive load.

CLT divides cognitive load which has an effect on learning performance into three forms: intrinsic, extraneous, and germane.

Brunken et al. (2003) define intrinsic cognitive load as the cognitive load which is generated by the innate difficulty, instead of the instructional design of the subject; thus, learning foreign language syntax may be more challenging than learning the vocabulary of that language since building sentences involves comprehending the vocabulary which constitutes this sentence in addition to some grammatical features such as tenses and word order (Antonenko and Niederhauser, 2010).

In contrast to the intrinsic load, extraneous load is defined by Cierniak et al. (2009) as unneeded processing of information which is the result of instructional designs. This effect is named the split-attention effect (Yeung et al., 1998, Sweller and Chandler, 1991). An example of the split-attention effect is when reading texts in language classes are followed by most reading comprehension questions, and thus students need to move between the reading text and the

comprehension questions, trying to keep remembering the questions while looking for the answers (Hung, 2009). The way instructional design may result in the split-attention effect is clarified in many research studies such as Kalyuga et al. (1998), Yeung et al. (1998), Leahy and Sweller (2004), Hung (2009), and Al-Shehri and Gitsaki (2010) which declared that instructional formats of learning materials either enhanced or restricted the learning process. Several instructional formats, whether online-based or paper-based, cause high cognitive load on learners' working memory. Because intrinsic and extraneous load are extra burden on working memory (Paas et al., 2003), researchers recommend that the instructional format in any learning material should be designed cautiously, mainly when the innate level of difficulty in the content, or the intrinsic load, is high. In addition, Paas et al. (2010) state that working memory resources which are accessible for dealing with intrinsic and germane cognitive load is reduced by the growing extraneous load caused by certain instructional formats.

Germane cognitive load is the third type of cognitive load. Germane cognitive load is considered to be useful and productive during the learning process. Germane cognitive load helps learners be more aware during the learning process and is connected with oversufficient schema construction and automation. Burkes (2007) affirms that skills such as reading, which are habitually practised, can occur spontaneously without a lot of conscious effort by learners in spite of the possible complexity of the accompanying tasks. Cierniak et al. (2009) advise that "… an instructional design should reduce extraneous load (i.e., information processing hindering learning) and increase germane load (i.e., information processing supporting learning)". Despite the unalterable nature of the intrinsic load due to it being an essential part of the subject matter, extraneous and germane cognitive load may be highly affected by instructional formats of learning materials. Consequently, various amounts of memory load may be generated by the very same learning material depending on the different methods manipulated for presenting the material (Brunken et al., 2003).

2.2.2 Cognitive Theory of Multimedia Learning (CTML)

CLT has been a powerful and helpful clarification of human cognition, how this system is built, and how its different elements interact with each other. Studies built on CLT have researched various methods of reducing students' cognitive load in order to enhance their learning outcomes. Since CLT is concerned with verbal as well as visual representations of information, and since both verbal and visual information representations are integrated into multimedia, researchers have come up with a multimedia-based learning theory with effective theoretical and practical suggestions and implications. Cognitive Theory of Multimedia Learning (CTML)

can result in successful learning process by helping learners choose from multiple cognitive processes based on multimedia (Moreno and Valdez, 2005). Cognitive processes involve choosing relevant information, arranging the selected information into consistent representations, and combining these representations with existing knowledge. Whilst on one hand, the cognitive relationship between short-term working memory and long-term permanent memory can be enhanced by the designs based on CTML, on the other hand, Moreno and Valdez (2005) argue that "the processing demands required by interactive multimedia may exceed the processing capacity of the cognitive system and prevent learning".

2.3 Working Memory and its Role in Second Language Acquisition

2.3.1 Introduction

A very important factor that is believed to affect second language acquisition is working memory. According to Miyake and Friedman (1998), working memory can be an essential determiner of language aptitude, and it may be necessary in accounting for individual differences in ultimate achievement among adult L2 learners.

The following part is a discussion of background information on working memory (WM), how to measure it, and how it is involved in L1 and L2 acquisition.

2.3.2 Working memory

The term Working Memory (WM) refers to a system with restricted capacity in control of maintaining and manipulating information while the brain has to perform a cognitive task such as solving a problem, using mathematical skills, using reasoning and verbal skills, or even playing chess (Baddeley and Hitch, 1974, Daneman and Carpenter, 1980). WM was first proposed by Baddeley and Hitch (1974) as a temporary storage where received information could interact with long term memory and then stored long term.

Storage and processing constituents are taken into consideration by most recent research about working memory. Baddeley and Hitch brought the most effective model of working memory into use (Baddeley and Hitch, 1974). The model they introduced was built on a vast amount of experimental data collected from healthy adult participants as well as various samples of selective impairments from patients with damaged brains. The model is made up of three subsystems; the phonological loop (PL) and the visuospatial sketchpad (VSS) which are two parallel slave systems, in addition to the supervising central executive (CE). A more recent

development of Baddeley's model has a fourth constituent; namely the Episodic Buffer (Baddeley, 2000) (Figure 2-1).

Figure 2-1 Baddeley's model of Working Memory (2000, p.418)

The phonological loop is split into a storage system which contains characterizations of acoustic information until they decay and a rehearsal system which restores traces of decaying memory. It is presumed that the visuospatial sketchpad is similarly responsible for storing both visual and spatial information temporarily. The Central Executive is the system which directs, controls, and coordinates the working memory and all its subsystems. In the early stages it was believed that the Central Executive had a storage capacity, but the whole concept changed after the episodic buffer was introduced. The episodic buffer is presented as a multimodal storage system which relates various types of information from multiple sources and varied modalities.

There is a limitation in the capacity of the working memory system, which can lead to a restriction in the amount of resources or activation which can be assigned to storing as well as processing information; thus, when a certain task is more demanding, the working memory is more probable to be overloaded; thus to process information slower or even to lose it completely (Daneman and Carpenter, 1980).

2.3.3 How to measure working memory capacity

Span tasks have been used to measure the working memory capacity as part of a long-established practice. These tasks have various versions, the simplest of which are the digit span

as well as the word span. These two simple versions of the span tasks are simply about the number of digits or words which can be correctly repeated following oral presentation. In other words, these tasks basically test the amount of acoustic information which can be saved in memory for a short period of time; thus, the short term storage or phonological loop capacity.

Daneman and Carpenter (1980) later came up with a more complicated span task that is supposed to do more with the central executive component of working memory. In Daneman and Carpenter's reading span task, participants are exposed to a set of unrelated sentences. After finishing reading the sentences aloud, participants are supposed to repeat the last word of every sentence in a set. The number of sentences in a set where all the last words can be remembered and repeated defines the actual size of the reading span. As the same resources are needed while reading aloud, this process can challenge the articulatory repetition of the vocabulary which has to be remembered in the phonological loop. The concept which this task is built on is that the compromise between the functions of processing information and storing it in the working memory is reflected by the individual differences in test performance; thus, when the task has more demanding and complicated constituents, the information storage is expected to be poorer due to the fewer resources available to store information.

2.3.4 Working memory and first language

Involving working memory in language processing appears to be naturally very reasonable. Language processing needs continuous observing of a continuous flow of information. It requires both processing and storage of this information. According to Just and Carpenter (1992), to understand language one should be able to carry out complicated symbolic computations and to make sure intermediate and final products are maintained at the same time. A practically smaller capacity of storage might result in lack of comprehension, especially when consecutively encountered words, phrases and sentences are represented in context (Daneman and Carpenter, 1980).

The relationship between working memory capacity and language acquisition is explored in many studies. A relationship between performance on a version of Daneman and Carpenter's reading span task and various on-line and off-line measures of language acquisition was explored and found to be existent in many studies (Just and Carpenter, 1992, King and Just, 1991, Friederici et al., 1998, Vos, 1999).

According to King and Just (1991) and Just and Carpenter (1992), when task requirements are high, the processing of sentences is particularly restricted by working memory capacity;

however, when the task expectations are higher than the accessible working memory capacity, the information which has been stored and processed in the past will not be active anymore and new information will be stored instead, and this replacement of information can lead to slower or less accurate sentence processing (Daneman and Carpenter, 1980).

2.3.4.1 Is working memory a domain-specific or a general capacity?

According to Daneman and Merikle (1996), complex span measures, particularly those which highlight both processing and storage demands of working memory, such as the reading span task, are usually better indicators of language comprehension. Whereas simple span tasks can only focus on a storage component rather than focussing on the processing component, the reading span task studies the compromise between both components which is believed to be the reason behind differences in reading comprehension (Daneman and Carpenter, 1980); nevertheless, Ellis (2005) predicts that a reading span task which demands sentence processing can indicate performance on a sentence processing task, and it is not obvious whether this is an indication of domain-specific common ground between the two tasks, and whether a complex span task which involves non-verbal information processing would be valid when a sentence processing task was involved. Simply put, do the correlations between performance on the reading span test and measures of sentence comprehension indicate a more general capacity of the working memory or is the working memory capacity measured by the reading span particularly verbal? There is no agreement on whether working memory has a general capacity or a specific one. The following are some of the research findings regarding this debate.

On one hand, Just and Carpenter (1992) claim that the reading span task only measures working memory for language, and that the existing working memory capacity that "corresponds approximately to that part of the central executive in Baddeley's theory that deals with language comprehension" is a verbal working memory which is unspecialised. On the other hand, Waters and Caplan (1996) and Waters and Caplan (2004) discuss the verbal working memory system used in sentence comprehension. They suggest that the verbal working memory is a system which is divided up into various types of language processing resources that store and manipulate information simultaneously. They question whether the various working memory resources which are responsible for the syntactic processing in sentence comprehension are the same as the resources used in verbally mediated tasks which require conscious controlled processing. The evidence in their research is collected from different sources: the relationship between individual differences in working memory and individual differences in the efficiency of syntactic processing, the effect of concurrent verbal memory load on syntactic processing,

and syntactic processing in patients with poor short term memory, patients with poor working memory, and patients with aphasia. Experimental results from normal participants as well as participants with brain damage agree on the conclusion that the verbal working memory, which is specialized in assigning the syntactic structure of a sentence and using the structure in assigning meaning to the sentence, is different from the working memory system which is responsible for the use of sentence meaning to perform other tasks. For them, the results of research which find a correlation between reading span and measures of online language processing are questionable. Alternatively, they argue that the reading span or other working memory tests measure the capacity for the language comprehension which happens consciously and under control. This capacity, according to them, is relatively unrelated to the resource capacity which is necessary to the online language tasks which involve language processing and interpretation. On the contrary, Engle et al. (1999) argue that working memory is a general capacity rather than a domain-specific one. According to them, working memory reflects a "capability for controlled, sustained attention in the face of interference or distraction" ; thus, they propose a graded working memory system with a general capacity factor which includes many minor domain-specific factors where the general capacity factor can explain the different processing ways during the language tasks , whereas the domain-specific factors can be specific to the information storage during the language task. This is why it may be said that working memory tasks can have a stronger correlation with complicated tasks in a field matching the field of the stored information than with tasks in different fields.

Supporting this conclusion, Daneman and Carpenter (1980) found that "the predictive power of the working measure does not depend on a process component that taps the *language* comprehension processes" p.430. They concluded that although the working memory measures which purely tested verbal skills resulted in the most predictively valid ones, the complex span measures which tested processing and storing mathematical calculations and recalling numbers in order (Turner and Engle, 1989) were nearly as valid in speculating linguistic comprehension as the traditional reading span tasks.

From the previous review, it can be accepted that working memory has a verbal element but is this verbal element language specific and in what ways are the scores of L1 reading span tests related to L2 reading span tests? In other words, to what extent do L1 and L2 share the working memory resources (Miyake and Friedman, 1998)? After reviewing two studies comparing the scores of L1 and L2 reading span tests for learners of high linguistic proficiency, Miyake and Friedman (1998) found high correlations (r=.72 to .85) between those scores (Miyake and Friedman, 1998, Osaka and Osaka, 1992, Osaka et al., 1993). They also conducted their own

research to compare working memory in L1 and working memory in L2 acquisition among Japanese learners of English, and found a statistically significant correlation between working memory in L1 and L2 (r=.58). Moreover, a correlation between the scores of L1 and L2 reading span tests for a mixed group of Chinese, Japanese and Spanish learners of English (r=.62) is reported by Juffs (2004). These findings show that although it cannot be confirmed yet what the correlations may be for L2 learners with lower linguistic proficiency level, it can be said that the reading span tasks in both L1 and L2 are highly related among L2 learners with intermediate or higher linguistic proficiency level.

2.3.5 Working memory and second language

It is believed that when language acquisition starts after childhood, it is much more difficult for language learners to achieve high levels of linguistic proficiency; therefore, language attainment levels can vary a lot amongst adult second language learners. Although it is known that some of the common factors which affect second language acquisition are motivation, attitude and personality, these individual differences can be highly affected by many cognition-related variables among adult second language learners. According to Hyltenstam and Abrahamsson (2004), and to DeKeyser (2000), in order to achieve high levels of L2 proficiency which are native-like, L2 learners should have high levels of language aptitude, which is a talent for language learning and acquisition that includes various constituents and that can be measured by an aptitude battery. One of the essential factors which are thought to affect the nature of language aptitude is working memory (Miyake and Friedman, 1998, Ellis, 2005) and it is expected that part of the difference in the ultimate level of language attainment can be accounted for by the individual differences in working memory. Additionally, Miyake and Friedman (1998) propose that systems like working memory can action comparatively more efficiently in second language acquisition than in first language acquisition because second language acquisition in adults can depend on conscious systems of learning in general to a greater degree than child language acquisition does. However, can working memory work efficiently as a predictor for L2 proficiency as it is for L1 proficiency? And what is its role in processing L2 during the learning process?

2.3.5.1 Can working memory work efficiently as a predictor for L2 proficiency ?

It is believed that working memory capacity can be associated with the effectiveness of processing a second language only because high working memory span is associated with quicker and more precise processing of sentences in learners' first language. Despite the little amount of research which has been conducted to study the relationship between working

memory and processing of L2, and although most of the studies so far support the prediction that working memory capacity may be related to the efficiency with which a second language can be processed, some methodological and theoretical issues in the studies remain unresolved or other factors, such as individual differences among the learners, may affect the role that working memory plays in second language learning (Miyake and Friedman, 1998, Harrington and Sawyer, 1992, Harrington, 1991, Geva and Ryan, 1993).

For example, in their study of 32 advanced Japanese learners of English as a second language, Harrington and Sawyer (1992) conducted a reading span task in both first and second languages in addition to a simple word or digit span in second language. They used the grammar as well as the reading and vocabulary sections of the Test of English as a Foreign Language (TOEFL) to measure the English language proficiency among learners. In line with the results for processing of first language, the results showed that both simple span tasks had weak correlation with the L2 reading measures, with correlation coefficient being less than .25. The L2 reading span scores worked well as predictors for performance on both TOEFL sections; namely, grammar section and reading and vocabulary section (.57 with the grammar section, and .54 with the reading and vocabulary section. The results of this study support the prediction that working memory capacity plays a role in second language learning, but its role depends on other factors like the age of L2 learners, the task, and the linguistic domain. A statistically significant correlation (r=.52) between L2 working memory capacity as measured by reading span tasks and L2 syntactic comprehension ability among Japanese learners of English was also found by Miyake and Friedman (1998); however, the results of this study also suggest that the individual differences are another factor affecting L2 comprehension.

Some research have attempted to detach the effect of working memory on language proficiency from other variables which may affect the correlation between the two. In addition to a reading span task, Harrington (1991) conducted tests of second language vocabulary and grammar acquisition among adult Japanese learners of English. The study showed that although the factor of measuring vocabulary and grammatical knowledge was removed, the correlation between working memory and L2 reading comprehension (r=.51) stayed statistically significant; however, although the preliminary data proposes that the working memory capacity plays an important role in highlighting the differences in L2 development, some theoretical and methodological issues need to be reconsidered.

In their comprehensive test battery to English school children (aged 10-12) learning Hebrew, Geva and Ryan (1993) tried to have a clearer idea of the relationship between L1 and L2 skills

and cognitive abilities. Their battery of tests involved tasks testing intelligence, tasks testing working memory capacity in L1 and L2, in addition to various tests to measure L1 and L2 linguistic proficiency. Findings of the study showed a statistically significant correlation between L2 measures and L2 working memory capacity (r=.52), and even when the variable shared with age and intelligence was separated out, the correlation stayed statistically significant; nevertheless, there was no correlation between L1 working memory capacity and L1 linguistic proficiency after removing the common variance shared with intelligence and age, so this absence of correlation could be affected by the intelligence factor. The findings of Geva and Ryan (1993) are supported by Miyake and Friedman (1998)'s results which show that working memory is more probable to play an effective role in L2 than in L1 acquisition.

Although all the research cited above indicates that working memory can be an essential predictor for successful L2 comprehension, different findings were revealed by Juffs (2004)'s research which focussed on the effect of working memory on on-line L2 processing among speakers of Chinese, Japanese and Spanish as first languages. Juffs used a self-paced reading task to measure language processing online. Learners were required to read complex garden-path sentences which are grammatically correct, but they start in a way which can mislead the readers and make them interpret the sentences incorrectly. Learners read the sentences word by word as part of the test, and no statistically significant correlation was found between word or reading span and reading times when the processing load was at its most at certain stages in the sentences. Learners in Juffs' study were divided into groups according to high or low span, and it was found that while word span was related to L2 reading times, reading span was not. However, Juffs states that the findings of this study should be looked at carefully because firstly, the study is believed to be the first one which researches the relationship between working memory and L2 processing online; secondly, there was not enough evidence for the role of working memory in the online L2 processing task. As a result, the findings of the study can be accounted for using Waters and Caplan (1996)'s suggestion that working memory resources are separate from the resources needed for language processing online. Moreover, it is hard to decide how reliable the reading span tasks and measuring tools used in the above-mentioned research are, thus more research is needed to investigate the role of working memory in online language acquisition.

2.3.5.2 What is the role of working memory in second language processing during the learning process?

According to the research reviewed above, there is a consensus in research that larger working memory capacity is more effective when learning a second language; however, are there reasons why the beneficial role of working memory may be thought to change during the learning process? To answer this question, the first thing to think about is focussing on some details of the language processes at different stages.

Built on Anderson (1983)'s ACT model of language acquisition, Towell and Hawkins (1994) proposed an L2 acquisition model which suggested that working memory interrelates with two different types of long-term memory. These two types are the declarative memory and the proceduralized memory. Whereas the declarative memory is responsible for storing declarative "knowing that" knowledge such as grammar and rules, the proceduralized memory is responsible for storing procedural "knowing how" knowledge. At the beginning, the declarative part of long term memory stores all knowledge. Throughout the learning process, further declarative knowledge will gradually be changed into proceduralized knowledge and stored in proceduralized memory. Working memory or short-term memory always works as the path for the input and output of information. Whereas the interaction between declarative memory and working memory, which primarily occurs in the first stages of the learning process, is slow and demanding, interaction between procedural memory and working memory is quick and effective.

Ellis (2005) emphasizes the concept of automatization of explicit knowledge and his suggestions are consistent to a large extent with the concept of moving from declarative to procedural knowledge. Ellis (2005) briefly argues that explicit learning (explicit deduction, hypothesis formulation) in the first place has a more essential role and makes it easier for the implicit learning (unconscious learning occurring during the use of language and the exposure to it) to happen. The procedures of conscious, explicit learning all happen within the working memory has a powerful control over them. This information leads to the belief that working memory has a more powerful role in the early phase of the language acquisition process, where learners depend more on explicit, declarative knowledge which is extremely controlled by short-term memory. At these early stages, learners cannot yet use grammatical knowledge, and their working memory is highly likely to become constantly overloaded, resulting in slow and difficult analysis of information. In more advanced stages of the learning process, when knowledge becomes proceduralized, information processing can become more automatic and

probably more independent from working memory, because controlled processing is more challenging for working memory than automatic processing. However, this does not mean that working memory is not expected to play any role in the advanced stages of the learning process, since it has been found that working memory can have an efficient role in L1 acquisition, which is potentially automatized and procedural (Just and Carpenter, 1992, King and Just, 1991, Friederici et al., 1998, Vos, 1999); thus, it can be predicted that the early stages of the language learning process may require a large working memory capacity, whereas later on during the learning process when information becomes automated, working memory may start to be helpful the same way it is during first language acquisition.

2.4 Vocabulary Acquisition

2.4.1 Overview

Tozcu and Coady (2004) argue that there is a universal model of vocabulary recognition that L1 and L2 reading share, and that this model is used by native and non-native speakers while they are reading. In other words, effective vocabulary acquisition during reading depends mainly on how often certain words are used. Tozcu and Coady (2004) also argue that "less frequent vocabulary items can be learned through incidental encounters in context (with some strategy training as an aid) via extensive reading, but only after automaticity in high-frequency or core vocabulary has been acquired" (op. cit.: 480). Nation (2022) states that frequency and range of occurrence are very beneficial to be used in order to differentiate various levels of vocabulary. Telling the difference between these levels can lead to learning vocabulary in the most beneficial order and thus receive the best advantages from the words they learn. In order to make sure that the teachers and learners handle vocabulary most efficiently, it is helpful to make distinctions among high-frequency, mid-frequency, and low-frequency words where the 3000 high-frequency words should be taught first, and then let the mid-frequency and low-frequency words be learned through extensive exposure alongside using some vocabulary learning strategies (Nation, 2022). These figures are obtained from the (British National Corpus (BNC)). Accordingly, repetition and high-frequency of vocabulary use in multimedia-based online reading materials can have the potential of increasing cognitive load, especially for non-native speakers who are trying to translate while they are processing information (Tozcu and Coady, 2004). On the contrary, Constantinescu (2007) concluded that the use of multimedia can be effective for remembering new vocabulary and making reading comprehension easier.

Similarly, on-screen reading, according to Tozcu and Coady (2004), can increase the readers' vocabulary acquisition and their reading comprehension to a great extent.

Learning a word involves much more than learning only its meaning, spelling, and pronunciation (Schmitt, 2000). As a matter of fact, knowing words has more implications. Words collocations, grammatical features, register, frequency, and associations are more aspects to learn when learning a word. According to Schmitt and Schmitt (2011), learners should master all of the above mentioned word knowledge aspects in order to feel confident using the words. Whereas the meaning and spelling aspects of words can be taught explicitly, other aspects like frequency and register information are only possible to be acquired by exposure to the words.

2.4.2 A background to the teaching and learning of vocabulary

Vocabulary has been an ignored side of language teaching and learning over the years despite its being an essential consideration of the language learner (Zimmerman, 1997). Some methods like grammar-translation method which was introduced at the end of the eighteenth century or audio lingual methodology which was developed during World War II viewed vocabulary as playing a supportive rather than central role in language teaching and learning (Zimmerman, 1997). However, in more recent years vocabulary has become recognised as an important component of language use and as a key element of teaching methodology; for example, communicative methodology emphasizes on the potential of vocabulary knowledge in fluency while Krashen's Input Hypothesis (Krashen, 1981) focusses attention on the role of comprehensible input, of which vocabulary comprehension constitutes a main part, rather than correct output (Zimmerman, 1997). According to Krashen and Terrell (1983), vocabulary is very important in second language (L2) settings because of the important role it plays in classroom success. Moreover, Henry Sweet, a nineteenth century English linguist, observes that "the real intrinsic difficulty in learning a foreign language lies in that of having to master its vocabulary" (Sweet, 1899, 1964). Finally, while Wilkins (1972) states that although very little can be expressed without grammar, nothing can be expressed without vocabulary, there is a tendency now to shift from regarding language production as a rule-based activity towards regarding it as a process involving retrieval of chunks of language from memory. This tendency is exemplified by Lewis's argument that "language consists of grammaticalised lexis, not lexicalised grammar" (Zimmerman, 1997).

2.4.3 What do second language vocabulary learners learn? The nature of vocabulary learning

2.4.3.1 Word meaning

Commonly, knowing a word is considered by learners as knowing its form whether it is spoken or written, in addition to knowing its meaning. Nation (2022) argues that it is not enough for learners to know the form and meaning of a certain word separately because the essence of world knowledge is the ability to connect both the form and the meaning of the word. An example given by Nation (2022) is the word *brunch* where the learner may be aware of the form itself or aware of the concept of a meal which replaces both breakfast and lunch; however, learning the word involves relating the form *brunch* to the concept of the meal which replaces breakfast and lunch. The stronger the connection between the word form and meaning knowledge, the more ready the learner is to recollect the word form while wanting to express the meaning, and to recollect the word meaning as soon as they see or hear the word form. Baddeley (2013) proposes that each time a learner retrieves the form or meaning of a certain word, the connection between the two is strengthened, so it is required that learners not only see the form and meaning of the word together, but also have multiple occasions of retrieving them.

It is believed that it is easier for learning vocabulary in a second language to relate the form to the meaning if almost the same form in the first language is related to almost the full meaning; thus, the learning load of relating the form to the meaning is light if the word which is being learnt is a cognate or a loanword which exists in both the first and the second languages. Learning can be made a lot easier when there are loanwords between languages.

According to Hall (2002), words which have the form of cognates with the L1 are dealt with as cognates by assigning them the L1 meaning in addition to the form. Hall's participants were Spanish learners of English, but because of the considerable numbers of loanwords used in languages that are totally different to English, there is a possibility that this effect is not associated with the close connection amongst languages.

The huge number of English loanwords in Japanese was noted by Daulton (1998) where up to 38% of the 2000 most frequent words of English were borrowed into Japanese. In addition, there is a borrowing of 26% in the University Word List. Studies in addition to Daulton's which focus on Japanese point out that loanwords facilitate the learning of the English language although sometimes these loanwords have limited meaning in Japanese and learners may need

to extend this meaning. Examples of these loanwords in Japanese are *paati* meaning party, *piano* meaning piano, *booru* meaning bell, *Waarudo Shiriizu* meaning World Series. Very expectedly, a lot of research has found that learning cognates has positive influence on learning a language. Tonzar et al. (2009) argue that cognates in three languages are perceived quicker than those in two languages which the learners know.

A different way to connect form to meaning easily is to have a first language link between the second language word form and the meaning. This is the main concept of what is called the keyword technique (Nation, 2022). In the keyword technique, the form of a word in a second language is connected to the form of a similarly pronounced/written word in the first language but has a different meaning. The learner then tries to link the meaning of both words in both languages and imagine a relationship between them. An example given by Nation is the word *cow* in English. A Thai learner of English tries to think of a similar word in Thai, like the word *khaw* meaning 'rice', and then thinks of a cow eating rice in order to relate the two words in first and second languages.

Webb (2007) made a comparison of learning words that had high-frequency L2 synonyms, like *locomotive* and *train* with learning words that did not, like *lick*. The process of learning used pairs of L1-L2 pairs' and the process of testing used 10 tests for each word focussing on a variety of productive and receptive aspects of word knowledge, such as: meaning in form, associations, orthography and grammar. Whereas the results showed significant differences for paradigmatic association, syntagmatic and orthography, very little differences were noticed across most aspects whether there was or was not a sentence to provide context.

2.4.3.1.1 Receptive vs. productive knowledge

Nation (1990) distinguishes between receptive and productive vocabulary learning. Whereas receptive vocabulary learning, according to Nation (1990) entails the understanding and recall of words when learners meet them, productive learning entails receptive learning in addition to the ability to produce the words orally or in writing when necessary. The question that can be asked here is "How does receptive learning entail productive learning?" Nation's answer is that "real vocabulary learning comes through actual productive and receptive use (p. 6). Thus, moving vocabulary knowledge from the receptive to the productive learning stage really comes through use of the vocabulary.

A considerable body of research clarifies that learners usually builds receptive mastery of vocabulary before productive mastery, although this cannot be applied to each and every item.

This is clarified by studies which have compared the number of words known receptively versus productively. Laufer (2005a) on one hand, compared learners' productive test scores on L1-L2 recall tests as a percentage of their receptive test scores on L2-L1 translation tests. She found productive/receptive ratios ranging from 16% at the 5000 frequency level to 35% at the 2000 level. Fan (2000), on the other hand, found a range from 53% to 81% (mean 69.2%) for words taken from the 2000, 3000, and UWL levels. Whereas the ratios depend extremely on the types of receptive /productive tests used (Laufer and Goldstein, 2004), L2 learners' receptive lexicons seem to be larger than their productive lexicons.

Moving from the no knowledge stage through the receptive towards the productive mastery stage is really crucial for the process of developing one's vocabulary knowledge. Despite the considerable body of research which supported the idea of the productive mastery being a later stage after the receptive mastery, the way this process proceeds is not clearly defined. According to Melka (1997), the relationship between the two as a continuum where gradually progressing vocabulary knowledge helps learners move from the receptive to the productive stage. Contrastingly, it is argued by Meara (1990) that both receptive and productive mastery of vocabulary might differ quantifiably, possibly depending on an item's status within the lexical network. The receptive and the productive mastery stages are not easily measured. The relationship between the two is hugely dependent on the measurement tools used (Waring, 1999).

2.4.3.1.2 Vocabulary breadth vs. vocabulary depth

Milton and Donzelli (2013) draw the contrast between three approaches to language learning: approaches that hypothesize that learners learn and store different words separately, approaches that hypothesize that learners acquire words as lemmas which are base forms that can be regularly inflected, as in Pienemann's processability theory (Pienemann, 1980, Pienemann, 1984, Pienemann, 1985), and approaches that take the language learning process further and analyse words as word families. If L2 learners learn and store each word separately, mastering a language requires several hundred thousand words. Seashore and Eckerson (1940) followed this approach and the vocabulary size of native speakers was estimated to be about 200,000 or 250,000 words, which is a frightening task for L2 learners. However, if words are treated as lemmas or words families, achieving a native-like fluency is measured using a totally different scale in terms of the size of vocabulary. Figures of 17,000 to 20,000 words were suggested by Goulden et al. (1990) who followed the word family approach to measure the vocabulary size of educated native speakers of English. Much less size of 9,000 words was suggested by Milton

(2009) among 18 year-olds entering university. Obviously, these suggestions make a much more applicable task to L2 learners than the hundreds of thousands of words suggested by earlier studies. The suggested vocabulary size in these studies is not the only factor in deciding the quality of L2 learners' lexicons since some learners may know a lot of words but still can't use them fluently and at the same time there are some learners who know few words but can use them in the right way. Meara and Wolter (2004) note that learners may have similar vocabulary sizes but still differ in the organisation degrees of their lexicon.

The above-mentioned three approaches to language learning focus only on linking the form of a word to its meaning, whereas learning a word involves more than this. Nation's list of what is involved in knowing a word (Nation, 2001) demonstrates that knowing a word involves: first, an understanding of the concepts, referents and associations related to the word, especially if they differ from the words in the learner's native language; second, knowledge of its grammatical functions and its use restrictions.

Nation's list is given in Table 2-1.

Table 2-1. What is involved in knowing a word (from (Nation, 2001))

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable 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?

Note: In column 3, R = receptive knowledge, P = productive knowledge.

However, while Nation's list of what is involved in knowing a word is quite complicated and multi-dimensional, Daller et al. (2007) think of word knowledge in terms of three dimensions:

- Lexical breadth: the number of words that learners know regardless of how well they know them.
- Lexical depth: how well the learners know the words they know.
- Fluency: the speed and ease of using words in speech or writing.

Daller et al. (2007) classification of the three dimensions of word knowledge is more efficient than any of the three approaches to language learning which only link the form of a word to its meaning. So lexical breadth, lexical depth and fluency should be looked at independently when measuring vocabulary knowledge. However, it should be taken into consideration that in order for learners to use their words fluently, or at least reasonably, they should have a large vocabulary to select from. Milton and Donzelli (2013) declare that "Depth is the end of the breadth dimension since choices in the vocabulary to be used can only realistically be made when the learner has a large vocabulary from which to choose". This idea is also confirmed by Vermeer (2001) considers the three dimensions to be only one since both the degree of depth and the ability to use words fluently require large lexicons. Other studies also showed high correlation between lexical breadth and lexical depth. An example is Gyllstad (2007) study which showed correlation of 0.93 between breadth and depth.

To conclude, acquiring a second language lexicon requires L2 learners to acquire a large number of words and be able to use them in a sophisticated way. This leads to the second aspect of building vocabulary knowledge discussed in this study in addition to word meaning knowledge, which is word family knowledge. This model of analysing vocabulary knowledge and researching it focussing on the three elements of: word meaning, word families, and word collocations is built on Schmitt and Schmitt (2011) analysis.

2.4.3.2 Word families

It has been shown in research that it is necessary for learners to know around 6000-7000 word families in order to understand spoken language in various situations. In addition, 8000-9000 word families are required so that learners can read a variety of authentic reading texts, such as magazines, newspapers, or novels (Nation, 2006). It is still possible for learners to understand verbal or written language with a smaller number of words, but going through so much unknown vocabulary may be an issue for them (Schmitt et al., 2011a). The numbers mentioned above can sound scary, but they are not an exact representation of how much learners need to know. Every word family involves many separate word forms; the root form, its inflections, and its regular derivations. Nation (2006) suggests that the most frequent word families have

an average of six items per family; and these figures can be reduced to around three items per family at the 9000 frequency level. In line with Nation's suggestion, whereas the amount of vocabulary necessary for listening, which is around 6000 word families, implies knowing 28015 separate word forms, the amount of vocabulary necessary for reading, which is 8000 word families, implies knowing 34660 separate words (Schmitt, 2008). Nevertheless, knowing a separate word form does not necessarily mean that the other word forms or family members are known or can be guessed by the learner. According to a study by Schmitt and Zimmerman (2002), generally speaking, English learners of advanced linguistic levels did not know all of the noun, verb, adjective, and adverb members of the word families they studied. According to a survey of some contexts of international teaching by Laufer (2000), it was observed that after 1800-2400 hours of teaching, several university students knew about 4000 word families, but most learners she surveyed knew only 1000-2000 word families. These learners hardly had any sufficient vocabulary sizes compared to the figures mentioned above since they either knew so little or had no knowledge at all of the most recurrent English word families. Although learners may know a few forms in a word family, they are not expected to know all of the forms of each word they have learnt. Nonetheless, learners are expected to know the right form of a word in its context (Schmitt and Zimmerman, 2002).

2.4.3.3 Collocations

"Knowing a word involves knowing what words it typically occurs with" (Nation, 2022). Pawley and Syder (1983) suggest that storing a lot of word sequences memorised in the learners' brains is the reason behind speaking the first languages fluently and helping learners choose word sequences which make them use vocabulary similarly to native speakers. By knowing these ready-formed clusters, learners can avoid assembling words every time they are using the language. There are patterns of collocations, and familiarity with these patterns may reduce the learning load on the learners' brains. Similarity between collocation patterns in first and second languages can make the learning process easier (Nation, 2022). Collocation is one of many connection forms which have to do with the proper reception and production of vocabulary. Miller (1999) states that knowing a word mainly involves a cognitive depiction of the group of contexts in which the word form can be employed to convey meaning. This awareness of context may refer, according to Nation (2022), to situational context, topical context, and local context. Collocation is mainly information of local context expressed by words occurring next to other words. Comprehension of the topic which is being written or spoken about is referred to as topical context. Situational information implies general knowledge in addition to comprehension of the specific context of the communication.

2.5 What is the Best Way to Learn Vocabulary? How Do L2 Learners Succeed to Have Extensive Lexical Knowledge?

2.5.1 Terminology

Before discussing the ways second language learners develop their lexical knowledge, some terminological issues will be highlighted. First, the words "vocabulary" and "lexicon" are used interchangeably in this research. Second, the term "extensive vocabulary knowledge" is used interchangeably with "large vocabulary knowledge"; and it should be distinguished from the term "extensive reading" which will be defined later in this research.

2.5.2 Overview

In their review, Milton and Donzelli (2013) state that even the foreign language learners who are receiving their education under good learning circumstances do not seem to be adding more than five words to their lexicon per classroom hour, thus classroom hours are limited compared to the large size of the lexicon of foreign language learners, which has led some researchers, such as Harris and Snow (2004) to claim that learners retain only few words of those taught or learned by direct instruction, and to suggest that learners "extend their vocabulary through sub-conscious acquisition" (Harris and Snow, 2004). Milton and Donzelli (2013) state that there is a debate "as to whether vocabulary is best learned incidentally or is best explicitly taught". Why do Milton and Donzelli contrast "incidental" and "explicit" vocabulary learning? And would it be more consistent to contrast: on one hand "incidental" vs. "intentional" vocabulary learning, and on the other hand "implicit acquisition" vs. "explicit acquisition"? It could be argued that a learner's encounter with a new word is incidental, but the following engagement can be conscious. Similarly, a learner could consciously engage with a new word which has been incidentally encountered without having an explicit learning intention (Ellis, 1994a). Incidental learning can be what he calls non-explicit, but it is not necessarily implicit. Is the incidental-intentional dichotomy therefore false and is the implicit-explicit analysis of vocabulary learning theoretically and practically stronger?

An important issue is defining the two separated, but related, concepts: incidental learning in language pedagogy, and implicit learning in psychology. They are superficially similar since they both occur unconsciously (Reider, 2003), yet they are separate. Incidental learning is defined as picking up new words from context with no intention of doing so (Barcroft, 2009); and with respect to vocabulary, it is "the learning of vocabulary as the by-product of any activity

not explicitly geared to vocabulary learning"; however, it does not account for the actual process of acquisition of knowledge. The more careful concept of implicit acquisition, which is defined by Ellis (1994b) as the "acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply and without conscious operation", can account for the learning process more accurately.

The same argument can apply to intentional learning in language pedagogy, and explicit learning in psychology. They are superficially similar since they both occur consciously, yet they are separate. Pedagogically speaking, intentional learning, as its name tells, is intentional. Hulstijn (2001) defines intentional vocabulary learning as "any activity geared at committing lexical information to memory"; however, it does not account for the psychological concept of acquiring information purposefully, whereas the concept of explicit acquisition, which is "a more conscious operation where the individual makes and tests hypotheses in a search for structure" (Ellis, 1994b), accounts for the actual process of acquiring information consciously.

2.5.3 Implicit vs. explicit L2 vocabulary learning

The previous paragraph includes definition of implicit and explicit learning in general. This paragraph reflects the above-mentioned definitions on L2 vocabulary learning in specific. Ellis (1994c) distinguishes two possibilities in relation to the processing of new vocabulary, the implicit vocabulary learning hypothesis and the explicit vocabulary learning hypothesis. The former is related to behaviourist approaches, and argues that new vocabulary is acquired without the language learners being aware of it, especially when the learners are reading or interacting orally. The main feature of this hypothesis is the lack of intentionality and the unconsciousness.

> *An implicit vocabulary learning hypothesis* would hold that the meaning of a new word is acquired totally unconsciously as a result of abstraction from repeated exposures in a range of activated contexts. (Ellis, 1994c).

His second hypothesis predicts that language learners follow a number of conscious and planned strategies in which they pay explicit attention to new words.

> *An explicit vocabulary learning hypothesis* would hold that there is some benefit to vocabulary acquisition from the learner noticing novel vocabulary, selective attending to it, and using a variety of strategies to try to infer its meaning from

> the context. Furthermore there may also be advantage from applying metacognitive strategies to remember new vocabulary, to consolidate a new understanding by repetition [...]. (Ellis, 1994c).

The two above-mentioned hypotheses demonstrate opposing views of vocabulary learning. Building his arguments on a large body of experimental psycholinguistic research in the fields of vocabulary and intelligence, implicit memory and global amnesia, Ellis claims that incidental vocabulary acquisition involves both implicit and explicit learning processes: whereas implicit processes are involved in the acquisition of words' forms, collocations and grammatical class information, explicit learning processes are thought to be involved in acquiring words' semantic properties and mapping word form to meaning. Moreover, while Ellis argues for a complete separation of implicit (i.e. formal) aspects from explicit (i.e. semantic) aspects of vocabulary acquisition, more recent research would argue that for second language learners at least, a combination of both implicit and explicit L2 vocabulary learning is necessary and should be considered complementary (Schmitt, 2000). Singleton (1999) for example criticises Ellis' idea of separated processes, arguing that even if learning forms and meanings of unknown words begin by different mechanisms, this does not inevitably mean that they are manipulated separately at all stages. Alternatively, Singleton would argue that there is a possible interaction between implicit and explicit systems. Borner (1997) in turn emphasizes the necessity of an alteration and differentiation of Ellis' model in a way that combines different degrees of explicitness and allows for both explicit and implicit learning of form features.

2.5.4 Inductive vs. deductive L2 vocabulary learning

Whether an inductive or deductive approach is used in teaching or learning a word is very much dependent on the degree of difficulty of defining the word (Nation, 1990). Nation (1990) argues that it is best to teach words which are difficult to define inductively by providing many examples of sentences including the words. This gives the students the opportunity to repeat the words and allows them to work harder towards learning the words. As for the words which are easy to define, they can be quickly explained deductively before doing some practice with them and allowing the students to repeat them making their own examples.

2.5.5 Incidental vs. intentional L2 vocabulary learning

Huckin and Coady (1999) state that most L2 vocabulary scholars seem to agree that apart from the first few thousand most common words, L2 vocabulary is mainly acquired through extensive reading when the learner guesses the meaning of most of the new words (for example (Carter and McCarthy, 1988, Coady and Huckin, 1997, Huckin et al., 1993, Lewis, 1993, McCarthy, 1990, Nation and Carter, 1989, Schmitt and McCarthy, 1997, Gairns and Redman, 1986, Nation, 1990)). Since this learning of new words is a secondary type of learning, it is said that it is happening incidentally. Huckin and Coady (1999) consider that incidental L2 vocabulary acquisition through reading has the following advantages over intentional L2 vocabulary learning:

a. Words are contextualized: Contrary to the traditional paired-associate exercises where the learner is required to learn to associate one syllable or word with another building on the stimulus and response similarity, incidental L2 vocabulary through reading exposes the learner to the word itself in addition to its meaning in a context. Similarly, Ooi and Lee Kim-Seoh (1996) observe that "vocabulary learned through reading would give the learner more opportunities to process language use at a deeper level and to develop semantic networks and other kinds of associative links that will ultimately enhance learning" (Ooi and Lee Kim-Seoh, 1996). So if the learner comprehends the words and relates them to his or her goals or ideas, it will be easier to retain those words (Rivers, 1983), thus learners' engagement in the words they are learning is very important. Gairns and Redman (1986) argue that it is better for learners to be stimulated to contribute items they are willing to learn. According to McKeown et al. (1985) and Channell (1981), the learning process would be easier if learners manage to build semantic networks around the words they want to learn. Moreover, it is very important according to Stahl (1983) for the effective learning to occur that the learners relate the new information to the information they already know, and for this to happen, a deep-level processing is required.

b. It is more efficient in terms of pedagogy: it encourages reading while new vocabulary is learnt and previous words are fixed due to their being contextualized.

c. It is more independent and learner-based because the reading materials being used are mainly selected by the learners themselves; thus, it encourages autonomy in language learning. Qian (2004) shows that the strategy of guessing

> new vocabulary from context is one of the most common strategies when L2 learners are exposed to unknown words. According to Qian (2004), guessing, from context, is even more frequent than dictionary use; thus, learning vocabulary through guessing new words in context encourages learning at the students' pace. However, Huckin and Coady's view about learners' choice of the reading materials is not necessarily true all the time. Incidental vocabulary learning can occur even if the learner does not choose the material he or she has chosen themselves.

However, there are still some problems concerning the operational definition of L2 incidental vocabulary acquisition. One of those problems is the suggestion that incidental vocabulary acquisition occurs unconsciously (Reider, 2003). According to Gass (1999), considering incidental vocabulary acquisition to happen unconsciously as the "side-effect" of another activity underestimates the active role of the learner in this process. Huckin and Coady (1999) talk about secondary learning and define the incidental vocabulary learning as the "by-product, not the target, of the main cognitive activity, reading" (p. 182). In other words, second vocabulary learning is subordinate to the major pedagogical activity. According to Huckin and Coady (1999)'s definition, the aimed focus of learning is somewhere else. However, this definition, according to Gass (1999), is inclined to ignore the active role of the learner and although it may be helpful for studying incidental vocabulary acquisition, it implicitly proposes the idea that "outside a classroom context, there is only incidental learning, and this is clearly not what one eventually wants to say" (p. 321). Ellis (1994a) describes the process of incidental vocabulary acquisition as non-explicit since it does not involve an explicit learning intention, but at the same time it cannot be considered as implicit since it is not completely non-conscious.

2.6 Academic Vocabulary

The needs of English as a second language or foreign language learners preparing to study in higher education, in addition to English for Academic Purposes (EAP) learners, have been the main focus of research investigating academic vocabulary (Coxhead, 2020). In Beck et al. (2013)'s three-tiered model of vocabulary, academic vocabulary are defined as the common words which are used among mature speakers and which are less frequent in conversations but more frequent in texts. Some examples of academic vocabulary are: *concerned, forest, trunk, dependant,* and *survive*. Academic vocabulary in Beck et al. (2013)'s model sit between Tier 1 words, also called conversational words, and Tier 3 words, also called subject-specific words.

2.6.1 The importance of academic vocabulary to EAP learners

EAP learners need academic vocabulary for many reasons. First, academic words are the main element of written academic texts. 10% (Coxhead, 2000) to 14% (Gardner and Davies, 2014) of words in academic texts are composed of academic vocabulary. It is suggested by these figures that in each line of a written academic text, one word in ten or one word in seven may be academic. In order for EAP learners to study in an academic institution teaching in English, a large amount of reading is required. EAP learners need a considerable number of vocabulary to understand the academic texts they read. To reach 98% coverage of academic written texts, learners need 8000 to 9000 word families in addition to proper nouns (Nation, 2006). Hu and Nation (2000); Schmitt et al. (2011b) argue that this level of lexical coverage is enough for comprehending written texts. Moreover, understanding academic spoken texts need a vast knowledge of vocabulary. For example, 4000 word families in addition to proper nouns and marginal words are required in order for learners to reach 95% understanding of lectures, whereas reaching 98% understanding requires up to 8000 word families in addition to proper nouns and marginal words (Dang and Webb, 2014). In order to comprehend spoken discourse, 95% lexical coverage is believed to be enough by van Zeeland and Schmitt (2013), but the number of listeners who can properly understand spoken language may become larger with coverage exceeding 95%. Academic word lists are believed to be of great importance to EAP learners' reading and listening because they can establish the basis to learning the same vocabulary EAP learners may come across during listening to or reading their academic materials. For example, learning the 1741 words of the Academic Spoken Word List (Dang et al., 2017) can help learners to understand 92% to 93% of the vocabulary in academic speech, and this number exceeds the coverage they can reach by learning the most frequent 2000 words of general vocabulary which is 91%.

It is suggested by vocabulary testing research that a lot of learners of English as a second language have low levels of vocabulary knowledge and slow rates of vocabulary growth over a wide range of contexts, such as Denmark (Henriksen and Danelund, 2015) and Taiwan (Webb and Chang, 2012). For this reason, EAP learners studying at institutions which used the English language for teaching need to enhance their vocabulary knowledge in general, but they need to specifically build their academic vocabulary knowledge because academic words occur in academic texts more than they occur in more general texts. Academic lexis, for example, is not found in general English texts such as fiction (Coxhead, 2000, Gardner and Davies, 2014) or newspaper (Gardner and Davies, 2014) at the same frequency as its occurrence in academic texts. Certain learners may find that comprehending academic texts is more challenging than

comprehending general language. According to Maxwell (2013), no one can master academic English or reach native-like level at it.

In their study of accountancy lectures for first year at a university in New Zealand, Basturkmen and Shackleford (2015) found that vocabulary is an essential part of university discourse, and that speaking about vocabulary was the most frequent language-related episode which occurred among learners who took part in the research. Concentrating on vocabulary can be spotted in a lot of areas of teaching and learning in English for Academic Purposes, which shows that the awareness of academic vocabulary in pedagogy and research is increasing. For example, in a lot of places over the world, English for Academic Purposes (EAP) programs have been adding academic word lists into their plans. Academic word lists have also been integrated into learners' dictionaries, online, in learning materials, or in vocabulary tests (Nation, 2016) because they are believed to be a very useful practical tool which can guide learners while deciding which words to focus on in EAP classes or even for independent learning (Nation, 2016).

2.6.2 Academic vocabulary in EAP learning programs

There is a growing interest in researching academic vocabulary in EAP learning programs because knowing more about the needs of EAP learning programs and how to support EAP learners and their teachers is important, but it is even more important to put the theoretical knowledge into practice and to evaluate the effectiveness of approaches to pedagogy. Storch and Tapper (2009) give details of increased and relevant use of academic words lists in EAP writing programs over time in an Australian context. In addition, Crossman and Pinchbeck (2012) focus in their study on Generation 1.5 learners in Canada who are people between the ages of 15 and 24 who are going through the milestones which can shape their life in Canada, such as moving from secondary to possibly post-secondary education or to workforce, developing or creating long-term partnerships, establishing an independent household, or even having children of their own. Generation 1.5 people often postpone or compromise these milestones because of their arrival into a new country. Crossman and Pinchbeck (2012) report that seven Generation 1.5 learners in a university preparation course in Canada have used academic vocabulary in an increased sophistication. Moreover, Luxton et al. (2017) describe gains in academic vocabulary knowledge over a six-month period among 2642 students in 35 secondary schools in Aotearoa/New Zealand. They focused in their study on responses on the academic section of Schmitt et al. (2001) Vocabulary Levels Test. Luxton et al. (2017) divided their subjects into monolingual speakers of English and bilingual and/or multilingual speakers

of English. Participants studied variety of subjects in secondary school, including English literature, English as a second language, physical education, religious education, technology, and science. Teachers were given supporting advice on focussing development of academic vocabulary in classes and were motivated to try to concentrate on teaching this lexis. There were statistically significant gains in test results among Pasifka students, Maori-medium students (kura Kaupapa) students, students who spoke other languages than English at home, speakers of English as a second language who had lived in New Zealand between three and five years, and students who attended English language support classes.

It is very interesting to research more the methods in which learners develop their knowledge, understanding, and use of academic multiword units. Jones and Haywood (2004) tried to include 80 lexical bundles chosen from Biber et al. (1999) in EAP coursework, and to measure the use of these multiword units in the learners' writing. The results of this study suggested that the learners' awareness of the lexical bundles was raised by the classroom activities; however, there was not enough evidence that this awareness was used in the actual practice of writing. Similarly, Li and Schmitt (2009)'s study of the way an MA student whose first language was Chinese used multiword units (lexical phrases) in writing over an academic year in a British University.

2.7 Vocabulary and Reading

According to Nation (2001),

> "Research on L1 reading shows that vocabulary knowledge and reading comprehension are very closely related to each other ... This relationship is not one directional. Vocabulary knowledge can help reading, and reading can contribute to vocabulary growth."

This leads to the question whether vocabulary is better being learnt intentionally during classroom hours or incidentally from extensive reading? Classroom vocabulary research has been especially prolific in two areas, namely: the area of vocabulary teaching and learning through extensive reading; and the approach based on the teaching and learning of carefully selected vocabulary matching some criteria of relevance, frequency and usefulness in performing certain tasks (Chacon-Beltran et al., 2010).

As mentioned earlier, Harris and Snow (2004) assume that vocabulary uptake in classroom is small compared to the large size of the lexicon of foreign language learners. However, some

other research shows different results from those of Harris and Snow (2004). Two issues can be dealt with in this situation. First, the incidental uptake of vocabulary, especially through extensive reading programmes; second the small amount of classroom vocabulary uptake.

2.7.1 The effectiveness of incidental vocabulary learning from reading

According to Sternberg (1987) and West and Stanovich (1991), it is well known in first language acquisition research that reading is one of the very important ways of learning new vocabulary, and that the more the learners read, the more vocabulary they acquire. Horst et al. (1998) assume that reading is essential for second language learning exactly as it is essential for first language development; however, Horst et al. (1998) state that the early studies have not done much to clarify how acquisition happens, and that they may have had some methodological flaws. For example, they may have had very little or too much amount of reading or even not too much management of text difficulty. In addition, they may have had weak measurement instruments, no delayed post-tests or inadequate number of target items. An example of these studies is the experiment run by Saragi et al. (1978) which claimed to prove that second language vocabulary learning occurs incidentally during reading. The participants in the study were native speakers of English who read *A Clockwork Orange* by Anthony Burgess. The experiment aimed at testing the participants' comprehension of many of the fictional register of 241 Russian-based slang words which are used in the novel by the teenage gang members. Burgess utilised this register to depict his characters as speaking a form of Russian-influenced English called Nadsat. The name comes from the Russian suffix which means *-teen* as in thirteen. Subjects were not told to try to learn or remember the Nadsat words, but they had a surprise multiple-choice test. The aim of using Nadsat for testing the subjects was to investigate whether a controlled book-length reading treatment could result in more incidental word learning and a higher pick-up rate than the rate of the other studies with shorter and less extensive tasks. The findings showed that the participants could correctly recognise the meaning of most of the words, especially the ones which occurred frequently; however, Horst et al. (1998) argue that the circumstances of this experiment do not resemble the realistic circumstances in which second language learning occurs. In this study, whereas native speakers of English used contexts they are familiar with to infer, for example, that *droog* meant *friend*, readers in a foreign language who know very few words in the context will possibly find making such connections much more difficult. The results of early research on vocabulary acquisition from incidental exposure in reading showed a low pick up rate where participants correctly identified about only one out of every twelve words tested, which was not encouraging at all (Horst et al., 1998).

Although the mean number of vocabulary items acquired by the participants was 68.4, which formed around three quarters of the 90 words tested, replications of this research have resulted in totally different findings. For example, in a study by Pitts et al. (1989), learners read *A Clockwork Orange* for an hour and were tested on 30 items. The mean score of the Nadsat words correctly identified was two. Similarly, Day et al. (1991) and Hulstijn (1992) in their research which uses the Clockwork Orange methodology, report learners identifying one, two or three words. A higher score of around seven words was reported by Dupuy and Krashen (1993); however, reading was not the only variable since their research involved watching a video. In their experiment, Dupuy and Krashen (1993) had two control groups: one control group was a group of learners enrolled in a third semester French class, and another control group was a group of learners enrolled in a more advanced class. They also had an experimental group of third semester college students of French in one intact class. The experimental group saw the first five scenes of *Trois Hommes et un Couffin* and read the next five scenes in class. After that, they had a surprise vocabulary test which included highly colloquial words from the text. Experimental group performed significantly better than both control groups. The results of Dupuy and Krashen (1993)'s study confirm that incidental vocabulary acquisition is possible in a foreign language situation.

Overall, the findings of these studies of L2 reading show that, on average, in every twelve words tested, one word was identified correctly. Horst et al. (1998) believe that these findings are not startling because of the limitations in the experiments where the chances of reading and acquiring new vocabulary are very limited because the reading treatments were an hour long or even less than that, so whereas the research by Saragi et al. (1978) involved reading a 60000-word novel over a few days, only 6700 words were required to be read by participants in Pitts et al. (1989)'s study, and the participants in Hulstijn (1992)'s study were asked to read an even shorter text of 907 words. Because there was no way to double check whether the participants indeed read the whole assigned texts, the actual amount of words read might be even less than stated in the research methodology of these studies. It was reported by Pitts et al. (1989) that more than 50% of their participants did not finish the reading task. There is no evidence that the other studies have not faced the same issues of limitations in the chances of acquiring new words from reading texts, lack of enough words in the tests, and lack of experimental control; hence, the topic of incidental vocabulary learning through reading needed more support (Horst et al., 1998).

Schmitt (2010) takes all these research weaknesses into consideration and adds that the results of more recent studies which have dealt with some or all of the above-mentioned issues have shown more gains from reading than the ones showed by the results of earlier research. Horst et

al. (1998) ran an experiment to investigate the effectiveness of extensive reading for learning new words. The study tested the effects of three factors: the frequency of the words in texts, the frequency of the words in the English language, and the average of the number of words the participants already know. The researchers aimed at helping teachers comprehend how the success at vocabulary learning may be influenced by word repetition in a text in addition to learner ability, so they tried to look at the average number of vocabulary the participants could acquire from reading a simplified novel. Participants were 34 low intermediate level university students in Oman who were on a 14-week course before taking the Cambridge Preliminary English Test (PET). Participants were given a simplified version of The Mayor of Casterbridge without using dictionaries, and text access was only allowed during the six hours sessions they needed to complete the reading. After analysing the text to find the words suitable for testing, two tests were created: a 45 multiple choice test and a 13-item word association test where participants were given sets of three words and asked to tell the "odd word out". Tests were run one week before the reading sessions, and then straightaway after the reading sessions. They compared the test scores in order to explore whether knowledge of new vocabulary was improved taking into consideration the amount of target words known by the participants at the time of starting the experiment. They also ran The Vocabulary Levels Test (Nation, 1990) before the reading in order to consider the participants' average vocabulary knowledge before reading and to decide how the participants' ability to learn new words through extensive reading was affected by this baseline knowledge. The results showed that learners acquired an average of five words out of the available 23 words. The pick-up rate was about one word in five, which is a larger pick-up rate than the one in the studies before this one. 16% more correct post-reading associations than pre-reading were produced by learners, proving that new vocabulary knowledge can be acquired by comprehension-focused reading.

A measurement study by Horst (2005) showed the advantages of extensive reading with the vocabulary acquisition of 21 adult ESL learners in Montreal and concluded that the learners acquired well over fifty percent of the unfamiliar words which were presented to them during their extensive reading of graded readers.

Pigada and Schmitt (2006) did a one-month extensive reading case-study to investigate the acquisition of spelling, meaning and grammatical characteristics. The result of their study showed that more vocabulary acquisition can occur through extensive reading than the previous studies have shown. In other words, 65% of the target words were enhanced in one of the above-mentioned word knowledge types, for a pickup rate of around 1 of every 1.5 tested words. While spelling was strongly enhanced, even from a little number of exposures, meaning and grammatical knowledge were enhanced to a less extent. Brown et al. (2008) discovered

promising amounts of substantial incidental vocabulary learning in including recognition of word forms as well as word meaning in a multiple choice test. However, they found that their participants were less able to produce the meaning of the words in a translation task. Waring and Takaki (2003) found that their Japanese participants could gain and retain the words more for recognition than for recall knowledge. In other words, the learners recognized the meaning of 10.6 out of 25 words on the immediate multiple-choice test. However, they could only translate 4.6 out of 25 target words. Whereas the recognition of meaning score dropped to 6.1 after three months, the translation score dropped much more dramatically to 0.9. This suggests that incidental vocabulary learning from reading can lead to a partial level of word mastery more than a full mastery level, and that the recall learning tends to result in forgetting more than the recognition learning does.

2.7.1.1 Incidental vocabulary acquisition among EAP Chinese L1 speakers

Several researchers have tried to study incidental vocabulary acquisition among Chinese learners. A lot of these studies have focused on incidental vocabulary acquisition during reading for comprehension (Li and Tian, 2005, Gai and Wen, 2013). Some researchers have tried full analysis of the factors which may affect incidental vocabulary acquisition among learners of English (Wang, 2020). Teng (2019) also studied the impact of different contexts, word occurrence frequencies and the combination of these variables on learning and retention of new vocabulary. The word occurrence rates in this study were 1, 5 and 10 times. Participants were 180 Chinese university students. Each 30 participants read in one of 6 different conditions. In each condition, learners were exposed to the same 15 target words. The results of the study showed that the more informative contexts had statistically significant effects on the productive and receptive knowledge of meaning, whereas an increase in word occurrences resulted in statistically significant effects on productive and receptive knowledge of word form. Regarding word meaning acquisition, only more informative contexts led to the cumulative effect of repeated encounters with new words. Teng (2019) concluded that in an informative context, it is realistic to introduce learners to 10 encounters of new vocabulary in order to trigger the incidental vocabulary acquisition. Because most of these studies focused on the explicit processes and external factors affecting the incidental vocabulary acquisition among Chinese L1 speakers, some researchers tried to investigate more the implications of the learners' internal factors.

The Involvement Load Hypothesis was proposed by Laufer and Hulstijn (2001) to explain how incidental vocabulary acquisition may be affected by different learning tasks. According to the

Involvement Load Hypothesis, the higher the learners' involvement load in the task, the better the effect of the task in enhancing vocabulary acquisition, and the better the vocabulary retention. According to the Involvement Load Hypothesis, the involvement load includes need, search, and evaluation. Need is triggered by the learners' motivation, and is divided into zero-need, the external need, and the internal need. If there is no need, the involvement load index is 0, and no vocabulary acquisition occurs. External needs, such as what is required by the teachers or what is asked for in the learning tasks, is not believed to carry high motivation and is represented by index 1. Index 2 represents higher involvement resulting from the internal need. Search is the cognitive involvement in the process of working out the meaning of a newly-introduced word. An example of search is looking up a word in the dictionary. Index is 0 if there is no search, and is 1 if search is involved. Evaluation is the process of working out whether a new word fits better in the context by comparing it with other words. The index is 0 when there is no evaluation. The index is 1 if the evaluation is intermediate and if it only involves finding out the particular meaning of a word in the context. The index is 2 if the evaluation is advanced and if it involves using target words in making sentences or writing passages. As a result, the involvement load produced by learning tasks may be referred to using particular indexes. The total of the three component indexes refers to the value of the involvement load of the task. This value of involvement load ranges between 0 and 5 during word processing in specific tasks.

Some empirical research on the effect of involvement load was carried out among Chinese L1 speakers. For example, Hou (2004) investigated the effects of involvement load on senior high school students' vocabulary retention. Wu et al. (2007) also studied the relationship between task-induced involvement load and incidental vocabulary acquisition by using reading comprehension and blank-filling activities. In addition, Li and Zheng (2008) used reading in English to research the incidental acquisition of English idioms by non-English major college students, and they claimed that their study has partially verified the Involvement Load Hypothesis. Chang and Li (2009) also researched the effect of various types of English reading tasks on incidental vocabulary acquisition among middle school students. Few of the above-mentioned studies fully confirm the Involvement Load Hypothesis; however, the statistical methods were found by other researchers to have some defects. This is why some more recent studies have tried to use the effect of multiple factors, like English proficiency and involvement load, on incidental vocabulary acquisition. For example, Fan (2021) observed the effects of EFL learners' English proficiency and involvement load which were induced by tasks on incidental vocabulary acquisition among 163 students of non-English specialities in a Chinese

university. Participants were divided into two groups according to their English language proficiency levels. The proficiency levels of the participants were decided according to their scores of College English Test Band 4 (CET-4). 10 target vocabulary items were involved in three tasks: reading for comprehension, blank-filling, and writing. Without being told about it, participants were asked to take an immediate vocabulary test of the target words fifteen minutes after finishing the task. A delayed post-test testing the same vocabulary items was taken two weeks later by the participants. The results of the study showed that whereas learners' English proficiency and involvement load mainly affected the immediate memory, the interactive effect of both factors on incidental vocabulary acquisition was not statistically significant in the immediate post-test. The results of the delayed post-test showed that whereas there was no statistically significant effect of learners' English proficiency on delayed memory, the effect of involvement load was statistically significant. Similarly to the results of the immediate test, the results of retention test showed that the interactive effect of the two factors were not statistically significant. In their study, Dang et al. (2022) ran a quasi-experimental study to investigate the incidental acquisition of collocations among EAP students. Their participants were 165 learners of English for academic purposes at a university in China. The participants were randomly divided into five experimental groups and a control group. The control group did not receive any treatment. Each experimental group was exposed to 19 target collocations in the same academic lecture. The input modes were: reading, listening, reading while listening, viewing, and viewing with captions. The results of the study showed that reading, viewing, and viewing with captions resulted in learning at the form recognition level; however, there were not statistically significant differences in the learning gains across these modes. Learning was affected by nonverbal elaboration, type of vocabulary, and type of verbal elaboration; however, it was not affected by frequency of occurrence, strength of association, comprehension, or previous knowledge of general vocabulary.

2.7.1.2 Incidental vocabulary acquisition among EAP Arabic L1 speakers

Incidental vocabulary acquisition among Arabic L1 speakers learning English for academic purposes was discussed and researched in several studies. Many of those studies have repeatedly reported that the vocabulary size of EFL learners in Saudi Arabia, for example, is small (AlSaif, 2011, Al-Akloby, 2001, Al-Hazemi, 1993, Masrai, 2015). According to these studies, students leave high school with about 1000 words, and these words are mostly very frequent ones. According to Alsaif and Milton (2012), the school English textbooks only expose the learners to half of the most frequent 5000 words; thus, learners in the school of some Arabic countries such as Saudi Arabia should be exposed to a larger volume of vocabulary through

extensive reading in order to help these learners achieve higher in their academic studies. Alahirsh (2014) conducted a study to investigate incidental lexical growth and retention by 80 Libyan university EFL majors who were involved in a two-month extensive reading programme. The participants were divided into the control and the experimental groups. Forty different novels from the university library were available for the participants to read but the learners were not allowed to take the novels home. Participants were not told about the post-test, and were not allowed to use dictionaries in order to avoid the process of intentional learning of the words. Following a 9-week treatment, only 18 participants in each group could finish the experiment due to the Libyan crisis. The results of the study showed statistically significant vocabulary growth by the experimental group compared to the control group. Whereas the mean scores of the control group were 0.47 (SD = 0.14) in the pre-test and 0.49 (SD = 0.11) in the post-test, the mean scores for the experimental group were 0.52 (SD = 0.14) in the pre-test and 1.52 (SD = 0.30) in the post-test. These results show that the extensive reading has led to significant building of the incidental vocabulary knowledge. Data from interviewing 14 participants in the experimental group of the study indicated that around 93% of them (13 out of 14) perceived extensive reading positively. Aiming to investigate the vocabulary acquisition gains from informal extensive reading activities among Arabic L1 learners of English, Alsaif and Masrai (2019) ran a study with a single participant who was instructed to read extensively for eight weeks outside the classroom as an informal activity. The participant's written receptive vocabulary knowledge was measured before and after the treatment. The results of the study showed that the participant's vocabulary acquisition was significantly affected by extensive reading. It was indicated by the results of this study that a vocabulary uptake of about eight words from extensive reading intervention has occurred compared to about two words every hour of contact during language classroom input where reading texts are short and not assembled in one place.

2.7.2 Extensive reading

According to Day and Bamford (1998), there are a lot of benefits which can be gained by using extensive reading programmes. These benefits include gains in reading and writing proficiency levels, oral skills, vocabulary knowledge, and motivation. In order to understand a text, readers need to know 98-99% of the words in it. This is why authentic texts, which are not aimed at readers at certain level, may not be easy for readers who are not advanced. Since the vocabulary load in authentic texts is not easy for developing learners, graded readers where the vocabulary is accommodated to suit the learner's level are recommended for them (Nation and Wang, 1999). A large body of research has shown that a considerable amount of vocabulary learning

can be derived from graded readers although they have been considered boring and poorly written (Schmitt, 2010). Horst (2005) found that her learners acquired more than 50% of the unfamiliar words they came upon in the graded readers they read as part of a measurement study which aimed at showing the advantages of extensive reading with the vocabulary acquisition amongst 21 adult ESL learners in Montreal.

Similarly, in their research, Al-Homoud and Schmitt (2009) compared an extensive reading class with another class which used the more traditional methods of reading intensively and giving out vocabulary exercises. The classes were part of a Saudi college pre-sessional language course. The setting of the classroom implied many issues for the extensive reading approach, especially with the students of relatively lower linguistic abilities who did not consider reading for pleasure to be ideal, especially that the course was over a short period of time. The circumstances in which the extensive reading class was carried out could be considered as challenging. Despite all that, gain scores, whether in reading comprehension ability, reading speed, or vocabulary acquisition, resulted in the extensive reading approach being as efficient as the intensive approach was despite the fact that the measurement tools which measured the three variables should have supported the intensive approach over the extensive one. All in all, the study found that over the course of ten weeks of the learners getting exposed to extensive reading and graded readers, the vocabulary of Saudi learners was increased at the 2000, 3000, and 5000 frequency levels, and their reading speed and attitude towards reading were improved. Expectedly, the amount of reading in an experiment is an essential factor which may lead to the change of the whole study results. Out of ten variables entered into a regression analysis, the only significant predictor of gain scores in overall language proficiency was the quantity of extensive reading done during a two-month course (Renandya et al., 1999).

2.7.3 L2 vocabulary acquisition during classroom hours

As mentioned earlier, Harris and Snow (2004) claim that vocabulary uptake in classrooms is small. However, there are carefully controlled studies which have focussed on classroom language and on the words acquired from learners' exposure to classroom language, and which have found that classroom hours might be the main source of vocabulary learning for most learners, particularly at the outset of learning, and that the uptake could be large. One example is Vassiliu (2001). The learners in Vassiliu's study were beginners at a frontisteria in Greece. Frontisteria are schools in Greece that operate within a rather loose legislative framework. The learners were exposed to about one thousand words in their first year of study, and another thousand words when they were tested at the end of the year. The learners could recognise about half of the words (on average). Although the rate of uptake can be considered

relatively small-about five words per class session, good learners learned most of the words they were exposed to during the study.

2.7.4 Other possibilities

As explained in the previous two paragraphs, there is a debate about whether it is enough to learn L2 vocabulary during the limited classroom hours or not; second whether incidental vocabulary acquisition through extensive reading is effective or not. A question may arise here about other ways of acquiring L2 vocabulary. Milton and Donzelli (2013) suggest that the answer to this question may lie in the informal language learning activities which "well-motivated" learners engage in. Examples of these activities are: listening to songs, watching movies or TV programmes, or even playing computer games in the foreign language they are learning. Some case studies have been run earlier at Swansea University to investigate and track the lexical gains from the above-mentioned activities. Milton (2008) conducted a study to investigate vocabulary uptake from listening to songs over the course of eight weeks. The learner involved was pre-tested on the lexis of the songs before listening to a CD of Greek ballads from films while reading the lyrics with a translation, and singing alone. The learner was tested every week on the vocabulary of the songs using Wesche and Paribakht (1996)'s Vocabulary Knowledge Scale. Surprisingly, the results showed that large volumes of lemmatised words, or different inflected forms of the same word which are grouped together, were learned, over 30 words per hour from the activity. In addition, the learner could also translate over 90% of the words and provide the lexical and grammatical contexts the words occurred in; however, it can be argued in this case that the pre-test in Milton (2008)'s study could have introduced the words to the learner who could then look the words up.

Another study reported by Milton (2008) was about a learner who watched a film with sub-titles. The study showed that the man learned sixteen words per hour. The previous examples show that a very large lexicon can be acquired by motivated learners while doing informal activities whose aims may not be necessarily teaching or learning.

2.8 Online Reading

For the past twenty years, the literature has paid particular attention to reading online texts and methods which enhance text (Akyel and Erçetin, 2009, Huang et al., 2009, Anderson, 2003). Different formats for reading texts and different methods to deal with information which may be hard for readers who are taught to infer meaning from traditional paper-based reading texts are provided by electronic documents (Coiro, 2003, Protopsaltis, 2008).

2.8.1 Why is online reading considered beneficial?

Online reading activities in second language acquisition (SLA) have been found to be advantageous primarily because they provide broad exposure to authentic language materials which are accessible through the Internet, the access to grammatical and vocabulary clarifications, connection of multimedia representation and simple reading text, access to self-paced reading (Levy and Stockwell, 2006), and online dictionaries (Chun, 2001).

In her empirical study, Chun (2001) tried to find out the ways in which L2 learners access information during reading L2 texts in an online learning environment. Two web-based reading texts in the netLearn program that the researcher developed were read and then summarized by twenty-three second-year German students. During the experiment, the participants' use of multimedia assistance resources, which involved a program-internal glossary, an online bilingual dictionary, and an audio narration of the text, was tracked. The summaries that the participants wrote after they read the texts were scored according to the number of propositions recalled. The tracker took notes of the number of vocabulary the learners looked up and the duration of time the learners spent reading the texts and then writing the summaries. A number of learners thought aloud protocols during the online reading, and some learners were interviewed following their use of the program. The interviews involved questions about these learners' metacognitive reading strategies. The analysis of both the qualitative data (for example analysing the reasons of looking words up and when the learners did that, and whether the audio was found to be supportive), and the quantitative data (for example the amounts of vocabulary looked up, the duration of time spent on tasks, and the amount of propositions recalled) gave insights into the way L2 learners use multimedia to help with their comprehension during reading. Comparing the learning of words which were hyperlinked online and learning the words which were not, Chun (2001) concluded that learners who were reading online performed better as long as they were hyperlinked with a website. Moreover, having the internal vocabulary glosses as a substitute to the external online dictionaries saved the learners' time and helped them read without interruption. Similarly, Tozcu and Coady (2004) concluded that a significantly larger amount of vocabulary was learned by students who used Tutorial CALL to learn highly frequent vocabulary than the amount learned by those who were not exposed to the CALL program, the reaction time for frequent word recognition was reduced amongst learners who were exposed to the CALL program, and reading comprehension was shown to be significantly better among those learners than those who were not exposed to the Tutorial CALL. Tozcu and Coady (2004) tried to explore the influence of direct vocabulary learning using Computer-assisted Language Learning (CALL) on vocabulary knowledge,

reading comprehension, and speed of word recognition by dividing their participants into two groups: the treatment group who used the computers for three hours per week for eight weeks in order to study around 2000 of the highly frequent words in English, and the control group who read texts and did reading comprehension exercises over the same period of time. The learners in the treatment group showed significantly better gains than those in the control group despite the improvement in the vocabulary gain and reading comprehension, and the decrease in reaction time for frequent word recognition in both groups.

2.8.2 Other views about online reading

Online English reading materials may include distractors such as pop-up windows which provide explanations, brightly coloured words, hyperlinks to different websites, explanatory pictures or videos, and many more. These distractors can increase the amount of cognitive load on the working memory. As a result to this increase in cognitive load, effective reading can be prevented (Akbulut, 2007). Differently, ESL reading skills were found to be improved by web pages with simple instructional design inspired by research results (Akbulut, 2007, Chun, 2001, Gerjets et al., 2004). According to Skulmowski and Xu (2022), although digital technology can introduce very exciting ways to facilitate learning, and although technology can be used by learners to search the virtual world and interact with their learning applications, cognitive load can be increased by many design factors involved in digital learning. For example, when learners feel that they are entirely engaged in a virtual world while they are learning using online materials, their experience becomes totally different from when they are using the traditional learning resources; however, learners' immersion in the virtual world may distract the learners and make them use their cognitive resources on the experience itself rather than on the learning process. In this case, the design factors which may aim at enhancing the learning process can lead to unwanted cognitive load (Frederiksen et al., 2020).

2.8.3 Factors affecting L2 online reading

The cognitive load related to online multimedia-based reading in general and L2 online reading in specific has been found to be affected by many factors. One of these factors is the individual differences amongst the learners. The second factor is the effectiveness of instructional formats. The third factor is the amount of text displayed on screen.

2.8.3.1 Individual differences

Individual differences among the learners are believed to be very important factors which can play an important role in online reading comprehension as well as vocabulary acquisition.

2.8.3.1.1 Learners' different abilities

An example of individual differences is the different abilities of the learners. In her empirical study which is reviewed in this research in 2.7.1., Chun (2001) tried to find out the ways in which L2 learners access information during reading L2 texts in an online learning environment. The results of Chun's study gave insights into the way L2 learners use multimedia to help with their comprehension during reading, but one of the aspects the study focused on was the different abilities of the learners where higher ability students spent less time on reading and looked up less words than lower ability students. Similarly, in their research, Ozono and Ito (2003) tried to concentrate on the logical connectives as catalysts for interactive reading. Their main goal was to demonstrate the ways the types of logical relations and the levels of L2 proficiency levels could affect the understanding of texts. They developed a special test called Logical Relations Reading Test and asked Japanese University students learning English as a second language to choose the appropriate logical connectives; namely "for example" for ILLUSTRATIVE, "therefore" for CASUAL, and "however" for ADVERSATIVE, for the target logical relations. The study concluded that students' various abilities affected their reading comprehension and their cognitive load. Oronzo and Ito's study found that low ability students have a tendency to guess the meanings of the words more frequently than high ability students and this increases the cognitive load.

2.8.3.1.2 Learners' memory capacity

Another example of individual differences among learners is their memory capacity. Chun and Payne (2004) tried to investigate the effect of learners' individual differences, such as their working memory capacity on their L2 vocabulary acquisition and their reading comprehension. Their study included 13 students in a second year German language course. The learners were asked to read a German short story on a multimedia CD-ROM, and were given access to multimedia annotations of hard vocabulary. The learners were tracked during the process of looking up words, and during completing a vocabulary test and comprehension exercises. A nonword repetition test and a reading span test were taken by the participants in order to assess their working memory capacity. Data analysis showed a relationship between phonological working memory (measured by the use of nonword repetition test) and the amount of times learners looked up vocabulary where more words were looked up by participants with lower

verbal working memory capacity. These results indicate that learners make use of the multimedia look-up properties of the CD-ROMS trying to make up for the working memory capacity limitations. Accordingly, Ardac and Unal (2008) investigated the impact of the changes in the amount of text displayed on screen on the learning from a multimedia instructional unit focussing on the main ideas of coordinate geometry. They applied two versions of the instructional unit; namely short-text and whole-text, and investigated how they affected the understanding amongst 101 seventh grade students of different abilities who were randomly assigned either the whole text or the short text. A pre-test, post-test and retention test scores were used for data collection. Memory capacity was measured by the use of the sub-tests of the Structure of Intellect-Learning Abilities Test. The retention scores showed that students with higher memory capacity who worked with the whole-text version acquired more information than students with lower memory capacity who also worked with the whole-text version; however, the differences between the results of the groups who worked with the whole-text version and those who worked with the short-text version were not statistically significant regardless of the memory capacity of those students. It can be concluded that memory capacity had an impact on learning in Ardac and Unal (2008)'s study; thus, the workability of design principles for multimedia instruction may depend on learners' characteristics.

2.8.3.2 Effectiveness of instructional formats (integrated and split-attention)

The effectiveness of instructional formats has been researched in various studies. Instructional design, whether paper-based or multimedia-based, may have a large effect on the learners' cognitive load, and consequently affect learning performance whether negatively or positively. This study focuses on the integrated and split-attention instructional formats in online reading texts.

2.8.3.2.1 What is the split-attention effect?

The split-attention effect can be defined as an effect in learning which is related to some instructional materials which are designed unsatisfactorily. The split-attention effect occurs when the same modality, for example, visual, is employed to express different sets of information in the same context. When the split-attention effect is employed, learners need to split their attention between different displays and various materials, such as a picture or a voice recording and a reading text, in order to comprehend the expressed information. The split-attention effect can be visual or auditory; thus, it can occur during reading or listening.

In the field of language learning, it has been found that various instructional designs affect language learning in various ways. For example, the split-attention effect occurs when the readers' visual focus is on various types of information. To acquire any linguistic knowledge, readers must split their attention between these different information types, and during this process a greater cognitive load is expected because the more-than-one need to be mentally integrated before any knowledge is acquired from the reading text. As a result, a considerable cognitive load on the short-term working memory may be happening during the learning process (Schnotz and Kürschner, 2007).

Mutlu et al. (2022) designed a study with the goal of evaluating the effect of the split-attention format in multi-media learning environments by using some objective measurements, such as EEG and eye-tracking, to compare subjective and objective cognitive load measurements. The researchers designed two different multimedia learning environments in a focused (integrated) and split-attention (separated) formats. 44 participants were divided into two groups for integrated and split-attention. Their eye movements and multichannel electroencephalography (EEG) signals were recorded. At the end of the experiment, the researchers collected self-report ratings and measured the retention performance of the participants. Pearson Correlation was used for data analysis. The eye-tracking and EEG results showed statistically significant differences between the groups due to the split-attention effect. Significant differences between the fixation, brain wave, and retention performance of the two groups were found, where fixations in the split-attention group were higher than those in the integrated group. Moreover, the integrated (focussed attention) group showed higher retention performance than the split-attention group; thus, more cognitive load was involved in the situations where the text was not integrated into the image. In addition to that, narrating the text instead of printing it helped with focussing the attention of the learners. To minimise or even prevent the split-attention effect, Mutlu et al. (2022) suggest integrating the text into the image when designing online learning materials.

2.8.3.2.2 Integrated and split-attention instructional formats

Various amounts of memory load can be caused by the various instructional formats designed to present the same learning material. Because there is a considerable amount of on-screen text in instructions based on multimedia, researchers have studied the various aspects of online reading in instructions of both first language (L1) and second language (L2) (Chun, 2001, Tozcu and Coady, 2004, Constantinescu, 2007).

Whereas the split-attention instructional format was believed to be increasing the cognitive load, the integrated instructional format was found by Yeung et al. (1998) to be helpful in reading comprehension for both first language (L1) and second language (L2) learners. In this study, Yeung et al. (1998) administered five experiments to investigate the influence of cognitive load management using explanatory notes in reading paragraphs for readers with various linguistic levels. The first experiment showed that while explanatory notes enhanced the comprehension of 5th-grade, first-language learners (high-level processing), it did not improve their vocabulary learning (low-level processing). The second experiment showed that whereas vocabulary definitions which were integrated within the reading passages (integrated format) improved 5th graders' comprehension in comparison to a separate vocabulary list (separated format), it reduced vocabulary learning. The third experiment showed that an integrated format reduced comprehension amongst adult readers but improved vocabulary learning. In the fourth experiment, participants were low-ability 8th-grade learners of English as a second language (ESL) whose reading comprehension and vocabulary learning were affected similarly by both vocabulary formats to the 5th graders in the second experiment. The fifth experiment found that both vocabulary formats affected high-ability ESL learners similarly to the way they affected adults in the third experiment. Yeung et al. (1998) argue that the degree to which instructional formats impose an extraneous cognitive load affects its efficiency; thus, the learners' abilities play a very important role on whether the same format; namely split-attention or integrated, may enhance performance or decrease it among learners.

In their study, van Schaik and Ling (2003) had a goal of adding to the intranet design guidelines for usability by designing an empirical study which used an existing "live" site in addition to some realistic tasks. They used an experimental design in order to study the effect of link colour on both performance and subjective measures. Participants were eighty-six undergraduates who were given an information retrieval task written by two different versions of the same online text. The study concluded that simpler and more familiar page layouts and font colours could reduce the cognitive load among users; thus, web designers responsible for instructional format are advised to use standard web designs and navigation methods so that students are more familiar with what they are browsing.

In an investigation of whether various hypertext architectures could be connected to an individual's cognitive style to make learning easier, Graff (2003) used three hypertext architectures, linear, hierarchical, and relational. Graff (2003) used the Cognitive Styles Analysis to assess the individuals' cognitive styles or, in other words, the way they thought, perceived and remembered information. Cognitive Styles Analysis (CSA) was developed by

Richard J. Riding in 1991. Riding's CSA is the most frequently used computerized measure of cognitive styles. It depends on a two-dimensional model which includes W-A, wholistic-analytic, and V-I, verbal-imagery. These two dimensions are considered the two basic learning models of the cognitive style (Riding and Cheema, 1991). The results of the Cognitive Style Analysis in Graff (2003)'s study suggest that the individuals' performance was affected by their cognitive style and by segmentation. According to Graff, particular designs, such as hierarchical architecture, can reduce cognitive load effectively and thus make online reading much easier. Furthermore, Keller et al. (2006) investigated whether information visualizations played a role in enhancing knowledge and acquisition. They also aimed at investigating the way information visualizations, from a cognitive perspective, could be designed in order to function as helpful learning tools. The experimental study found that knowledge acquisition was enhanced by information visualizations; moreover, the empirical results found that three-dimensional designs lead to higher cognitive load than two-dimensional ones and that color-coded information visualisations slightly increased performance in a knowledge test compared to monochromatic ones.

Similarly, a correlation between vocabulary learning and various formats of multimedia instructions was found by Akbulut (2007) who tried to research how incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language affect the immediate and delayed influence of various hypermedia glosses. Participants, who were sixty-nine freshman EFL students at a Turkish university were randomly divided into three groups; each group of learners was assigned to one type of annotations: the first was definition of words, the second was definitions coupled with associated pictures, and the third was definitions coupled with associated short videos. They were all required to read an annotated test with the aim of comprehending the text. The researcher collected the data using a vocabulary pre-test, a vocabulary post-test, a delayed vocabulary test in addition to a reading comprehension test. Since the experiment aimed at measuring incidental vocabulary learning, the participants were not told about the vocabulary tests in advance. The study found that significantly higher vocabulary scores were achieved, in both immediate and delayed post-tests, by learners who had access to definitions accompanied by visuals than by learners who only had annotations on their own; thus, a correlation between the effectiveness of instructional formats online and vocabulary learning, particularly for advanced L2 learners. Moreover, students' cognitive styles could also be influenced by the various layouts of web pages and the way online reading texts are presented. However, no differences in the test results of the reading comprehension were found.

In their study which aimed at exploring the effect of split-attention and integrated instructional formats on students' cognitive load and on the way these instructional formats could affect online reading and vocabulary learning in second language, Al-Shehri and Gitsaki (2010) randomly assigned twenty students studying English as a Second Language (ESL) at an Australian language institution to four groups; namely, SAND (Split-Attention No Dictionary), SAOD (Split-Attention with Online Dictionary), IFND (Integrated Format No Dictionary), and IFOD (Integrated Format with Online Dictionary). Participants were given an online reading comprehension task. Whereas a typical reading text followed by comprehension questions was given to SAND and SAOD participants, a reading text with comprehension questions inserted within it was given to IFND and IFOD participants. An online dictionary was accessed by SAOD and IFOD participants. The results of the study showed that the learners' reading comprehension was made easier when the integrated format was used than when the split-attention one was used. In addition to that, learners in the SAOD and IFOD groups had better results in the vocabulary test; however, the reading task took them more time than it took the other two groups. Moreover, more vocabulary was looked up by learners in the split-attention groups than learners in the integrated format groups.

Genç and Gülözer (2013) studied the effect of different cognitive load caused by different instructional formats (split-attention and integrated formats) and text presentation methods (paper-based and online) on learners' second language reading comprehension performance. Whereas comprehension question were positioned after the reading text in split-attention format, the integrated format involved dividing the reading text into paragraphs and inserting comprehension questions between the suitable paragraphs. The researchers ran reading comprehension tests to collect quantitative data from forty pre-service teachers in an English Language Teaching department. The participants were divided into four groups: Online Reading Split-Attentions Format, Online Reading Integrated Format, Paper-based Reading Split-Attention Format, and Paper-based Reading Integrated Format. The experimental design of the study involved a post-test only, and the results of this post-test revealed no statistically significant difference among the four groups with regards to the scores of learners' second language reading comprehension tests. However, a statistically significant difference in L2 comprehension between participants who read online text and those who read paper-based text was suggested by the results.

2.8.3.3 Amount of text displayed on screen

Ardac and Unal (2008) found that low memory students' cognitive load was reduced by short-text reading while more information was acquired from whole text reading by students with high memory who also performed better while working on whole text reading than low memory students. Similarly, Mikk (2008) found that reading comprehension was affected by the length of sentences when a high correlation was found between long sentences and overloading working memory. Moreover, Hung (2009) found that an integrated format (where questions were actually integrated into related paragraphs) successfully reduced any irrelevant cognitive load, whereas a split-attention format (comprehension questions following reading passages) increased the irrelevant cognitive load. In other words, the integrated format basically divides the reading text into multiple sections, where each section is followed by relative comprehension question(s), while the split attention format follows the used-to method of a reading text preceding the comprehension question(s).

2.9 Summary

A very large vocabulary knowledge is required in order for learners to be able to use any language. Certain learning strategies should be used to help learners build their vocabulary knowledge.

Research has found that one of the most important points to consider is that learning a word involves much more than only learning its meaning. In fact, amongst other vocabulary learning criteria, the knowledge of word families and derivations, in addition to the knowledge of collocations are very important criteria of learning a word in addition to knowing its meaning.

The findings of various studies reveal that it is very beneficial to learn the words within context rather than on their own, and that reading plays a very important role in incidental vocabulary learning.

The role of online reading in reading comprehension and vocabulary learning has been controversial.

Some researchers find that reading online may facilitate vocabulary learning and reading comprehension because of using some tools which are not accessible when reading on paper.

In contrast, other researchers believe that online reading can include materials which may increase the cognitive load on the brain.

A large body of research has found, though, that the cognitive load related to online multimedia-based reading in general and second language online reading in specific may be influenced by other factors. One of these factors is the individual differences among the learners and their L2 proficiency, the second factor is the effectiveness on instructional formats and the existence of the split-attention effect, and the third factor is the amount of text displayed on screen.

This study aims at filling some gaps in the reviewed literature. First, many studies which investigated the potential of multimedia in building vocabulary knowledge among English as an additional language learners, including the research which looked at the three above-mentioned factors and even focused on the impact of certain instructional formats on students' cognitive load and their building of vocabulary knowledge, mainly used multimedia glosses and vocabulary knowledge, multimedia annotations and vocabulary acquisition, or even multimedia-enhanced dictionaries; however, none of the studies that I have come across so far have focused particularly on the impact of the different instructional formats; such as online split-attention and integrated instructional formats on vocabulary acquisition among EAP learners.

Second, although Al-Shehri and Gitsaki (2010) tried to investigate the impact of split-attention and integrated instructional formats on students' cognitive load by focusing on the effect of different instructional formats on reading comprehension and reading speed, the main focus in their study is not vocabulary. Moreover, Al-Shehri and Gitsaki did not take into consideration the other two factors of individual differences and the amount of information presented to L2 learners.

Third, the lack of research which aims at investigating the relation between cognitive load and building vocabulary knowledge; thus, on the role of different instructional formats in vocabulary learning among ESOL students whether they are EAP or not, and the lack of studies which involved the three factors and their effect on cognitive load encouraged the researcher to use the cognitive load theory as a background for this study and contribute to the field by investigating the impact of integrated and split-attention online reading formats on building adult EAP vocabulary knowledge.

It is hoped that this contribution can shed the light on this topic. The method used in order to investigate the topic of this study is quasi-experimental. This method was used to collect quantitative data needed for the research. The treatments are integrated and split-attention formats, and the measurements to collect data are the pre- and post-tests. All experiments were

held at a language Centre in a UK university. The next chapter will present the methodology of this study.

Chapter 3. Research Methodology

3.1 Introduction

This chapter provides an overview of the research focus and research questions. It presents and discusses the methodological approach and research design best suited to answer the research questions. A quantitative method design is proposed in order to answer the research questions. An overview of the research design then follows, starting with an explanation of the sampling procedures and the participants in the study, and then an outline of the key method employed; namely, quasi-experiment. Because the design and the reliability and validity are important in the choice of research instruments, the used method is justified, and the reliability and validity checks are all explained. The next section illustrates the specific process of data collection. After that, the technological features within the study are all explained before the overview of the methods used for data analysis. Ethical issues concerning the research process are presented in the subsequent section. The chapter ends with a brief summary of the written sections.

3.2 Research Focus and Research Questions

Most of the research which focused on the potential of multimedia in building vocabulary knowledge among English Speakers of Other Languages (ESOL), whether they are English for Academic Purposes (EAP) learners or not, even the ones which looked at the effect of certain instructional formats on students' cognitive load and their building of vocabulary knowledge, focused on multimedia glosses and vocabulary knowledge, multimedia annotations and vocabulary acquisition, or even multimedia-enhanced dictionaries. However, none of the studies that I have come across so far have considered the influence of the online split-attention and integrated instructional formats on EAP learners' vocabulary acquisition in particular. So, this study is designed to investigate whether there is a relation between cognitive load and building vocabulary knowledge; thus, whether vocabulary among EAP students can be learnt more effectively depending on the different instructional formats of the online reading texts; namely, split-attention and integrated instructional formats.

Although Al-Shehri and Gitsaki (2010) tried to investigate the effects of split-attention and integrated instructional formats on students' cognitive load and the way they might facilitate second language online reading and vocabulary learning, their main focus was the effect of

different instructional formats on reading comprehension and reading speed, while the focus of this study is the effect of the instructional formats used in Al-Shehri and Gitsaki's study on building vocabulary knowledge; thus, the first research question of this study is whether an online reading task in the integrated format can be more effective than in the split-attention format in helping adult EAP learners build their vocabulary knowledge.

One of the most important points to take into consideration is that learning a word implies more than the knowledge of the word meaning, spelling, and pronunciation only (Schmitt, 2000). As a matter of fact, additional criteria to word knowledge involve, amongst other criteria, word collocations, grammatical characteristics, and word family knowledge. Learners should have an understanding of all the forms of word knowledge in order for them to use words confidently. The target vocabulary in this study is introduced within reading texts in addition to vocabulary exercises accompanying the texts. Moreover, the vocabulary exercises in this study follow Schmitt and Schmitt (2011)'s model of concentrating on the three aspects of word meaning, word families (derivative forms of words), and collocations in order to achieve the maximum teaching goal of the exercises. According to Schmitt and Zimmerman (2002), learners do not necessarily know all of the categories of a word family although they may be aware of some of its forms. Nevertheless, learners should know the right form of a word (noun, verb, adjective, or adverb) in a certain context, and hence the focus on derivative forms of words in this study. Learning word collocations is also very essential when learning new vocabulary. Although it is hard to teach all word pairings, integrating the element of collocations knowledge into the exercises in this study aims at helping the learners become more aware of collocations and thus enhancing their intuitive understanding of collocations for individual words. This is why the first research question is being addressed through investigating three subcategories of the word knowledge and be addressing the three sub-questions of whether an online reading task in the integrated format can be more effective than in the split-attention format in helping adult EAP learners build their word meaning knowledge, word families knowledge, and collocations knowledge.

There are two variables which might play an important role in the effect of certain instructional formats on learners' vocabulary acquisition which Al-Shehri and Gitsaki (2010) did not focus on. These variables are the proficiency level of the learners themselves and their first language (L1).

Aiming to give recommendations about the methods L2 learners use multimedia to help with their reading comprehension, Chun (2001) tried to find out the ways in which L2 learners access

information during reading L2 texts in an online learning environment. One of the aspects the study focused on was the different abilities of the learners where higher ability students spent less time on reading and looked up less words than lower ability students. Similarly, aiming to clarify the ways the types of logical relations and the levels of L2 proficiency levels could affect the understanding of texts, Ozono and Ito (2003) tried to concentrate on the logical connectives as catalysts for interactive reading. Their study concluded that students' various abilities affected their reading comprehension and their cognitive load; thus, low ability students have a tendency to guess the meanings of the words more frequently than high ability students and this increases the cognitive load. Considering the results of these studies, this research also aims at investigating whether adult EAP learners at different proficiency levels acquire new vocabulary items from the reading text in both split-attention and integrated formats differently.

On one hand, several researchers have tried to study incidental vocabulary acquisition among Chinese learners. For example, Teng (2019) studied the impact of different contexts, word occurrence frequencies and the combination of these variables on learning and retention of new vocabulary. Wang (2020) also investigated some factors which might affect incidental vocabulary acquisition among learners of English. Fan (2021) observed the effects of EFL learners' English proficiency and involvement load which were induced by tasks on incidental vocabulary acquisition. Moreover, Dang et al. (2022) ran a quasi-experimental study to investigate the incidental acquisition of collocations among EAP students. On the other hand, incidental vocabulary acquisition among Arabic L1 speakers learning English for academic purposes was discussed and researched in several studies, such as AlSaif (2011), Al-Akloby (2001), Al-Hazemi (1993), Masrai (2015), and Al-Akloby (2001). One of the studies which investigated incidental lexical growth and retention was run by Alahirsh (2014). Another study which focused on vocabulary acquisition gains from informal extensive reading activities among Arabic L1 learners of English was conducted by Alsaif and Masrai (2019). None of the studies above compared the vocabulary gains of Arabic L1 and Chinese L1 EAP learners. In addition, although some of the studies focused on the involvement load theory, the cognitive load and the effect on the working memory of the learners was not investigated. This research is using some of the methodology used in the previous studies to investigate whether adult EAP learners whose first language is Arabic acquire new vocabulary items from the reading texts in both split-attention and integrated formats differently from adult EAP learners whose first language is Chinese.

Although Al-Shehri and Gitsaki (2010) tried to investigate whether a reading task in the integrated format was more effective than in the split-attention format in making L2 reading easier for ESL readers, and whether an online dictionary with either of the two above-mentioned formats could affect students' reading comprehension and speed, they only implemented a post-test which measured the receptive vocabulary of the participants in their study, thus their study did not include any pre-test to measure the vocabulary knowledge of their learners before the experiment and compare the results with the post-test that they ran.

So, it can be said that Al-Shehri and Gitsaki's study was preliminary, and although it had some problems with the methodology that was used, part of the methodology used in their study was adopted and built on in this research.

The literature reviewed above was all considered to design this study and to test the following research questions and sub-questions:

1. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their vocabulary knowledge?

1.1. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word meaning knowledge?

1.2. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word families knowledge?

1.3. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their collocations knowledge?

2. Do adult EAP learners at different proficiency levels acquire new vocabulary items from the reading text in both split-attention and integrated formats differently?
3. Do adult EAP learners whose first language is Arabic acquire new vocabulary items from the reading text in both split-attention and integrated formats differently from adult EAP learners whose first language is Chinese?

3.3 Methodological Approach

3.3.1 A quantitative approach

Patton (1990) has advocated the importance of understanding that "different methods are appropriate for different situations" and thus the aim of a study, the research questions or

research hypotheses being investigated, and the available sources largely determine designing a study which is appropriate for a specific situation.

Usually, a quantitative approach is required if the researcher is focusing on a certain group as "representative of a population, with a focus on common rather than individual features" (Richards et al., 2011). A quantitative approach is generally related to describing what people do (Hammersley, 2018). On the other hand, a qualitative approach is adopted if the goal of the research is to "understand the behaviour, beliefs, understandings, values, etc. of a particular group (or individual), or explore aspects of a world in its context" (Richards et al., 2011). The central aim of a qualitative approach is to "document the world from the point of view of the people studied (Hammersley, 2018), and to know "how people define their situations" (Marshall and Rossman, 2016). Dörnyei (2010) draws the distinction between the two approaches talking about meaning in the particular (qualitative) and meaning in the general (quantitative). Thus, whereas qualitative research focuses on understanding a natural context, quantitative research focuses on "establishing the parameters or conditions relating to a specific group or situation"(Richards et al., 2011).

Since this study is not focusing on the participants' beliefs, values, or understanding, and since the groups of participants are not meant to be studied as individual cases, but as representative of the population of EAP adult learners, quantitative research methods are believed by the researcher to be the right methods to reach "a macro-level generalization supported by empirical evidence" (Richards et al., 2011) when collecting the data needed for this research.

A quantitative approach is systematic and focused. The variables were defined in advance, a logical scale of values was assigned to them, and they were expressed in numbers because the nature of the research involves accurate measurement in order to produce reliable and replicable data which can be generalized to other contexts. The nature of this research requires a large enough sample which enables the researcher to filter out the idiosyncratic differences related to specific individuals and to look at the commonalities that are found in the data. Moreover, this research aims at testing the effect of two different online reading formats on building vocabulary knowledge among adult EAP learners. Because of the two above mentioned reasons, a quantitative approach was found by the researcher to be the best and most practical approach to try to answer the research questions of this study.

The main aim of using the quantitative approach to answer the research questions of this study is to result in reliable and replicable data which can be generalizable to other contexts.

3.4 Data Collection and Procedures

3.4.1 Participants

The participants in this study were a group of adult EAP learners (n=60) who were English for University Studies learners at INTO, Newcastle upon Tyne. 60 was the maximum number of participants which could participate in the study. A large number of students was addressed, but 60 volunteered to participate in the study. These learners (35 Arabic speakers, 25 Chinese speakers) were males and females. They were all between 20 and 30 years old. All of them had studied English as a foreign language in their home countries. Their English Proficiency level was between 4.5 and 6.5 in IELTS. Assigning the participants into low proficiency level learners and high proficiency level ones was according to their IELTS scores where learners with 5.5 IELTS score considered as the separating score between lower proficiency level learners and higher proficiency level learners. According to www.ielts.org, 5.5 is the score between 5 (which refers to the skill level "modest") and 6 (which refers to the skill level "competent"). The participants took the test at the start of their studies at INTO, and they were allocated into their classes according to their IELTS scores; thus, they participated in the study before completing any course after taking the IELTS test. No buffer zone was considered when allocating the participants to their groups because 5.5, which is the separating score is considered by the official classification of IELTS scoring to be a modest score; thus, any score up to- and including 5.5 was considered a lower proficiency score compared to the scores which were 6 and above. All participants attended morning and afternoon classes at INTO. However, the experiment did not interfere with the teaching hours. The researcher met the participants outside learning hours in small groups and introduced them to the website. The website did not have any of the texts from the curriculum.

The criteria used to select participants were their IELTS scores and their L1 only. Since their ages were between 20 and 30, it was understood that some of them were intending to study in an undergraduate program. It was also known by the researcher that they were all planning to study different majors. Although the researcher did not keep a record of all their majors, she was told that none of them was planning to study English or TESOL.

3.4.1.1 Sampling procedures

A combination of convenience/opportunity sampling and a quota sampling approach was followed to employ the participants in this research. The procedures started with setting a

sampling frame and then the main proportions of the subgroups were decided according to the criteria set within the frame. The actual sample was selected in a way which reflected the required proportions. No random selection was made. Instead, the researcher was given permission to address the population; i.e. INTO students, explain about the experiment, and ask for volunteers with certain criteria. The researcher contacted the head of the programme which the participants were attending. During an initial meeting, the researcher explained the study to the head of program, and asked for permission to employ participants. The researcher was introduced to the students by the head of programme before the start of one of the classes. The experiment was explained briefly by the researcher, and the learners were given the opportunity to express interest and register their names to participate. They were also informed that they could change their minds at any time, and the experiment would not affect their marks at the programme. While the researcher was running the tests at INTO, she had the opportunity to employ more learners until the maximum number of 60 participants was reached.

The sampling frame of this study specifies that 50 percent of the participants should have IELTS scores of 5.5 or less, and the other 50 percent should have IELTS scores of more than 5.5. The data was collected over five weeks for each participant, and recruiting participants was going on at the same time. The process of addressing INTO students and recruiting more participants from the same population continued over two terms. Over that time, the word of mouth among the participants and the time the researcher spent with them helped in selecting more EAP learners who met the certain practical criteria of selecting participants. Data collection continued until the total number of participants was completed for each proficiency level group in both Split-Attention and Integrated groups.

The researcher explained to all the participants about the aims of this study and made it very clear that the results of the tests were anonymous, and the only two people who could have access to the identities of participants were the researcher and the programmer who wrote the website. The participants were fully aware that the results of the tests would not be revealed to English language teachers at INTO, and they would not affect their marks at the programs they were studying at INTO.

3.4.2 Research design

The method used in order to investigate the topic of this study is quasi-experimental. This method was used to collect quantitative data needed for the research. The following is a brief

discussion and rationale for choosing to design a quasi-experimental intervention study to collect data for this research.

In most educational settings, it is rarely possible for researchers to randomly assign participants into groups; thus, researchers mostly use a quasi-experimental design. Quasi-experiments and experiments are similar in everything except for the random assignment factor. Whereas experimentation is defined by Campbell and Stanley (1963) as "the portion of research in which variables are manipulated and their effects upon other variables observed", using quasi-experiments is described by Cook and Campbell (1979) as a design researchers use when they do not randomly assign participants into groups to make comparisons to help them conclude the change caused by the treatment.

Many approaches to teaching methods are adopted in classrooms by teachers and teaching assistants. While some approaches used in classrooms are sometimes reported to be beneficial and fruitful, others are reported to have controversial results.

In the field of education, researchers sometimes try to investigate the effects of certain teaching methods looking for the different effects of the method in hand, and trying to build on and improve that method suggesting solutions to some problems and giving recommendations for other researchers and teaching staff.

While qualitative research involves asking the participants about their understandings, opinions and response towards the problem being investigated, quantitative research can objectively trace the phenomena looking at explanations and causality through experimentation. If the dependent and independent variables in any research are defined, the independent variables are manipulated and the rival variables are controlled, researchers can define and explain the causality in their research very clearly and objectively. "While the other research designs provide useful information, the experimental design provides the most rigorous test of a hypothesis which specifies that X causes Y." (Bouma et al., 1995). Despite all this, sometimes there are practical restraints which force the researcher to work with non-equivalent groups. In these cases, the causal claims can be based on a quasi-experimental approach; however, the initial group differences should be considered in these cases. Heinsman and Shadish (1996) made a meta-analysis comparison between the full experimental and the quasi-experimental approaches, and they concluded that if both approaches were well designed in equal ways, they would deliver comparable results. The results of the meta-analysis stressed two particular ways which can improve the design of quasi-experimental studies: first, avoiding the situations where

the participants can choose themselves or volunteer to be in the treatment group they choose; second, making sure the pre-test differences between the groups are kept to the minimum.

Although adopting the experimental approach can help researchers in the field of education, and although it can highlight, clarify and explain many debated issues in the field of education, there are some limitations of using the experimentation method. According to Campbell and Stanley (1963), "it is a refining process superimposed upon the probably valuable cumulations of wise practice". Since it is well known that the ideal setting for experimentation is a laboratory where the subjects are inanimate and all the variables can be controlled perfectly, adopting the experimentation approach with humans requires a lot of ethical and legal considerations. However, in the field of education, participants are mostly students, teachers or administrative staff. This means they are rational humans who are aware of the aim of the research they are participating in, the variables which affect the results and findings of the experiment to a large extent.

Moreover, although a full 'pre-test post-test control-group design' can control the various threats to the internal validity of the experiment (Dörnyei, 2010), this design can reduce the external validity and the generalizability of the study as a result of trying to control each and every variable in the experiment by following highly artificial frameworks in laboratory conditions (Clarke, 2004). On the other hand, a quasi-experimental design threatens the validity of a study; however, Christensen (2004) confirms that even if a threat to validity is possible, it is not necessarily plausible. A properly designed and administered quasi-experimental study could result in scientifically credible findings.

Building on what was mentioned above, the quasi-experimental design for the sake of this research took into consideration the interests and circumstances of the participants. Moreover, the validity and reliability of the research procedures, the instruments used in collecting data, and the data analysis tools were checked very carefully by piloting the experiment with students and teachers of English as a second language, and by making sure all the technological issues found in the website were solved properly.

3.4.3 Data reliability

The term reliability refers to the "consistencies of data, scores or observations obtained using elicitation instruments, which can include a range of tools from standardized tests administered in educational settings to tasks completed by participants in a research study" (Chalhoub-

Deville, 2006). In other words, reliability refers to the consistency of the results if the tool (the treatment) is run again with the use of similar methodology, and if the population studied is represented accurately again. According to Joppe (2000), if the results of a study can be reproduced by the use of a similar methodology, the study can be counted as reliable.

To obtain the experiment's reliability, the researcher used more than one item in each test of the treatment to measure the same variable instead of using only one question or one of testing item. Moreover, the treatment was repeated three times with three different reading texts and vocabulary knowledge exercises, and results were compared using descriptive and inferential statistics. This is important in terms of meeting the criteria of the internal consistency reliability.

According to Bachman (2004b), it is required by the professional international standards that 'the reliability of each total score, sub-score or combination of scores that is to be interpreted' should be estimated and reported by researchers (AERA, APA, and NCME., 1999, p. 31). To do this, the general approach described in detail by Bachman (2004b) was done by the researcher to estimate the reliability of data in this research.

Before running the reliability tests, the collected data was prepared for analysis by reducing the number of test questions and dividing them into three variables with equal numbers of questions in each variable; thus, the sixty questions in each of the Pre- and Post-tests which aimed at testing learners' vocabulary knowledge were reduced to three variables; namely, word meaning, word families, and collocations, with the answers of twenty questions forming each variable.

Internal consistency reliability was measured by using the Cronbach Alpha coefficient for Pre-test and Post-test. Table 3-1 and Table 3-2 below refer to the internal consistency reliability measurement of the pre-test and post-test used in the study for all groups. Reliability coefficients were .84 and .82 respectively, so both tests could be considered as reliable tests. Usually, when Cronbach Alpha is used to test the reliability of a questionnaire, it is only used with the scales that make up the variables tested by the questionnaire; however, with the pre- and post-tests in this research, Cronbach Alpha was used even with the whole tests since the main objective of the sixty questions of the pre- and post-tests is to test the participants' vocabulary before and after the treatment.

Table 3-1 Pre-test reliability

Reliability Statistics	
Cronbach's Alpha	N of Items

.838	60

Table 3-2 Post-test reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.817	60

In addition to using Cronbach Alpha with the pre- and post-tests, it was used with vocabulary testing questions of each treatment session, as referred to in table 3-3, table 3-4, and table 3-5 below. Reliability coefficients were .71, .72, and .63 of Test1, Test2, and Test3 respectively, so the three tests could be considered as reliable tests.

Table 3-3 Test 1 Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.713	30

Table 3-4 Test 2 Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.717	30

Table 3-5 Test 3 Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.629	30

Reliability coefficients were less than .7 but still over .6 for the three variables; namely, word meaning, word families and collocations in Pre- and Post-tests. Lower reliability coefficients in the three factors can be caused by the number of items which constitute each variable. Whereas the whole number of items in the whole pre- and post-tests is 60, and the whole number of item in each within-treatment-test is 30, the number of items in each variable is 20. In their study which aimed at assessing the reliability and internal consistency of a multi-dimensional scale by Coefficient Alpha, Serbetar and Sedlar (2016) showed that a higher overall value of

Cronbach Alpha (.83) could be secured by eliminating some items. Tables 3-6, 3-7, and 3-8 refer to reliability coefficients for the three factors of knowing a word focused on in this study; namely, word meaning, word families, and collocations in the pre-test. 20 questions are assigned to test each of these factors.

Table 3-6 Pre Word Meaning Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.656	20

Table 3-7 Pre Word Families Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.671	20

Table 3-8 Pre Collocations Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.633	20

Tables 3-9, 3-10, and 3-11 refer to reliability coefficients for the three factors of knowing a word focused on in this study; namely, word meaning, word families, and collocations in the post-test. 20 questions are assigned to test each of these factors.

Table 3-9 Post Word Meaning Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.604	20

Table 3-10 Post Word Families Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.657	20

Table 3-11 Post Collocations Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
.605	20

3.4.4 Validity

Validity in the quantitative research involves two parallel components; one involves 'construct validity', and the other involves the 'internal-external validity' dichotomy (Dörnyei, 2010). In order to assess the validity of a certain research, both research validity and measurement validity should be looked at.

3.4.4.1 Research validity

Campbell and Stanley (1963) distinguish 'internal research validity' from 'external research validity', where the internal validity focusses on whether the results of a study are really caused by the variables and the treatment factors of that study, and the external validity is involved with the generalizability of the results of the study (Bachman, 2004a).

For the sake of this research, the researcher made sure she ruled out potential threats to validity by excluding any factors which might affect and invalidate the results of the study, and by using what is believed to be a suitable research tool for data collection. To maintain the external validity, any participant who had special circumstances which might affect the generalizability of the results was excluded from the pool of participants. For example, a participant who was found to have lived in the UK when they were younger was excluded from the pool of participants because they were exposed to the English language differently from the other participants at a younger age. The were excluded while employing other participants was going on; thus, the proportions were maintained properly between groups. Moreover, although there were some drop outs during the experiment, the researcher made sure the data analyzed was collected from groups with proportionate characteristics. It can be argued that the external validity element of the trustworthiness of this study was maintained because the EAP learners who participated in the study can be considered to be a representative of the EAP population in the UK. First, the EAP learners who participated in each L1 group came from different environments whether in the Arabic or the Chinese speaking countries. For example, it is well known for the researcher that Arabic L1 speakers come from countries with different cultures; thus, a Syrian is not the same as a Libyan culture. However, the participants came from different

Arabic speaking countries, so their results could be generalised to the Arabic speaking population. Second, it is realised by the researcher that INTO students come from a population which is able to pay the tuition; however, some participants are at INTO because they are high achievers on a scholarship, and not because they are rich. These two varieties mean that not all the participants come from the same economic or educational background. This is also believed to make the results of the study more transferable to the wider population; and thus, more trustworthy.

The Hawthorne effect, where the performance of participants tends to be different when they realize they are being studied, is believed to be one of the threats to research validity (Dörnyei, 2010). To minimize the Hawthorne effect, the researcher did not tell the participants that the results of participants with different proficiency levels or with different L1 would be compared.

3.4.4.2 Measurement validity

Measurement validity refers to the degree of accuracy with which the used instrument measures what it is supposed to measure (Weir, 2005, Hughes, 2003). According to Henning (1987), validity is the "… appropriateness of a given instrument or any of its component parts as a measure of what it is proposed to measure". However, there has been a change of this concept over the past decades (Dörnyei, 2010). In the 1960s, validity was considered to be one of the characteristics of a language test (Chapelle, 1999), thus many types of validity were defined: 'criterion validity' which refers to the correlation of the test used with another similar tool, 'content validity' which is maintained by having the test content judged by experts in the field, and 'construct validity' which refers to whether the test outcomes comply with the theoretical background according to which the test is designed. However, the previous definition of three validities was replaced with a unitary concept 'construct validity' in 1985 in the main international guidelines for educational and psychological measurement sponsored by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education-the AERA/APA/NCME Standards for Educational and Psychological Testing (AERA, APA, and NCME, 1999). This change resulted from the change of looking at validity as a quality of the test to looking at it as the credibility of the interpretation of the test scores. According to Lynch (2001), 'When examining the validity of assessment, it is important to remember that validity is a property of the conclusion, interpretations or inferences that we draw from the assessment instruments and procedures, not the instruments and procedures themselves'. As a result of the new approach to the concept of validity, content- and criterion- related validity were considered to contribute to the concept of

construct validity in addition to interpretation and use of scores. It can be said here that construct validity can be mainly considered as a general term explaining the process of theory validation which include clearly identified test validation operations (Smith, 2005).

To assess the content validity of the online experiment as the only source of data designed for this study, the final version of the website was reviewed by three technicians and web designers, two faculty experts in the field of Applied Linguistics and TESOL, and in Second Language Acquisition. Faculty experts stated that the website was designed properly with clear language and instructions used. They also stated that the pre-test and post-test are designed adequately for the purpose of the research, and that the questions equally tested the three word knowledge factors which were meant to be tested; namely, word meaning, word families, and collocations. They approved that the tests following or integrated into each reading text included test items focussing on the three word knowledge factors focused on in the study. They also double checked that each word tested in the pre-test and post-test existed in one of the three texts and the questions following or integrated into them.

3.5 Data Collection Steps

Collecting the quantitative data was done by the use of intervention. The intervention included a pre- and post-tests treatment. By post-test, it is meant that an immediate post-test was conducted. All experiments were held at INTO, Newcastle.

The rationale behind adopting quasi-experimentation as the main method of investigation in this study is to examine whether vocabulary knowledge of adult EAP learners is built more effectively when the online reading text they are exposed to is in the integrated format than when it is in the split-attention format, and also to examine whether the proficiency levels of those learners and their first languages (L1) would be factors affecting their vocabulary knowledge.

Assigning the participants into low proficiency level learners and high proficiency level ones was according to their IELTS scores, so sixty participants were randomly assigned to four groups for the different levels of the treatment.

a. 15 low proficiency level learners exposed to split-attention format of online reading text.

b. 15 high proficiency level learners exposed to split-attention format of online reading text.

c. 15 low proficiency level learners exposed to integrated format of online reading text.

d. 15 high proficiency level learners exposed to integrated format of online reading text.

The L1 make up of all the four groups was 35 Arabic L1 speakers and 25 Chinese L1 speakers.

While the split-attention format involves vocabulary testing questions following the online reading text, the integrated format involves vocabulary testing questions integrated into the reading text dividing it into shorter paragraphs (between three and five lines each). The reading texts used consisted of about nine paragraphs each; otherwise they would be too long for learners with limited or modest linguistic skill levels.

The aim of pre- and post- testing all four groups and comparing their test results was to know whether integrated format online reading texts would decrease cognitive load, and thus help learners build their vocabulary knowledge more effectively than those exposed to split-attention format online reading texts which were hypothesised to cause cognitive overload. On the other hand, pre- and post- testing the four groups aimed at giving a clearer idea whether both formats of reading texts would have similar or different effects on both low and high proficiency level students.

These findings could not be achieved if different measuring instruments to measure breadth and depth of vocabulary in pre- and post- tests were used because using different measuring instruments would affect the internal validity of the study, as stated by Campbell and Stanley (1963). On the contrary, obtaining the data for this research was achieved by using the same measuring instruments in both pre- and post- tests, which were intended to measure the breadth and depth of vocabulary both before and after the treatment.

The treatment involved five sessions. Data was collected in available rooms or computer clusters at INTO. The researcher was with the participants for the whole time during each session, and she logged them in without giving them their passwords because their access to the program without the interpreter guiding the experiment would be a real threat to the validity of the test. The data was collected in groups, but each individual would be doing something

different from the others depending on the date they started and on which session they reached. All the participants from both groups started with a sixty-question pre-test, and ended with the post-test which involved the same sixty questions.

Before the administration of the pre-test, all the groups were introduced to the website which included the bank of reading texts in both integrated and split-attention formats. Reading texts were chosen from Schmitt et al. (2011a). There was a selection of reading texts out of which the researcher chose three for this study. The texts and the vocabulary exercises were written in the website which was designed especially for the sake of this study, and called The Vocabulary Learning Assistant. Link for the website is: http://vocabulary-language-assistant.com/ (The original link had "vocabulary-learning-assistant" since this is the name which was given to the programme, but due to technical issues, the website was rebuilt and the programmer gave it this name. Figure 3-1 shows the website main page.

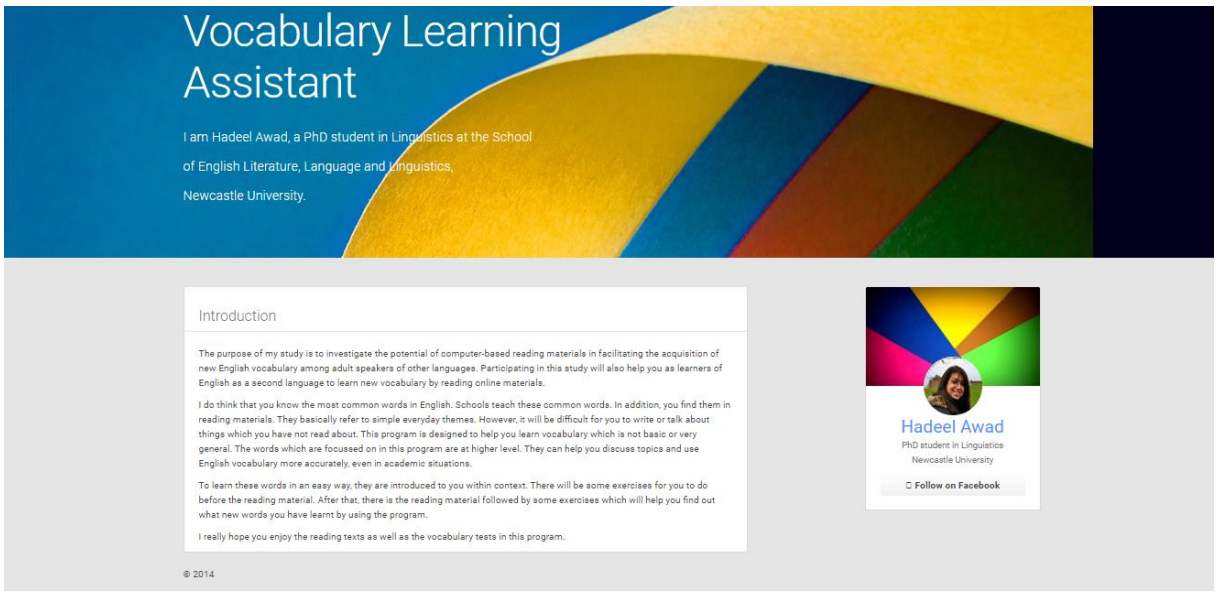


Figure 3-1 Website main page

Deciding on writing the reading text and the accompanying tests in a website involved, among other issues, taking into consideration the feasibility and practicality of inserting the vocabulary exercises after or within the texts without splitting them from each other; otherwise answering the vocabulary exercises would be a memorization test, the thing which was avoided because the study was not intended to test participants' memorization skills. It should be mentioned here that treatment for all groups was not only exposing them to one reading text. Treatment involved exposing learners in both integrated and split-attention groups to three reading texts and accompanying questions. The texts and the questions were the same for both groups. The only

difference was the instructional formats. Learners in the split-attention group read a text every week and answered the vocabulary questions located after the text. Learners in the integrated group read the same reading text split into smaller paragraphs with the vocabulary questions inserted between the paragraphs also on a weekly basis. Another important point to be mentioned here is that the experiment was not run during teaching hours. Participants volunteered to do it outside their learning hours. Because the study involved repeated testing over five weeks while the participants were attending classes, practice effect could affect the validity of the research. Practice effect occurs when the learners are exposed to the same items being researched during the course of the treatment; however, in the case of this research, the vocabulary used is not specific to any major or to any environment, and the words are not repeated every session; thus, the possibility of encountering the same words tested could not be high. In addition, encountering the same words in the test and the classroom is possible among the EAP learners, so it can be generalised into the wider population. Running the experiment was monitored by the researcher. The data collected from participants who did not complete the experiment was taken out of the data pool. However, to minimise the probability of participants not completing the experiment, the researcher offered a prize draw of shopping gift vouchers for the learners who completed the exercises over the five weeks. Providing the participants with a voucher can be seen as a threat to validity; however, the voucher was not given to the learners who answered correctly. This is why it can be argued that it should not be a threat to the validity of the research. In contrast to that, presenting the learners who complete the five weeks with a voucher can be argued to be a factor to encourage the participants not to drop out; and thus reduce participants mortality or attrition which is considered to be one of the main threats to research validity in studies in which collecting different sets of data, such as pre-test, post-test, or multiple tests is involved (Dörnyei, 2010).

3.6 Piloting

According to Mackey and Gass (2005), piloting is essential for first testing and then finalizing the methods and materials used in any study, in addition to ensuring that each measure will be useful in addressing the research questions in hand.

All the items used in the website designed for the sake of the experiment were piloted before they were presented to the participants in this study. Some of the criteria taken into consideration during piloting are: wording of the treatment instructions, their clarity and simplicity, in addition to the time needed to answer the questions and participate in the

treatment. Following this, the researcher and the programmer made some changes to the software after piloting it with friends and colleagues. A major change was made to the introductory page as well as the login page of the website. The layout of both pages was found to be dull and not attractive. Thus, both pages layout was changed into a more coloured and user-friendly design, with some illustrative photos to make it more attractive to the students.

Another point related to the content was the wording of the instructions. Some instructions were seen to be somehow complicated for the lower proficiency level students. So, they were changed by the researcher to be clearer and more understandable for the learners. An example of the changes administered is the following. Within the texts, there are multiple choice questions asking the participants to choose the right meaning of the word in the text. In the multiple options, there is sometimes more than one correct answer, but only one of them is correct in the reading text. Questions like these cannot be in the pre-test or the post-test where there is no reading text to choose the most appropriate answer from. Despite this fact, there were three questions with more than one correct answer in the pre- and post-tests. Because such questions require a reading test to conclude the meaning of the word from, they were replaced by other questions with one possible answer for each.

Technically, the website was found to be working well on all web browsers.

3.7 Technical Features

3.7.1 Software requirement specification (SRS)

Experiments were implemented online. For the sake of the treatment, a website was designed and written with the help of a specialized programmer.

As a result of piloting, some issues were taken into consideration while editing the website. One of the most important issues was the lack of control on some variables because the exercises were intended to be done as homework with no direct monitoring.

In order to minimise the trials for obtaining extra help from other resources, the researcher confirmed to students that the reading texts and the vocabulary exercises were not to test their English proficiency, and they were not related to their exam results. Students were also informed that it was always better for teachers as well as the researcher to know what

knowledge they lacked in order to help them improve their vocabulary knowledge. So, learners were encouraged to complete all exercises honestly.

3.7.2 Online test system purpose

Since the purpose of this study is the investigation of the impact of different online reading formats on building adult EAP learners' vocabulary knowledge, the purpose of the project is to provide an online facility for the researcher to conduct an on-screen experiment, and to participants to be able to take the treatment as an extra activity online rather than taking it during the teaching hours.

All the reading tests and vocabulary exercises can be modified and edited easily by the researcher. In addition, the lists and groups of students can be controlled online in terms of entering all the learners' detailed information. In addition, the answers can be saved in an Sql Server Database (.mdf file). Researcher can control the content by using the dashboard designed for this purpose.

3.7.3 Scope

The scope of this project is very broad in terms of other manually taken tests. Although it is intended to be used in an educational institution, it can as well be used in corporate world. It can also be used anywhere and anytime as it is a web-based application (user location does not matter). Moreover, there are no restrictions that force the researcher to be present while participants are taking the experiment.

3.7.4 Features

Some features of the software are:

- It is secure and controlled by the researcher herself who sorts out all participants' profiles. After listing all the participants' names and details, they were all given login details to be able to register their personal accounts and login to the website to do their homework.
- It is easy to use and very user friendly. The font, pictures and colours used are suitable for the purpose of the website. Clear font against colourful background is used for the introductory page. There are some inviting educational pictures which make the website look very user friendly. Instructions are clear and participants have found it easy to

follow the steps. However, during piloting, some changes were being made to the instructions when they were found to be somehow complicated for lower level learners.

- It gives reliable and accurate results since all the information were saved and sorted within the assigned database, and then were exported to SPSS for analysis.

3.7.5 User types and roles

- **Anonymous users:** who can only access the introductory page of the website and read the general aims of it, but cannot see any of the tests.
- **Participants:** who each have a username and password. Participants are allowed to login to their dashboard, where they can take the tests and read the reading texts in the format assigned for them by the researcher. Because participants were divided into two groups, and each group was assigned certain online reading format, half of them could access only the split-attention formatted reading texts and exercises, and the other half could only be able to access the integrated format texts and exercises. They could also view their answers and had a feedback of right and wrong answers from the researcher after the experiment was finished.
- **Researcher:** who is the administrator of the website. She can do the following:
  - Manage participants' groups: The researcher divided the participants into groups of two different proficiency levels and assigned each group a certain format of the reading texts and exercises. She administered the four groups separately.
  - Manage participants individually: Each learner's work was managed, and the results were sorted and saved in the assigned database.
  - Conduct test type and reading texts types after choosing them so that texts can be edited and modified online before they are activated to be seen by learners.
  - Activate each test at when needed so that participants can take it, and deactivate it again when finished with.

3.7.6 Technologies used

Front end as: HTML & JavaScript

Back end as: ASP.NET and C# using VS2010 and .NET Framework 4.0

Server: IIS 7.0

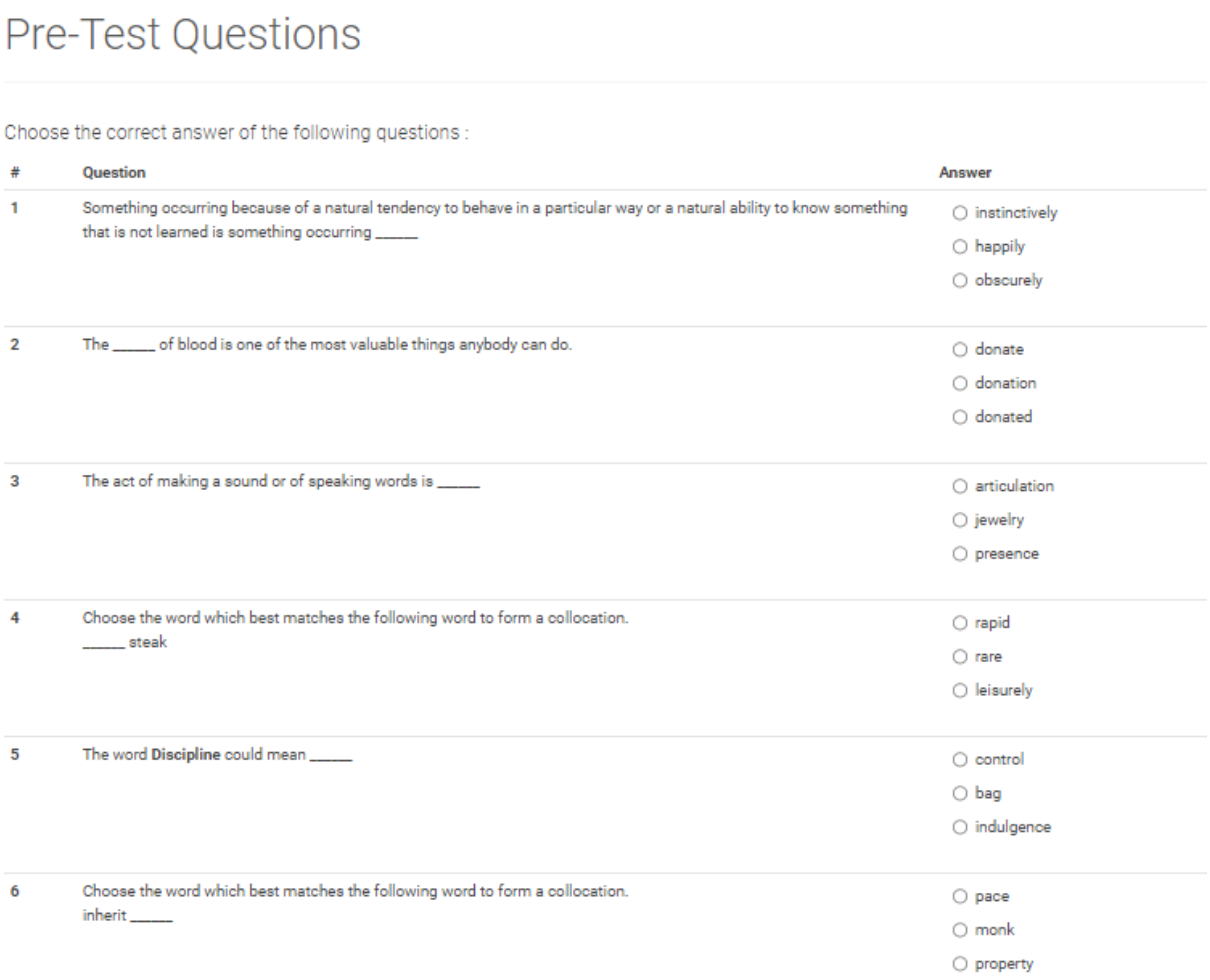
Database: Microsoft SQL Server 2008

Querying language: SQL

3.7.7 Description of the website and the tests

The website involves an introductory page containing the researcher's personal information and some points about the research itself and its aims. The first page can be seen by anonymous users, but only the participants in the research can register and login to their personal web pages where they can start taking tests.

When learners log in, they choose the type of experiment assigned for them; namely, split-attention or integrated formats. Before they can access the reading text, there is a pre-knowledge collection of questions. Learners are given a list of words and a list of choices to match with them. Choices range on a scale of four choices from "I don't know this word" until "I know this word and can use it in my own speaking and writing". Learners then receive a confirmation message of willingness to submit their answers before they can move into the next step which is a further investigation of their knowledge of the target words. An important point here is that participants are not given the chance to navigate between the pages until they have finished the pre-test and submitted it. Learners also receive an alerting red shade on any missing answer. This step is to help avoid receiving unanswered questions. Figure 3-2 below is a screenshot showing 6 out of the 60 questions in the pre-test.



Pre-Test Questions

Choose the correct answer of the following questions :

#	Question	Answer
1	Something occurring because of a natural tendency to behave in a particular way or a natural ability to know something that is not learned is something occurring _____	○ instinctively ○ happily ○ obscurely
2	The _____ of blood is one of the most valuable things anybody can do.	○ donate ○ donation ○ donated
3	The act of making a sound or of speaking words is _____	○ articulation ○ jewelry ○ presence
4	Choose the word which best matches the following word to form a collocation. _____ steak	○ rapid ○ rare ○ leisurely
5	The word **Discipline** could mean _____	○ control ○ bag ○ indulgence
6	Choose the word which best matches the following word to form a collocation. inherit _____	○ pace ○ monk ○ property

Figure 3-2 An example of the questions in the Pre-test

After submitting the answers of the vocabulary knowledge test, participants have the choice to move to the next page which includes the reading text in either split-attention or integrated formats. Each group can see their own format of the reading text. On one hand, the group of learners taking the integrated format are able to read the text divided into shorter paragraphs with integrated vocabulary exercises to test the target vocabulary. Figure 3-3 below is an example of integrated instructional format where a paragraph from the reading text appears on the screen with the first question next to it. When the participants answer the question, they choose the following question. Questions keep on appearing next to the paragraph until the questions about the next paragraph start. At that point, the paragraph automatically changes.

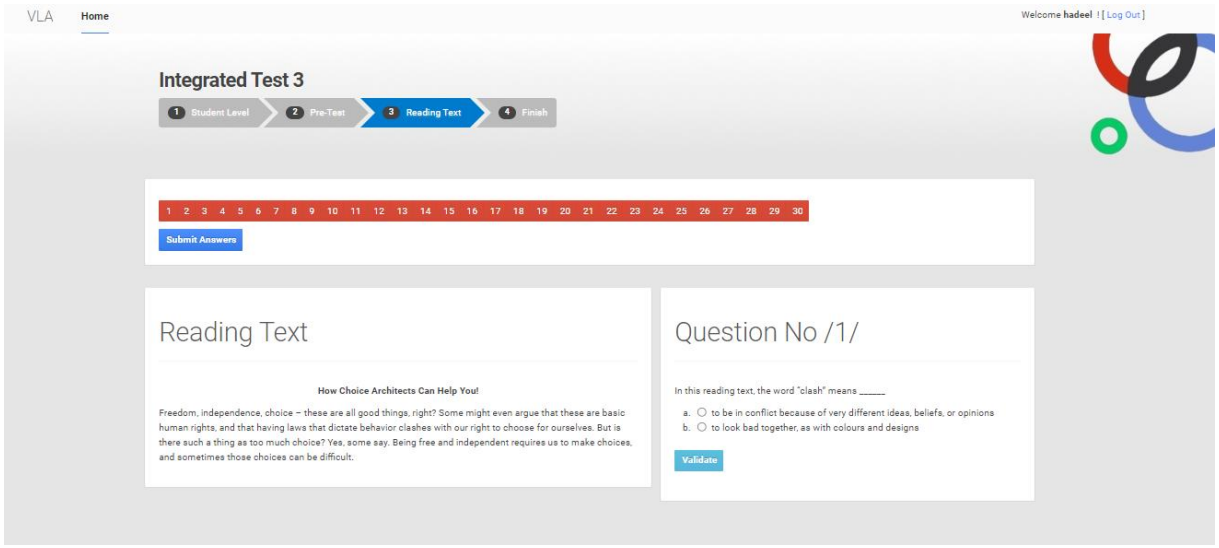


Figure 3-3 An example of integrated instructional format

On the other hand, participants taking the split-attention format reading texts are able to read the text as a whole before they start answering the exercises. Figure 3-4 below is an example of split-attention instructional format where the reading text as a whole appears on screen and the question is next to it. After participants answer the question, they choose the following one from the list of questions on top of the page.

After learners answer each question, they validate their answer and move to the next one. Again, participants are not given the choice to navigate between the pages while they are taking the treatment in order not to be able to go back to the pre-test and look at the answers or even change them.

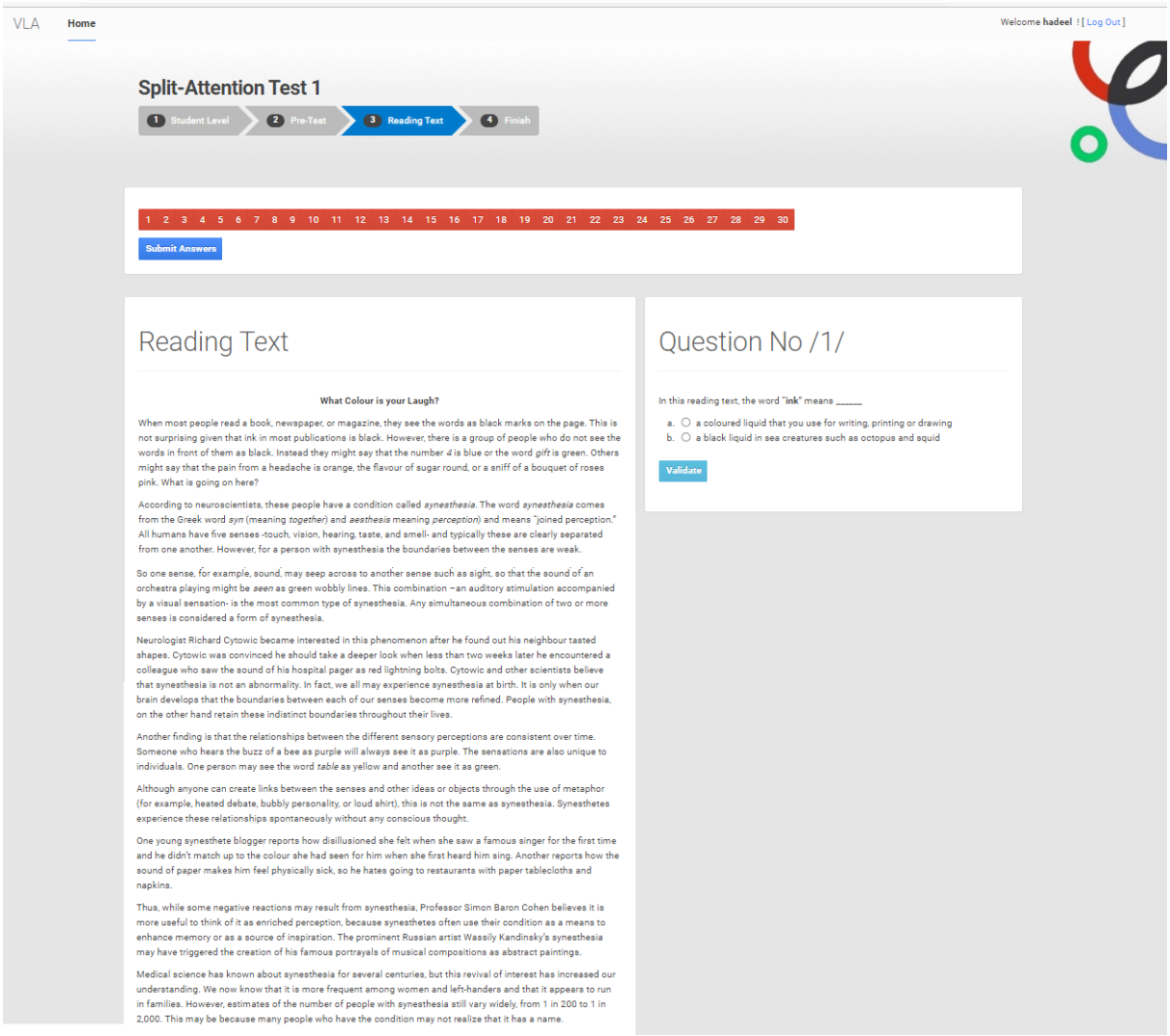


Figure 3-4 An example of split-attention instructional format

Vocabulary exercises in both reading text formats are the same. The vocabulary exercises in this study follow Schmitt and Schmitt (2011)'s model of concentrating on the three aspects of word meaning, word families (derivative forms of words), and collocations in order to achieve the maximum teaching goal of the exercises. According to Schmitt and Zimmerman (2002), learners do not necessarily know all of the categories of a word family although they may be aware of some of its forms. Nevertheless, learners should know the right form of a word (noun, verb, adjective, or adverb) in a certain context, and hence the focus on derivative forms of words in this study. Learning word collocations is also very essential when learning new vocabulary. Although it is hard to teach all word pairings, integrating the element of collocations knowledge into the exercises in this study aims at helping the learners become more aware of collocations and thus enhancing their intuitive understanding of collocations for individual words.

Word meaning knowledge is tested by the use of two exercises. In the first exercise, learners are given fourteen phrases, and they have to choose one from three target words to define each phrase. In the second exercise, learners are given the target word with four choices out of which they need to choose the word which is not a synonym of that target word. The exercise involves nine target vocabulary.

Word family knowledge is tested by giving the learners ten sentences which contain words written in bold. With each of the ten sentences, there is another sentence with a space to be filled by another correct form of the word written in bold belonging to the same word family and completing the sentence correctly.

Collocations knowledge is tested by giving learners six target words and other six words that can be matched to them to form a correct collocation. Learners are asked to match the words to each other so that they make correct collocations out of them.

For each of the exercises, learners are given a solved example to make the instructions clearer for them.

After piloting, changes were made to the exercises by the researcher and the programmer so that they were introduced to the learners in the best and the most attractive way to encourage them to continue participating in the experiment till the end, and thus help the researcher collect the data needed to finish this study. For example, the writing was plain when the website was first designed without any illustrations or colours. After receiving some feedback about how dull the pages were, some colourful illustrations were added to the pages. It was taken into consideration though not to make the design too colourful or distracting to the participants.

A flowchart showing how the participants in both groups progressed through the website is in Figure 3-5.

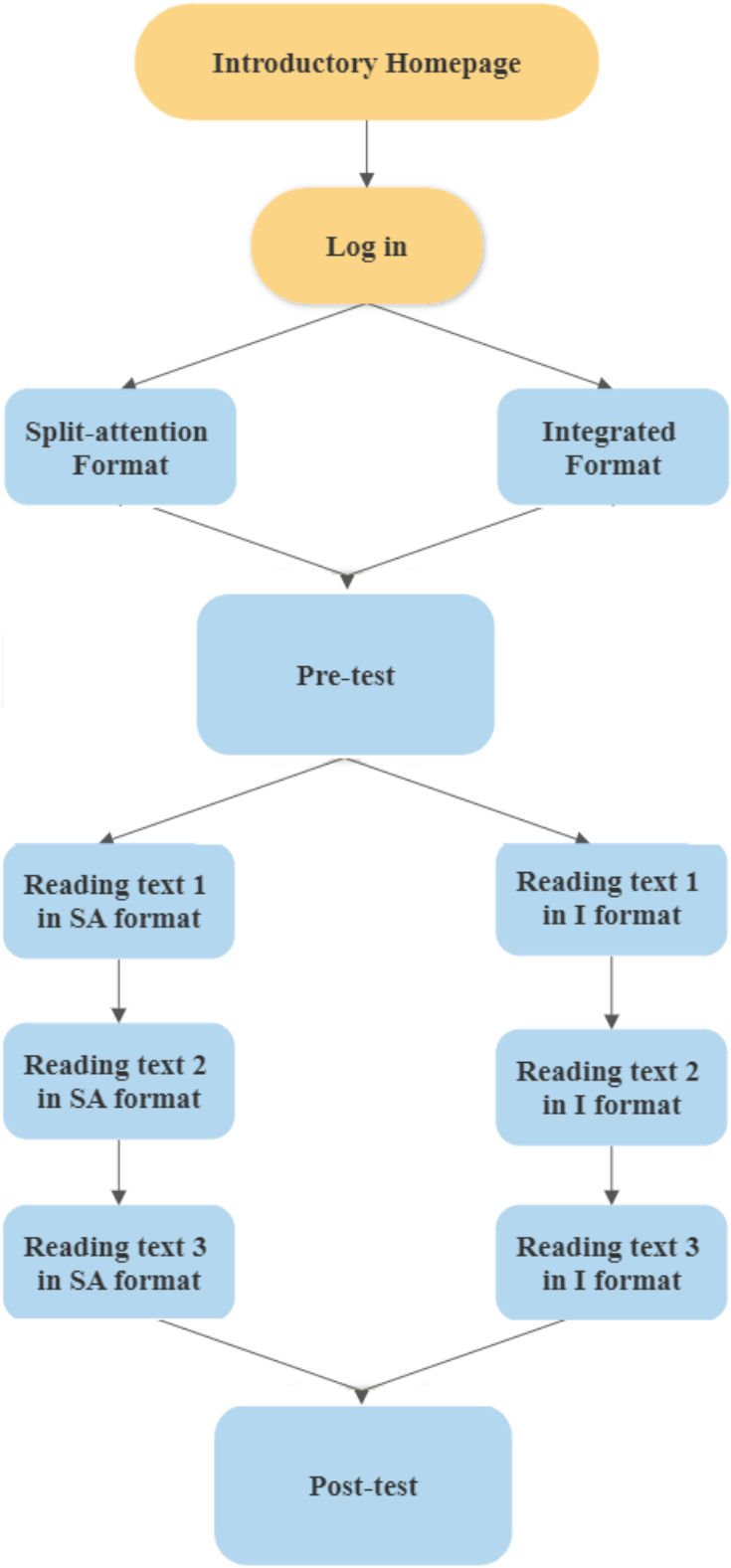


Figure 3-5 How learners progressed through the website

3.8 Analysis of the Data

Data analysis methods used in this research were chosen building on the nature of the research questions leading the study, and the data collection methods followed by the researcher. The experiment test answers were analysed quantitatively using IBM SPSS Statistics. Frequency, percentages and means were calculated. Crosstabulation, Chi Square, Paired-Samples T-test, Wilcoxon Signed Rank Test, Independent-Samples T-test, Mann-Whitney U Test, One-Way Between-Groups Analysis of Variance, and Kruskal-Wallis Test were conducted to test if there is any statistically significant relationship between various data categories. The findings and results of the study were stated and illustrated by tables and/or graphs.

3.8.1 Descriptive analysis

3.8.1.1 Frequency

Frequency analysis is a descriptive statistical method which displays the number of occurrences of each response chosen by the respondents. It is used for describing categorical data. In this study, Frequency was used to check the categorical variables, such as L1 frequency, proficiency level frequency, the frequency of the recoded scores of vocabulary knowledge in pre- and post-tests, the frequency of the recoded scores of the word meaning knowledge items in pre- and post-tests, the frequency of the recoded scores of the word family knowledge items in pre- and post-tests, and the frequency of the recoded scores of the collocations knowledge items in pre- and post-tests.

3.8.1.2 Descriptives

Descriptives analysis is used for continuous variables. It provides the researcher with ‘summary’ statistics such as mean, median and standard deviation. Descriptives also helps with some information about the distribution of scores on continuous variables (skewness and kurtosis). This information is particularly needed if the variables are being used with parametric statistical techniques, such as t-tests or analysis of variance. While the Skewness value indicates the symmetry of the distribution, Kurtosis indicates how much peaked the distribution is. If skewness and kurtosis value is 0, which is very uncommon in social sciences, the distribution is perfectly normal.

Whereas positive skewness values mean positive skew or scores clustered to the left at the low values, negative skewness values mean clustered scores to the right at the high values. While positive kurtosis values mean that the distribution is clustered in the centre or rather peaked,

negative kurtosis values mean that the distribution of the scores is fairly flat with a lot of cases in the extremes.

In this study, Descriptives analysis was used for the vocabulary knowledge reflected in the total scores of the pre- and post-test, the word meaning knowledge reflected in the total scores of the pre- and post-test, the word family knowledge reflected in the total scores of the pre- and post-test, the collocations knowledge reflected in the total scores of the pre- and post-test.

3.8.1.3 Explore to assess normality

The word 'normal' is used to report a symmetrical, bell-shaped curve with the greatest frequency in the centre and the smaller frequencies extended to the extremes (Gravetter and Wallnau, 2004). Although skewness and kurtosis values can help with assessing normality, the Explore option in SPSS is a more accurate way of reporting the normality of the data. The results of the Kolmogorov-Smirnov statistic assess the normality of the distribution of scores. Normality is indicated by a statistically non-significant result (Sig. value of more than .05).

In this study, Kolmogorov-Smirnov analysis was used for the normality of the data collected from the pre- and post-tests testing vocabulary knowledge, and the normality of the data collected from the items testing word meaning knowledge, word family knowledge, and collocations knowledge reflected in the scores of the pre- and post-tests.

3.8.1.4 Chi-square test for independence

Chi-square test for independence is used to explore the relationship between two categorical variables. Each variable can have two or more categories. Chi-square test provides a comparison between the observed frequencies with the values which would be expected if there was no association between the two variables being measured. It is displayed in a crosstabulation table. When there are two categories in each variable, an additional correction value; namely, Yates' Correction for Continuity, is included in the output from Chi-square. One of the assumptions of chi-square regarding the 'minimum expected cell frequency' is that the frequency should be 5 or greater in at least 80 percent of cells. If this assumption is violated, Fisher's Exact Test should be referred to. The main value in the input of Chi-square is Pearson Chi-square; however, in 2 by 2 tables, Continuity Correction should be used. In order for the result to be statistically significant, Sig. value should be .05 or smaller.

In this study, Chi-square was used to investigate the association between the different reading instructional formats and second language vocabulary acquisition including second language word meaning learning, word family learning, and collocations learning.

3.8.2 Inferential analysis

3.8.2.1 Paired-samples t-test

A paired-samples t-test is used when data is collected from one group of people under two different conditions. This technique works well in situations of pre-test/post-test experimental designs if the data collected is normally distributed. This test tells the researcher whether there is a statistically significant difference in the mean scores of the different conditions. Sig. (2-tailed) is the probability (p) value. If p is less than .05, it can be concluded that there is a statistically significant difference between the two scores.

In this study, a paired-samples t-test was used to investigate the effect of the treatment on general vocabulary knowledge building.

3.8.2.2 Wilcoxon Signed Rank test

The Wilcoxon Signed Rank test is a non-parametric test used with repeated measures; that is, when the participants are measured under two different conditions if the data collected is not normally distributed. This test is used as an alternative to a Paired-Samples T-test which compares the means of two tests. Instead Wilcoxon Signed Rank compares the ranks. The two things to look at in the output are the Z value and the significance level which is displayed as Asymp.Sig (2-tailed). If the probability value, P, is less than or equal to .05, the results of the test are statistically significant.

In this study, Wilcoxon Signed Rank test was used to investigate the effect of the treatment on word meaning knowledge, word family knowledge, and collocations knowledge since the data collected for these items is not normally distributed.

3.8.2.3 Independent-samples t-test

An independent samples t-test is used for comparing the mean score, on a continuous variable, for two different groups of participants when the data collected is normally distributed. The test tells whether there is a statistically significant difference in the mean scores for the two groups. The results of an independent-samples t-test display the mean and standard deviation for each of the groups studied. The first section of the output box displays the results of Leven's test for

equality of variance. Levene's test reveals whether the variance of scores for both groups is the same. The result of this test determines which t-value is the right one to be used. If the Sig. value for Levene's test is larger than .05, the first line in the table should be used. To assess the difference between the groups, Sig. (2-tailed) is used. If this value is equal to or less than .05, there is a statistically significant difference in the mean scores.

In this study, an independent-samples t-test was used to investigate the effect of the online reading split=attention and integrated instructional formats on vocabulary knowledge building.

3.8.2.4 Mann-Whitney U test

The Mann-Whitney U test is a non-parametric test used to investigate the differences between two independent groups on a continuous measure if the data collected is not normally distributed. This test is used as an alternative to Independent-Samples T-test which compares the means of two tests. Instead Mann-Whitney U compares the medians. The main values to look at in the output are the Z value and the significance level, which is displayed as Asymp.Sig (2-tailed). If the probability value, P, is less than or equal to .05, the results of the test are statistically significant.

In this study, Mann-Whitney U test was used to investigate the effect of the online reading different instructional formats on word meaning knowledge, word family knowledge, and collocations knowledge since the data collected for these items is not normally distributed.

3.8.2.5 One-way between-groups analysis of variance

One-way between-groups ANOVA is used when there is one independent variable with three or more groups and one dependent continuous variable if the data collected is normally distributed. 'One-way' indicates that there is only one independent variable, and 'between-groups' indicates that there are different participants in each group. The test reveals whether there are statistically significant differences in the mean scores on the dependent variable across the three or more groups.

The results table informs the researcher about information of each group. The homogeneity of variance displays Levene's test for homogeneity of variances, which tests whether the variance in scores for each one of the groups is the same. If the significance value for Levene's test is greater than .05, the assumption of homogeneity of variance in the test is not violated.

Regarding the statistical significance of the results, the p value should be considered. If Sig. is equal to or less than .05, there is a statistically significant difference somewhere among the mean scores on the dependent variable for the three or more groups.

In this study, a one-way between-groups analysis of variance test was used to explore the effect of different proficiency levels on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats, and also to explore the effect of L1 on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats

3.8.2.6 Kruskal-Wallis Test

The Kruskal-Wallis test is the non-parametric alternative to a one-way between-groups analysis of variance. It facilitates comparing the scores on some continuous variable for three or more groups. Scores are converted to ranks and the mean rank for each group is compared. This type of analysis is a between-groups analysis so different people from each of the different groups compared. The main values to focus on from the output of this test are: Chi-Square value, the degree of freedom (df) and the significance level (Asymp. Sig). If the significance level is less than .05, it can be concluded that there is a statistically significant difference in the continuous variable across the groups.

In this study, Kruskal-Wallis test was used to investigate the effect of different proficiency levels on word meaning, word family, and collocations acquisition in both online reading split-attention and integrated instructional formats. It was also used to investigate the effect of L1 on word meaning, word family, and collocations acquisition in both online reading split-attention and integrated instructional formats. The reason of using Kruskal-Wallis is that the data collected in the groups is not normally distributed.

3.9 Ethical Issues

In accordance with the ethical guidelines of Newcastle University, privacy and confidentiality of the participants information and their test results were considered during the whole research process**.**

The researcher met the participants at the start of one of their classes, and was introduced by the class teacher. The aim and nature of the research were thoroughly explained. The students were asked to volunteer in the experiment and were told that their results would not be shared with any of the INTO staff, and would not affect their INTO curriculum tests. They were also

told that the experiment would take place out of class time, so it would not affect their actual program schedule. Participants were also informed that they could withdraw from the experiment at any time, but were told that there was a shopping voucher prize draw at the end of the experiment for those who completed the five sessions. All of the guarantees of anonymity and freedom to withdraw were verbally given at the initial introduction of the research and were repeated immediately before the treatment.

3.10 Summary

This chapter has explained the research focus, clarified the research questions, and outlined the methodological approach, research design and the data collection procedures in detail. A quantitative approach was adopted in an attempt to fill in a gap in literature about the effect of different instructional formats of reading texts on vocabulary learning among EAP adult learners. The quantitative data was collected from INTO EAP students by logging in to the experiment website designed by the researcher and developed by a specialized programmer. Data validity and reliability were considered, achieved and tested. Cronbach Alpha test showed reliability of at least .6 for all tests and variables used in the treatment. Finally, ethical considerations were taken throughout the whole research process to make sure that the data from all participants was kept private and confidential, and that all participants participated willingly and felt free to opt out of the research any time they wanted to.

Chapter 4. Results

4.1 Introduction

This chapter provides a presentation of the description, analysis and results of the quantitative data collected by the use of quasi-experimentation method. The chapter starts with an overview of the research topic and the research questions presented in Ch.3 before addressing these questions by stating and explaining the quantitative data findings.

4.2 Overview

The main purpose of this study is to investigate whether there is a relation between cognitive load and building vocabulary knowledge; thus, whether vocabulary among EAP students can be learnt more effectively depending on the different instructional formats of the online reading texts; namely, split-attention and integrated instructional formats.

The dependant variables in this study are the total scores of the pre-test, post-test, and vocabulary tests accompanying reading texts in both integrated and split-attention formats. The vocabulary tests aim at testing three aspects of acquiring new vocabulary, so they test: word meaning knowledge, word family knowledge, and collocations knowledge.

The independent variables in the study are:

- Reading instructional formats; namely, integrated format and split-attention format.
- Participants' linguistic proficiency level where >5.5 IELTS score is considered lower proficiency level, and <5.5 IELTS score is considered higher proficiency level for the sake of this study.
- Participants' first language (L1) which is either Arabic or Chinese.

Therefore, the study aims at answering the following major research question, noted in 3,

1. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their vocabulary knowledge?

1.1. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word meaning knowledge?

1.2. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word families knowledge?

1.3. Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their collocations knowledge?

2. Do adult EAP learners at different proficiency levels acquire new vocabulary items from the reading text in both split-attention and integrated formats differently?

3. Do adult EAP learners whose first language is Arabic acquire new vocabulary items from the reading text in both split-attention and integrated formats differently from adult EAP learners whose first language is Chinese?

4.3 Data Preparation

4.3.1 Coding and inputting the data

The data for this study was all collected online, and the results were all saved in a Structured Query Language (SQL) database. SQL is a domain-specific language used in programming and designed for managing data held in a relational database management system. They were also displayed as a table on a dashboard which could be seen only by the researcher and the programmer. Figure 4-1 below shows an example of the results displayed in the database. The green fields represent correct answers, and the red ones represent incorrect answers.

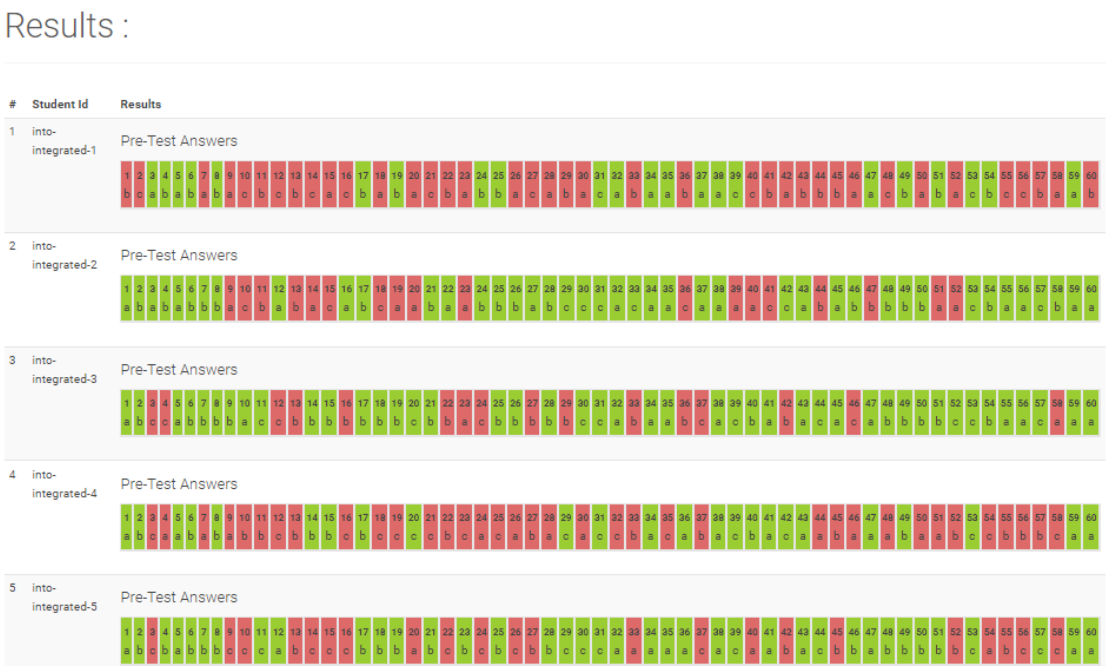


Figure 4-1 an example of the results as displayed in the database

The results, as could be seen on the website, were filtered by test, participant, or reading group. Figure 4-2 below shows the page of the tests which can be seen by using any of the filters, so they can be filtered by test, students groups, or students.

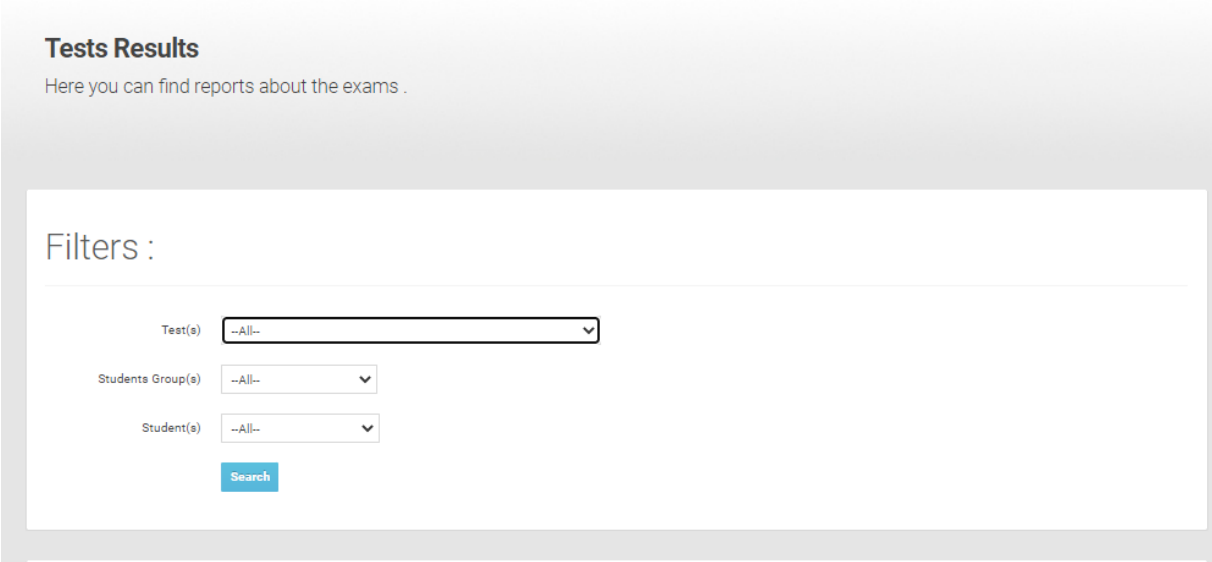


Figure 4-2 Test filters

Each correct answer in the tests was given the score of 1, and each wrong answer was given the score of 0. The maximum score in the pre- and post-tests is 60, whereas the maximum score in each of the three tests within the treatment is 30. To analyse the data, the results were all exported as a (.XLS) file, which in turn was exported into SPSS.

4.3.2 Data screening and cleaning

Before analysing the data, it was checked for errors as an extra check although no impossible data or incorrectly entered values were expected because the data was all transferred electronically from the website where it was collected. A Frequency test was run nonetheless, and the data had no out-of-range values, incorrectly entered values, or contradicting data.

There were drop out cases who were deleted from the whole list of participants and replaced by others during the data collection time. Originally, there were 35 participants in the “Integrated” group, and 34 participants in the “Split-attention” group. 9 participants from both groups dropped out, so they were replaced by others who met the requirements of the experiment in order to have 30 participants in each group.

4.3.3 Handling missing data

The website designed for this research was programmed so that participants were notified if they missed an answer; however, they were still given the choice not to answer. The main aim of the notification is reminding the participants in case they forgot to answer the question, but at the same time encouraging them to answer all the questions without putting any pressure on them. Luckily, there were not missing answers.

4.3.4 Recoding the data

To prepare the data for descriptive analysis, the total scores of Pre-test, post-test, and each of the word meaning, word family, and collocation knowledge dependant variables in both pre- and post-tests were calculated before the mean of each test total score was calculated using SPSS. Afterwards, the data from each test was recoded into nominal variables where total scores with the range 0 - 0.44 were given the value 0, and total scores with the range 0.45 – 1 were given the value 1.

4.4 Descriptive Findings

4.4.1 First language (L1) and linguistic proficiency

Table 4-1 L1 Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Arabic	35	58.3	58.3	58.3
	Chinese	25	41.7	41.7	100.0
	Total	60	100.0	100.0	

Table 4-2 Proficiency Level Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	High	30	50.0	50.0	50.0
	Low	30	50.0	50.0	100.0
	Total	60	100.0	100.0	

On one hand, the output of the proficiency frequency is expected because the sampling procedures aimed at 30 participants in each group, out of which there were 50% participants with >5.5 IELTS score and 50% participants with = or <5.5 IELTS score. On the other hand,

the output of L1 frequency, shows that there are 35 (58.3 per cent) speakers of Arabic as a first language, and 25 (41.7 per cent) speakers of Chinese as a first language.

4.4.2 Vocabulary knowledge in pre- and post-tests

Table 4-3 Vocabulary knowledge in Pre-test

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	8	13.3	13.3	13.3
	1	52	86.7	86.7	100.0
	Total	60	100.0	100.0	

Table 4-4 Vocabulary Knowledge in Post-test

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	6.7	6.7	6.7
	1	56	93.3	93.3	100.0
	Total	60	100.0	100.0	

The output of the frequency of the recoded scores of vocabulary knowledge in pre- and post-tests shows that whereas 86.7 per cent of the participants in both split-attention and integrated groups answered correctly in the pre-test, 93.3 per cent of the participants answered correctly in the post-test, so there was 6.6 per cent increase in the number of participants who answered correctly after the treatment.

Table 4-5 Descriptive Statistics for Pre- and Post-tests

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pre Total	60	17.00	56.00	36.2667	8.60206	-.158	.309	-.412	.608
Post Total	60	17.00	56.00	39.5667	7.96468	-.350	.309	.033	.608
Valid N (listwise)	60								

The descriptive statistics of the vocabulary knowledge reflected in pre- and post-tests show that the maximum total score of correct answers in both pre- and post-tests was 56 (out of 60), and the minimum total score of correct answers in both tests was 17 (out of 60) with a mean of 36.27 and a standard deviation of 8.6 for pre-test, and a higher mean of 39.57 with less standard deviation of 7.96 for post-test. The total score of the pre-test was slightly skewed, with negative skewness of -0.16 (SE = 0.31) and negative kurtosis of -0.41 (SE = 0.61). The total score of the

post-test was also slightly skewed with negative skewness of -0.35 (SE = 0.31) and positive kurtosis of 0.03 (SE = 0.61).

Table 4-6 Tests of Normality for Pre-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
PreTotal	.054	60	.200*	.990	60	.900
*. This is a lower bound of the true significance.						
a. Lilliefors Significance Correction						

Table 4-7 Tests of Normality for Post-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
PostTotal	.078	60	.200*	.986	60	.717
*. This is a lower bound of the true significance.						
a. Lilliefors Significance Correction						

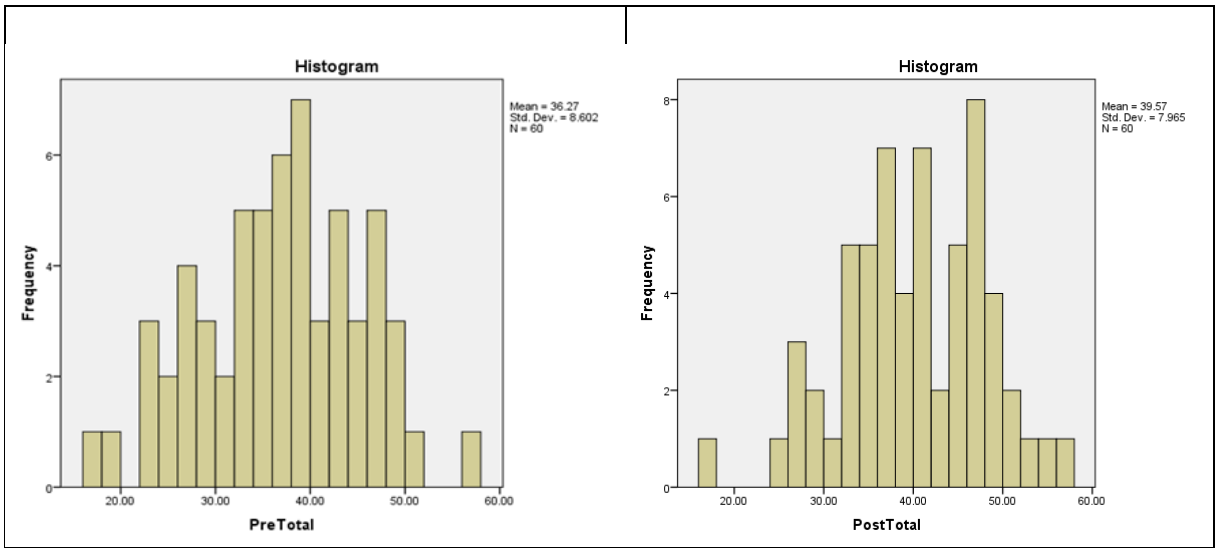


Figure 4-3 Normality for pre and post test

The statistically non-significant result of the Kolmogorov-Smirnov normality test of 0.2 for both pre- and post-tests results indicates that the data collected from the pre- and post-tests is normally distributed.

4.4.3 Word meaning in pre- and post-tests

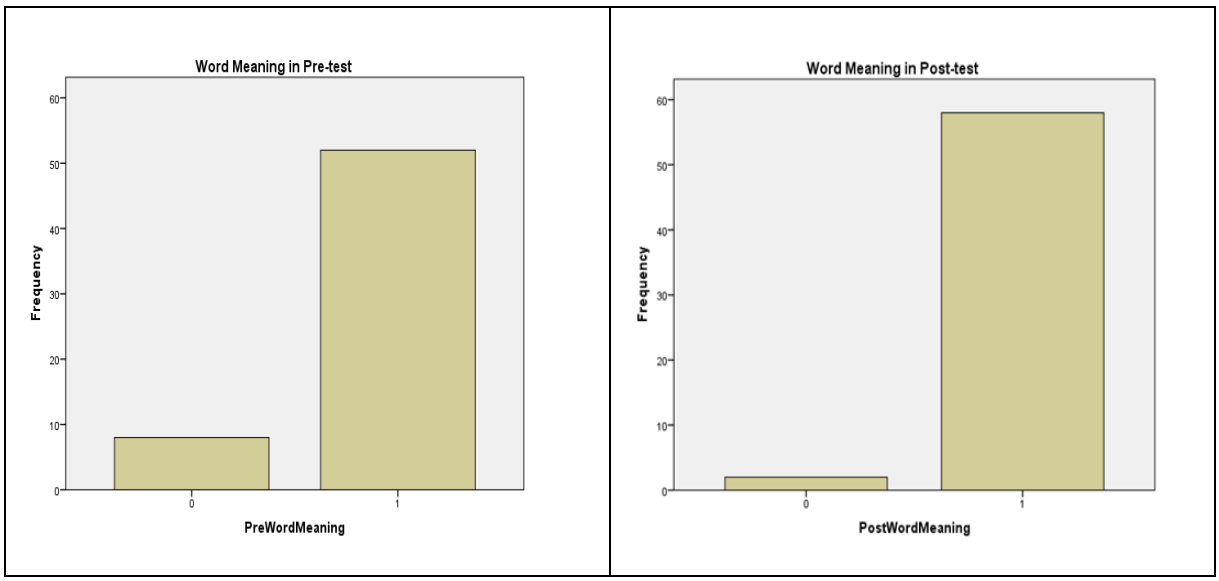


Figure 4-4 Word meaning in pre- and post-tests

The output of the frequency of the recoded scores of word meaning knowledge in pre- and post-tests shows that whereas 86.7 per cent of the participants in both split-attention and integrated groups answered correctly in the pre-test, 96.7 per cent of the participants answered correctly in the post-test., so there was 10 per cent increase in the number of participants who answered the word meaning variable questions correctly after the treatment.

Table 4-8 Descriptive Statistics for Word Meaning Pre- and Post-tests

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pre Word Meaning Total	60	6.00	20.00	12.2833	3.42519	.213	.309	-.418	.608
Post Word Meaning Total	60	6.00	19.00	14.3833	3.00334	-.558	.309	-.244	.608
Valid N (listwise)	60								

The descriptive statistics of the word meaning knowledge reflected in pre- and post-tests show that whereas the maximum total score of correct answers in word meaning pre-test was 20 (out of 20), the maximum total score of correct answers in the word meaning post-test was 19 (out of 20); however, the minimum total score of correct answers in both tests was 6 (out of 20) with a mean of 12.28 and a standard deviation of 3.43 for word meaning pre-test, and a higher mean

of 14.38 with less standard deviation of 3 for word meaning post-test. The total score of the word meaning pre-test was slightly skewed, with positive skewness of 0.21 (SE = 0.31) and negative kurtosis of -0.42 (SE = 0.61). The total score of the word meaning post-test was moderately skewed with negative skewness of -0.56 (SE = 0.31) and negative kurtosis of -0.24 (SE = 0.61).

Table 4-9 Tests of Normality for Word Meaning Pre-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
Pre Word Meaning Total	.116	60	.042	.970	60	.148
a. Lilliefors Significance Correction						

Table 4-10 Tests of Normality for Word Meaning Post-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
Post Word Meaning Total	.155	60	.001	.952	60	.019
a. Lilliefors Significance Correction						

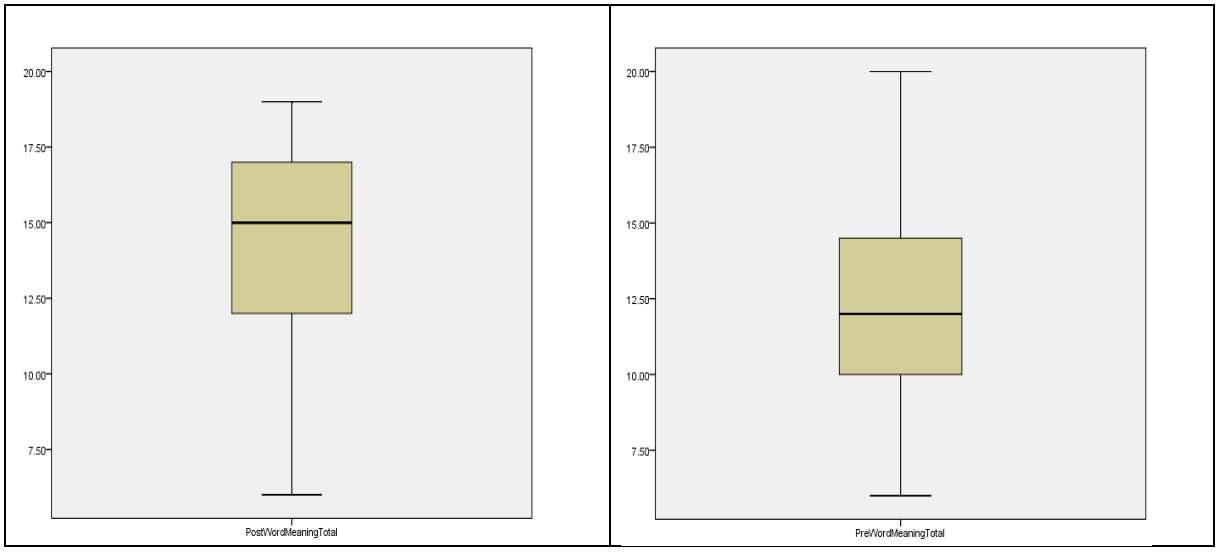


Figure 4-5 Normality for word meaning pre and post test

The statistically significant result of the Kolmogorov-Smirnov normality test of 0.04 for word meaning pre-test and 0.001 for word meaning post-test results indicates that the data collected from the word meaning pre- and post-tests is not normally distributed.

4.4.4 Word family in pre- and post-tests

Table 4-11 Word Family in Pre-test

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	15	25.0	25.0	25.0
	1	45	75.0	75.0	100.0
	Total	60	100.0	100.0	

Table 4-12 Word Family in Post-test

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	13	21.7	21.7	21.7
	1	47	78.3	78.3	100.0
	Total	60	100.0	100.0	

The output of the frequency of the recoded scores of word family knowledge in pre- and post-tests shows that whereas 75 per cent of the participants in both split-attention and integrated groups answered correctly in the pre-test, 78.3 per cent of the participants answered correctly in the post-test., so there was 3.3 per cent increase in the number of participants who answered the word family variable questions correctly after the treatment.

Table 4-13 Descriptive Statistics for Word Family Pre- and Post-tests

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pre Word Family Total	60	5.00	17.00	11.0833	3.52854	-.251	.309	-.955	.608
Post Word Family Total	60	5.00	18.00	11.3667	3.48832	-.130	.309	-.755	.608
Valid N (listwise)	60								

The descriptive statistics of the word family knowledge reflected in pre- and post-tests show that whereas the maximum total score of correct answers in word family pre-test was 17 (out of 20), the maximum total score of correct answers in the word family post-test was 18 (out of 20); however, the minimum total score of correct answers in both tests was 5 (out of 20) with a

mean of 11.08 and a standard deviation of 3.53 for word family pre-test, and a mean of 11.37 and a standard deviation of 3.49 for word family post-test. The total score of the word family pre-test was slightly skewed, with negative skewness of -0.25 (SE = 0.31) and negative kurtosis of -0.96 (SE = 0.61). The total score of the word meaning post-test was also slightly skewed with negative skewness of -0.13 (SE = 0.31) and negative kurtosis of -0.76 (SE = .61).

Table 4-14 Tests of Normality for Word Family Pre-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
Pre Word Family Total	.152	60	.001	.947	60	.011
a. Lilliefors Significance Correction						

Tests of Normality for Word Family Post-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
Post Word Family Total	.122	60	.027	.969	60	.128
a. Lilliefors Significance Correction						

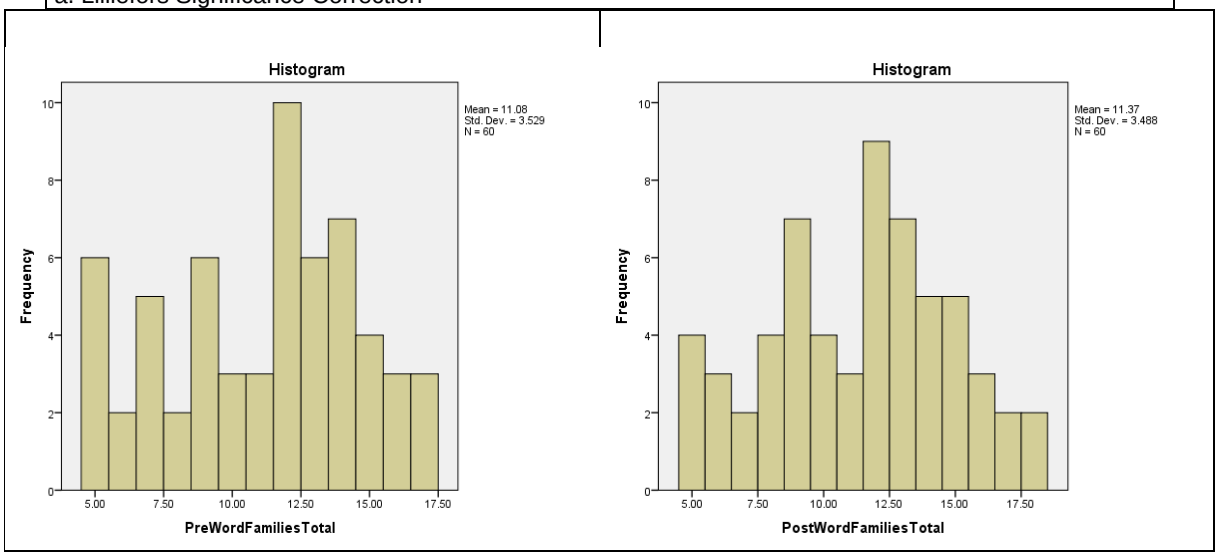


Figure 4-6 Normality for word families pre and post test

The statistically significant result of the Kolmogorov-Smirnov normality test of 0.001 for word family pre-test and 0.03 for word family post-test results indicates that the data collected from the word family pre- and post-tests is not normally distributed.

4.4.5 Collocations in pre- and post-tests

Table 4-15 Collocations in Pre-test

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	6.7	6.7	6.7
	1	56	93.3	93.3	100.0
	Total	60	100.0	100.0	

Table 4-16 Collocations in Post-test

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	6.7	6.7	6.7
	1	56	93.3	93.3	100.0
	Total	60	100.0	100.0	

The output of the frequency of the recoded scores of collocations knowledge in pre- and post-tests shows that 93.3 per cent of the participants in both split-attention and integrated groups answered correctly in both the pre-test and the post-test.

Table 4-17 Descriptive Statistics for Collocations Pre- and Post-tests

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pre Collocations Total	60	4.00	20.00	12.9000	3.21385	-.277	.309	.335	.608
Post Collocations Total	60	6.00	20.00	13.8167	3.08903	-.400	.309	-.225	.608
Valid N (listwise)	60								

The descriptive statistics of the collocations knowledge reflected in pre- and post-tests show that the maximum total score of correct answers in both collocations pre-test and post-test was 20 (out of 20); however, whereas the minimum total score of correct answers in the collocations pre-test was 4 (out of 20), the same score was 6 (out of 20) in the collocations post-test with a mean of 12.9 and a standard deviation of 3.21 for collocations pre-test, and a higher mean of 13.82 and a standard deviation of 3.09 for collocations post-test. The total score of the collocations pre-test was slightly skewed, with negative skewness of -0.28 (SE = 0.31) and

positive kurtosis of 0.34 (SE = 0.61). The total score of the collocations post-test was also slightly skewed with negative skewness of -0.4 (SE = 0.31) and negative kurtosis of -0.23 (SE = 0.61).

Table 4-18 Tests of Normality for Collocations Pre-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Pre Collocations Total	.112	60	.057	.976	60	.282
a. Lilliefors Significance Correction						

Table 4-19Tests of Normality for Collocations Post-test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
Post Collocations Total	.157	60	.001	.967	60	.107
a. Lilliefors Significance Correction						

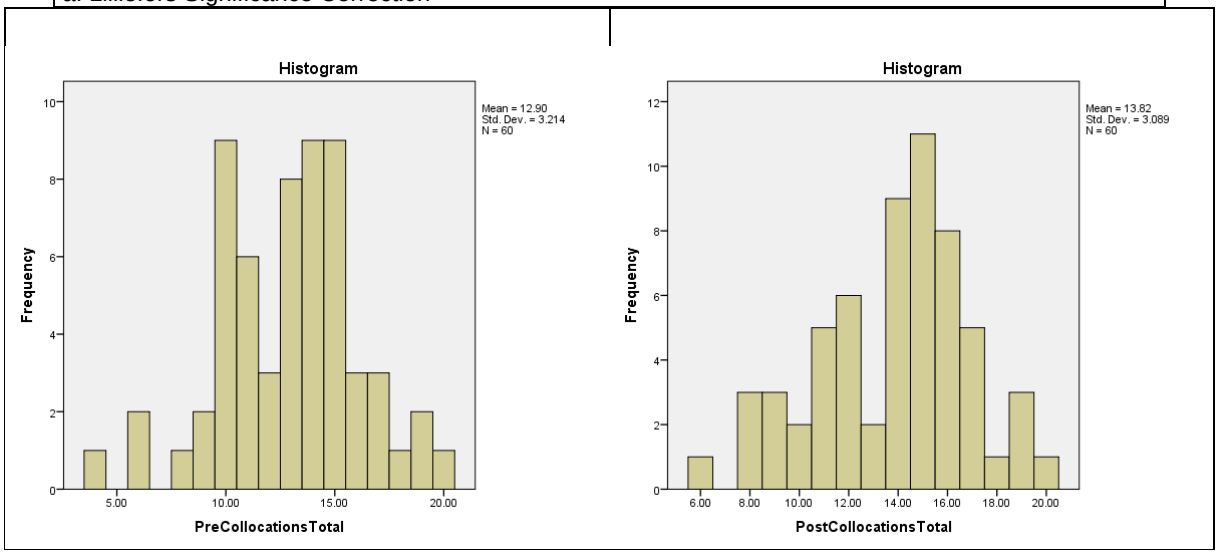


Figure 4-7 Normality for collocations pre and post test

The statistically non-significant result of the Kolmogorov-Smirnov normality test of 0.06 for collocations pre-test indicates that the data collected from the collocations pre-test is normally distributed, while the statistically significant result of 0.001 for collocations post-test indicates that the data collected from the collocations post-test is not normally distributed.

4.4.6 Is there an association between different reading instructional formats and second language vocabulary learning?

A crosstabulation test showed that 100 % of participants in the integrated group answered the post-test correctly, while 86.7 % of participants in split-attention group answered correctly;

however, a Chi-square test for independence (with Yates Continuity Correction) indicated no statistically significant association between different reading instructional formats and second language vocabulary learning, p = .12, *phi* = -.27. Because the table violates the Chi-square assumption of having at least 80% of cells counting more than 5, the Fisher's Exact Test was considered where P= .112.

Table 4-20 Chi-Square Tests for Groups and Post-test

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.286[a]	1	.038		
Continuity Correction[b]	2.411	1	.121		
Likelihood Ratio	5.831	1	.016		
Fisher's Exact Test				.112	.056
Linear-by-Linear Association	4.214	1	.040		
N of Valid Cases	60				
	a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 2.00.				
	b. Computed only for a 2x2 table				

Table 4-21 Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-.267	.038
	Cramer's V	.267	.038
N of Valid Cases		60	

4.4.7 Is there an association between different reading instructional formats and second language word meaning learning?

A Crosstabulation test showed that 100 % of participants in the integrated group answered the word meaning post-test correctly, while 93.3 % of participants in split-attention group answered correctly; however, a Chi-square test for independence (with Yates Continuity Correction) indicated no statistically significant association between different reading instructional formats and second language word meaning learning, p = .47, *phi* = -.19. Because the table violates the Chi-square assumption of having at least 80% of cells counting more than 5, the Fisher's Exact Test was considered where P= .492.

Table 4-22 Chi-Square Tests for Groups and Word Meaning Post-test

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.069[a]	1	.150		
Continuity Correction[b]	.517	1	.472		
Likelihood Ratio	2.842	1	.092		
Fisher's Exact Test				.492	.246
Linear-by-Linear Association	2.034	1	.154		
N of Valid Cases	60				
a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 1.00.					
b. Computed only for a 2x2 table					

Table 4-23 Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-.186	.150
	Cramer's V	.186	.150
N of Valid Cases		60	

4.4.8 Is there an association between different reading instructional formats and second language word family learning?

A Crosstabulation test showed that 86.7 % of participants in the integrated group answered the word family post-test correctly, while only 70 % of participants in split-attention group answered correctly; however, a Chi-square test for independence (with Yates Continuity Correction) indicated no statistically significant association between different reading instructional formats and second language word family learning, $p = .21$, $phi = -.2$

Table 4-24 Chi-Square Tests for Groups and Word Family Post-test

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.455[a]	1	.117		
Continuity Correction[b]	1.571	1	.210		
Likelihood Ratio	2.506	1	.113		
Fisher's Exact Test				.209	.105
Linear-by-Linear Association	2.414	1	.120		
N of Valid Cases	60				
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.50.					

b. Computed only for a 2x2 table

Table 4-25 Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-.202	.117
	Cramer's V	.202	.117
N of Valid Cases		60	

4.4.9 Is there an association between different reading instructional formats and second language collocations learning?

A Crosstabulation test showed that 96.7 % of participants in the integrated group answered the collocations post-test correctly, while only 90 % of participants in split-attention group answered correctly; however, a Chi-square test for independence (with Yates Continuity Correction) indicated no statistically significant association between different reading instructional formats and second language word family learning, $p = .61$, $phi = -.13$

Table 4-26 Chi-Square Tests for Groups and Collocations Post-test

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.071[a]	1	.301		
Continuity Correction[b]	.268	1	.605		
Likelihood Ratio	1.118	1	.290		
Fisher's Exact Test				.612	.306
Linear-by-Linear Association	1.054	1	.305		
N of Valid Cases	60				

Table 4-27 Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-.134	.301
	Cramer's V	.134	.301
N of Valid Cases		60	

4.5 Inferential Findings

4.5.1 The effect of the treatment on general vocabulary knowledge building

Because the comparison is for the same group, a paired-samples t-test was conducted to evaluate the effect of the treatment on participants' scores of the vocabulary test. There was a statistically significant increase in the test scores from pre-test (M = 36.3, SD = 8.6) to post-test (M = 39.6, SD = 7.96), t (59) = 4.9, p < .001 (two-tailed). The main increase in test scores was 3.3 with a 95% confidence interval ranging from 1.95 to 4.64.

Table 4-28 Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PostTotal	39.5667	60	7.96468	1.02824
	PreTotal	36.2667	60	8.60206	1.11052

Table 4-29 Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	PostTotal – PreTotal	3.30000	5.22153	.67410	1.95113	4.64887	4.895	59	.000

4.5.2 The effect of the treatment on word meaning

Because the data collected from the word meaning pre- and post-tests was not normally distributed, a non- parametric test (Wilcoxon Signed Rank Test) was used to investigate the effect of the treatment on word meaning test scores.

A Wilcoxon Signed Rank Test revealed a statistically significant increase in word meaning test scores following the treatment, z = -4.8, p < *.001*. The median of the word meaning total score increased from word meaning pre-test (MD = 12) to word meaning post-test (MD = 15).

Table 4-30 Test Statistics[a]

	Post-Word Meaning Total – Pre-Word Meaning Total
Z	-4.801[b]
Asymp. Sig. (2-tailed)	.000
a. Wilcoxon Signed Ranks Test	
b. Based on negative ranks.	

Table 4-31 Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
Pre-Word Meaning Total	60	10.0000	12.0000	14.7500
Post-Word Meaning Total	60	12.0000	15.0000	17.0000

4.5.3 The effect of the treatment on word family

Because the data collected from the word family pre- and post-tests was not normally distributed, a non- parametric test (Wilcoxon Signed Rank Test) was used to investigate the effect of the treatment on word family test scores.

A Wilcoxon Signed Rank Test revealed a statistically non-significant increase in word family test scores following the treatment, z = -.9, p = .39. The median of the word family total score stayed the same (MD = 12) for the word family pre-test and the word family post-test.

Table 4-32 Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
Pre-Word Family Total	60	8.2500	12.0000	14.0000
Post-Word Family Total	60	9.0000	12.0000	14.0000

Table 4-33 Test Statistics[a]

	Post-Word Families Total – Pre-Word Families Total
Z	-.863[b]
Asymp. Sig. (2-tailed)	.388
a. Wilcoxon Signed Ranks Test	
b. Based on negative ranks.	

4.5.4 The effect of the treatment on collocations

Because the data collected from the collocations pre- and post-tests was not normally distributed, a non- parametric test (Wilcoxon Signed Rank Test) was used to investigate the effect of the treatment on collocations test scores.

A Wilcoxon Signed Rank Test revealed a statistically significant increase in collocations test scores following the treatment, z = -2.3, p = .02. The median of the collocations total score increased from collocations pre-test (MD = 13) to collocations post-test (MD = 14).

Table 4-34 Test Statistics[a]

	Post-Collocations Total – Pre-Collocations Total
Z	-2.282[b]
Asymp. Sig. (2-tailed)	.023
a. Wilcoxon Signed Ranks Test	
b. Based on negative ranks.	

Table 4-35 Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
Pre-Collocations Total	60	10.2500	13.0000	15.0000
Post-Collocations Total	60	12.0000	14.0000	16.0000

4.5.5 The effect of the online reading split-attention and integrated instructional formats on vocabulary knowledge building

Because two different groups of participants were being compared, an independent-samples t-test was conducted in order to compare the post-test scores of split-attention group and integrated group. There was no statistically significant difference in scores for integrated group (M = 40, SD = 7.02) and split-attention group (M = 39.13, SD = 8.91); t (58) = .42, p = .68, two-tailed). Mean difference = .87, 95% CI: -3.28 to 5.01. The result of the T-test run to compare the means of the post-test results of the split-attention and integrated groups indicated that although the integrated group had higher scores than the split-attention group in the general vocabulary knowledge post-test, that difference was not statistically significant.

Table 4-36 Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Post-Total	Integrated Group	30	40.0000	7.02213	1.28206
	Split-attention Group	30	39.1333	8.90847	1.62646

Table 4-37 Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means

		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Post-Total	Equal variances assumed	.772	.383	.418	58	.677	.86667	2.07100	-3.27889	5.01222
	Equal variances not assumed			.418	55.000	.677	.86667	2.07100	-3.28371	5.01704

4.5.6 The effect of the online reading split-attention and integrated instructional formats on word meaning knowledge

Because the data collected from the word meaning post-test was not normally distributed, a non- parametric test (Mann-Whitney U Test) was used to investigate the effect of the online reading split-attention and integrated instructional formats on word meaning knowledge.

A Mann-Whitney U Test revealed no statistically significant difference in the word meaning post-test results between the split-attention group (*MD* = 16, *n* = 30) and the integrated group (*MD* = 14, *n* = 30), *U* = 360, *z* = -1.34, *p* = .18; thus, it cannot be claimed that the different reading instructional formats had a statistically significant effect on word meaning post-test scores.

Table 4-38 Report

Post-Word Meaning Total		
Group	N	Median
Integrated Group	30	14.0000
Split-attention Group	30	16.0000
Total	60	15.0000

Table 4-39 Test Statistics[a]

	Post-Word Meaning Total
Mann-Whitney U	360.000
Wilcoxon W	825.000
Z	-1.340
Asymp. Sig. (2-tailed)	.180
a. Grouping Variable: Group	

4.5.7 The effect of the online reading split-attention and integrated instructional formats on word family knowledge

Because the data collected from the word family post-test was not normally distributed, a non-parametric test (Mann-Whitney U Test) was used to investigate the effect of the online reading split-attention and integrated instructional formats on word family knowledge.

A Mann-Whitney U Test revealed no statistically significant difference in the word family post-test results between the split-attention group (*MD* = 10, *n* = 30) and the integrated group (*MD* = 12, *n* = 30), *U* = 346.5, *z* = -1.54, *p* = .12; thus, it cannot be claimed that the different reading instructional formats had a statistically significant effect on word family post-test scores.

Table 4-40 Report

Post-Word Family Total		
Group	N	Median
Integrated Group	30	12.0000
Split-attention Group	30	10.0000
Total	60	12.0000

Table 4-41 Test Statistics[a]

	Post-Word Families Total
Mann-Whitney U	346.500
Wilcoxon W	811.500
Z	-1.537
Asymp. Sig. (2-tailed)	.124
a. Grouping Variable: Group	

4.5.8 The effect of the online reading split-attention and integrated instructional formats on collocations knowledge

Because the data collected from the collocations post-test was not normally distributed, a non-parametric test (Mann-Whitney U Test) was used to investigate the effect of the online reading split-attention and integrated instructional formats on collocations knowledge.

A Mann-Whitney U Test revealed no statistically significant difference in the collocations post-test results between the split-attention group (*MD* = 15, *n* = 30) and the integrated group (*MD* = 14, *n* = 30), *U* = 449.5, *z* = -.007, *p* = .99; thus, it cannot be claimed that the different reading instructional formats had a statistically significant effect on collocations post-test scores.

Table 4-42 Report

Post-Collocations Total		
Group	N	Median
Integrated Group	30	14.0000
Split-attention Group	30	15.0000
Total	60	14.0000

Table 4-43 Test Statistics[a]

	Post-Collocations Total
Mann-Whitney U	449.500
Wilcoxon W	914.500
Z	-.007
Asymp. Sig. (2-tailed)	.994
a. Grouping Variable: Group	

4.5.9 The effect of different proficiency levels on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats

A one-way between-groups analysis of variance was conducted to explore the impact of different proficiency levels on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats. Participants were divided into four groups according to their proficiency level and the treatment they were receiving. There was a statistically significant difference at the $p < .05$ level in general vocabulary knowledge post-total scores for the four groups: F (3, 56) = 4.1, p = .01. Post hoc comparisons using the Tukey HSD test indicated that the mean score for split-attention low proficiency group (M = 35.47, SD = 9.54) was significantly different from integrated high proficiency group (M = 42.93, SD = 7.37), and that the mean score for split-attention low proficiency group (M = 35.47, SD = 9.54) was significantly different from split-attention high proficiency group (M = 42.8, SD = 6.68). The results from this test show that the proficiency level affected the vocabulary post-test scores of the split-attention group significantly, but did not have statistically significant effect on the vocabulary post-test scores of the integrated group.

Table 4-44 ANOVA

Post-Total					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	672.733	3	224.244	4.090	.011
Within Groups	3070.000	56	54.821		

Total	3742.733	59			

Table 4-45 Descriptives

Post-Total							
					95% Confidence Interval for Mean		
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Integrated High	15	42.9333	7.37241	1.90355	38.8506	47.0160	29.00	56.00
Integrated Low	15	37.0667	5.41778	1.39887	34.0664	40.0669	31.00	49.00
Split-Attention High	15	42.8000	6.67832	1.72434	39.1017	46.4983	28.00	54.00
Split-Attention Low	15	35.4667	9.53839	2.46280	30.1845	40.7489	17.00	50.00
Total	60	39.5667	7.96468	1.02824	37.5092	41.6242	17.00	56.00

Table 4-46 Multiple Comparisons

Dependent Variable: Post Total						
Tukey HSD						
					95% Confidence Interval	
(I) Group and Proficiency	(J) Group and Proficiency	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Integrated High	Integrated Low	5.86667	2.70361	.144	-1.2922	13.0255
	Split Attention High	.13333	2.70361	1.000	-7.0255	7.2922
	Split Attention Low	7.46667*	2.70361	.038	.3078	14.6255
Integrated Low	Integrated High	-5.86667	2.70361	.144	-13.0255	1.2922
	Split Attention High	-5.73333	2.70361	.159	-12.8922	1.4255
	Split Attention Low	1.60000	2.70361	.934	-5.5589	8.7589
Split Attention High	Integrated High	-.13333	2.70361	1.000	-7.2922	7.0255
	Integrated Low	5.73333	2.70361	.159	-1.4255	12.8922
	Split Attention Low	7.33333*	2.70361	.043	.1745	14.4922
Split Attention Low	Integrated High	-7.46667*	2.70361	.038	-14.6255	-.3078
	Integrated Low	-1.60000	2.70361	.934	-8.7589	5.5589
	Split Attention High	-7.33333*	2.70361	.043	-14.4922	-.1745
*. The mean difference is significant at the 0.05 level.						

4.5.9.1 The effect of different proficiency levels on word meaning acquisition in both online reading split-attention and integrated instructional formats

Because the data collected from the word meaning post-test was not normally distributed, a non- parametric test (Kruskal-Wallis Test) was used to investigate the effect of different

proficiency levels on word meaning acquisition in both online reading split-attention and integrated instructional formats.

A Kruskal-Wallis test revealed a statistically significant difference in the word meaning post-test total scores across the four groups which were divided according to treatment and different proficiency levels. χ^2 (2, n = 60) = 12.94, p = 0.005. The split-attention group with high proficiency level recorded the highest median score (Md = 42.03) amongst the four groups, while the integrated group with low proficiency level recorded the lowest median score (Md = 21.27).

Table 4-47 Ranks

	GroupandProficiency	N	Mean Rank
PostWordMeaningTotal	Integrated High	15	33.73
	Integrated Low	15	21.27
	Split Attention High	15	42.03
	Split Attention Low	15	24.97
	Total	60	

Table 4-48 Test Statistics[a,b]

	PostWordMeaningTotal
Chi-Square	12.937
Df	3
Asymp. Sig.	.005
a. Kruskal Wallis Test	
b. Grouping Variable: GroupandProficiency	

4.5.9.2 The effect of different proficiency levels on word family acquisition in both online reading split-attention and integrated instructional formats

Because the data collected from the word family post-test was not normally distributed, a non-parametric test (Kruskal-Wallis Test) was used to investigate the effect of different proficiency levels on word family acquisition in both online reading split-attention and integrated instructional formats.

A Kruskal-Wallis test revealed a statistically non-significant difference in the word family post-test total scores across the four groups which were divided according to treatment and different proficiency levels. χ^2 (2, n = 60) = 6.82, p = 0.078. The integrated group with high proficiency

level recorded the highest median score (*Md* = 37.80) amongst the four groups, while the split-attention group with low proficiency level recorded the lowest median score (*Md* = 21.57).

Table 4-49 Ranks

	GroupandProficiency	N	Mean Rank
PostWordFamiliesTotal	Integrated High	15	37.80
	Integrated Low	15	30.10
	Split Attention High	15	32.53
	Split Attention Low	15	21.57
	Total	60	

Table 4-50 Test Statistics[a,b]

	PostWordFamiliesTotal
Chi-Square	6.817
Df	3
Asymp. Sig.	.078
a. Kruskal Wallis Test	
b. Grouping Variable: GroupandProficiency	

4.5.9.3 The effect of different proficiency levels on collocations acquisition in both online reading split-attention and integrated instructional formats

Because the data collected from the collocations post-test was not normally distributed, a non-parametric test (Kruskal-Wallis Test) was used to investigate the effect of different proficiency levels on collocations acquisition in both online reading split-attention and integrated instructional formats.

A Kruskal-Wallis test revealed a statistically non-significant difference in the collocations post-test total scores across the four groups which were divided according to treatment and different proficiency levels. χ^2 (2, n = 60) = 7.23, p = 0.065. The treatment (integrated) group with high proficiency level recorded the highest median score (*Md* = 37.13) amongst the four groups, while the treatment (integrated) group with low proficiency level recorded the lowest median score (*Md* = 23.83).

Table 4-51 Ranks

	GroupandProficiency	N	Mean Rank
PostCollocationsTotal	Integrated High	15	37.13

	Integrated Low	15	23.83
	Split Attention High	15	35.83
	Split Attention Low	15	25.20
	Total	60	

Table 4-52 Test Statistics[a,b]

	PostCollocationsTotal
Chi-Square	7.233
Df	3
Asymp. Sig.	.065
a. Kruskal Wallis Test	
b. Grouping Variable: GroupandProficiency	

4.5.10 The effect of L1 on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats

A one-way between-groups analysis of variance was conducted to explore the impact of L1 on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats. Participants were divided into four groups according to their L1 and the treatment they were receiving. There was a statistically non-significant difference in general vocabulary knowledge post-total scores for the four groups: F (3, 56) = 1.08, p = .37. The results of this test show that L1 did not have any statistically significant effect on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats.

Table 4-53 ANOVA

Post-Total					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	203.879	3	67.960	1.075	.367
Within Groups	3538.855	56	63.194		
Total	3742.733	59			

Table 4-54 Descriptives

Post-Total								
					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Integrated Arabic	18	38.2222	6.85613	1.61601	34.8127	41.6317	31.00	56.00
Integrated Chinese	12	42.6667	6.66515	1.92406	38.4318	46.9015	29.00	53.00

Split Attention Arabic	17	38.0000	9.97497	2.41929	32.8713	43.1287	17.00	54.00
Split Attention Chinese	13	40.6154	7.41101	2.05544	36.1370	45.0938	24.00	48.00
Total	60	39.5667	7.96468	1.02824	37.5092	41.6242	17.00	56.00

4.5.10.1 The effect of L1 on word meaning acquisition in both online reading split-attention and integrated instructional formats

Because the data collected from the word meaning post-test was not normally distributed, a non- parametric test (Kruskal-Wallis Test) was used to investigate the effect of different L1s on word meaning acquisition in both online reading split-attention and integrated instructional formats.

A Kruskal-Wallis test revealed a statistically non-significant difference in the word meaning post-test total scores across the four groups which were divided according to treatment and different L1s (Integrated Arabic, n = 18, Integrated Chinese, n = 12, Split-attention Arabic, n = 17, Split-attention Chinese, n = 13). χ^2 (2, n = 60) = 6.7, p = 0.082. The split-attention group whose L1 is Chinese recorded the highest median score (Md = 37.73) amongst the four groups, while the integrated group whose L1 is Arabic recorded the lowest median score (Md = 22.64).

Table 4-55 Ranks

	GroupandLangauge	N	Mean Rank
PostWordMeaningTotal	Integrated Arabic	18	22.64
	Integrated Chinese	12	34.79
	Split Attention Arabic	17	30.26
	Split Attention Chinese	13	37.73
	Total	60	

Table 4-56 Test Statistics[a,b]

	PostWordMeaningTotal
Chi-Square	6.697
Df	3
Asymp. Sig.	.082
a. Kruskal Wallis Test	
b. Grouping Variable: GroupandLangauge	

4.5.10.2 The effect of L1 on word family acquisition in both online reading split-attention and integrated instructional formats

Because the data collected from the word family post-test was not normally distributed, a non-parametric test (Kruskal-Wallis Test) was used to investigate the effect of different L1s on word family acquisition in both online reading split-attention and integrated instructional formats.

A Kruskal-Wallis test revealed a statistically significant difference in the word family post-test total scores across the four groups which were divided according to treatment and different L1s (Integrated Arabic, n = 18, Integrated Chinese, n = 12, Split-attention Arabic, n = 17, Split-attention Chinese, n = 13). χ^2 (2, n = 60) = 9.09, p = 0.028. The integrated group whose L1 is Chinese recorded the highest median score (Md = 43.04) amongst the four groups, while the split-attention group whose L1 is Arabic recorded the lowest median score (Md = 23.94).

Table 4-57 Ranks

	GroupandLangauge	N	Mean Rank
PostWordFamiliesTotal	Integrated Arabic	18	27.89
	Integrated Chinese	12	43.04
	Split Attention Arabic	17	23.94
	Split Attention Chinese	13	31.12
	Total	60	

Table 4-58 Test Statistics[a,b]

	PostWordFamiliesTotal
Chi-Square	9.085
Df	3
Asymp. Sig.	.028
a. Kruskal Wallis Test	
b. Grouping Variable: GroupandLangauge	

4.5.10.3 The effect of L1 on collocations acquisition in both online reading split-attention and integrated instructional formats

Because the data collected from the collocations post-test was not normally distributed, a non-parametric test (Kruskal-Wallis Test) was used to investigate the effect of different L1s on collocations acquisition in both online reading split-attention and integrated instructional formats.

A Kruskal-Wallis test revealed a statistically non-significant difference in the collocations post-test total scores across the four groups which were divided according to treatment and different L1s (Integrated Arabic, n = 18, Integrated Chinese, n = 12, Split-attention Arabic, n = 17, Split-attention Chinese, n = 13). χ^2 (2, n = 60) = 0.03, p = 0.998. The treatment (integrated) group whose L1 is Chinese recorded the highest median score (Md = 31.13) amongst the four groups, while the treatment (integrated) group whose L1 is Arabic recorded the lowest median score (Md = 30.06).

Table 4-59 Ranks

	GroupandLangauge	N	Mean Rank
PostCollocationsTotal	Integrated Arabic	18	30.06
	Integrated Chinese	12	31.13
	Split Attention Arabic	17	30.71
	Split Attention Chinese	13	30.27
	Total	60	

Table 4-60 Test Statistics[a,b]

	PostCollocationsTotal
Chi-Square	.032
Df	3
Asymp. Sig.	.998
a. Kruskal Wallis Test	
b. Grouping Variable: GroupandLangauge	

4.6 Summary

The quantitative data collected in this study was analysed using descriptive and inferential statistics. Frequency, percentages and means were calculated. Crosstabulation, Chi Square, Paired-Samples T-test, Wilcoxon Signed Rank Test, Independent-Samples T-test, Mann-Whitney U Test, One-Way Between-Groups Analysis of Variance, and Kruskal-Wallis Test were conducted to test if there is any statistically significant relationship between various data categories. The findings and results of the study were stated and illustrated by tables and/or graphs.

The results showed that there was an increase in the number of participants who answered the vocabulary questions correctly after the experiment in both split-attention and integrated groups. This increase was mainly in word meaning and word family questions. A paired-samples t-test showed a statistically significant increase in the test scores from pre-test to post-test, and the results showed that the increase was statistically significant in word meaning and

collocation knowledge, but not in word family knowledge; however, there was no statistically significant difference between the results of the split-attention and the integrated group in the post-test although the scores of the integrated group were higher than those of the split-attention group. Moreover, a cross-tabulation showed that the integrated group gave more correct answers in the post-test but the chi squared test showed the difference was not statistically significant in either word meaning, word family or collocation knowledge tests, so it could be concluded that the different instructional formats did not significantly affect the vocabulary knowledge building in this study. While participants of higher proficiency level gave significantly more correct answers than those of lower proficiency level in the split-attention group, the difference in the integrated group was not statistically significant, L1 did not have any statistically significant effect on acquiring new vocabulary items in either split-attention or integrated groups.

These findings will be discussed in the next chapter in relation to other research findings.

Chapter 5. Discussion of the Results

5.1 Introduction

This research investigated the impact of different instructional formats; namely, integrated and split-attention online reading formats on building vocabulary knowledge of adult EAP learners studying at INTO, Newcastle. These learners were studying English as a second language, their first languages being Arabic and Chinese. As mentioned in the literature review, there is a lack of research which aims at investigating the relation between cognitive load and building vocabulary knowledge; thus, on the role of different instructional formats in vocabulary learning among EAP students in specific. In this chapter, the findings of this study will be discussed with reference to other important studies which have investigated this topic or related topics.

Cognitive Load Theory (CLT) and Cognitive Theory of Multimedia Learning (CTML) were adopted as the theoretical framework for this research.

In order to answer the research questions mentioned in Chapter Three, sixty students studying English at INTO, Newcastle participated in the experiment. Thirty students were in the split-attention instructional format group and thirty students were in the integrated instructional format group. The students spoke either Arabic or Chinese as first languages, and they were divided into lower proficiency and higher proficiency groups. All the students were pre-tested and post-tested for word meaning knowledge, word families knowledge and collocations knowledge. Data collected were analysed and the results were presented in the previous chapter.

In summary, the analysed data revealed that:

1. There was an increase in the number of participants who answered the vocabulary questions correctly after the experiment in split-attention and integrated groups, mainly in word meaning and word family questions.
2. There was also a statistically significant increase, mainly in word meaning and collocations knowledge, in the test scores from pre-test to post-test.
3. There was no statistically significant difference between the results of the split-attention and the integrated group in the post-test although the scores of the integrated group were higher than those of the split-attention group.

4. The different instructional formats did not significantly affect the vocabulary knowledge building in this study. While participants of higher proficiency level gave significantly more correct answers than those of lower proficiency level in the split-attention group, the difference in the integrated group was not statistically significant.
5. L1 did not have any statistically significant effect on acquiring new vocabulary items in either split-attention or integrated groups.

These findings are discussed in the following sections with reference to the findings of relevant research discussed in the literature review.

5.2 The Effect of the Online Reading Tasks on Vocabulary Knowledge

The first point investigated when the data was analysed was the effect of the online reading task on vocabulary knowledge among the learners. Descriptive as well as inferential findings were used for this investigation. As explained in 4.4.2. and 4.5.1., descriptive statistics were used to look at vocabulary knowledge before and after the learners were exposed to the reading texts and the accompanying vocabulary-testing questions. This comparison showed 6.6 per cent (from 86.7 percent to 93.3 percent) increase in the number of participants who answered correctly in both the split-attention and the integrated groups. After that, inferential statistics were used to investigate the effect of the treatment on vocabulary knowledge among the learners. The effect of the treatment on participants' scores of the vocabulary test was measured by the use of paired-samples t-test, and the results showed a statistically significant increase of 3.3 in the test scores from pre-test to post-test.

All of the above indicates that the online reading texts have helped increase the vocabulary knowledge among the learners in general in this experiment; however, according to Schmitt (2000), learning a word involves more than only learning what it means, how it is spelled, or how it is pronounced. As a matter of fact, additional criteria to word knowledge involve, in addition to other criteria, words' collocations, grammatical characteristics, and word family knowledge. Because learners should have the understanding of all the forms of word knowledge in order for them to use the words confidently, the vocabulary exercises in this study follow Schmitt and Schmitt (2011)'s model of concentrating on the three aspects of word meaning, word families (derivative forms of words), and collocations in order to achieve the maximum teaching goal of the exercises. The sixty questions in the pre- and post-tests were divided into: twenty questions to test word meaning knowledge, twenty questions to test word families knowledge, and twenty questions to test collocations knowledge. In addition to the comparison

between the output of the frequency of the recoded scores of vocabulary knowledge in general in the pre- and post- tests, a comparison was made between the output of the frequency of the recoded scores of word meaning, word families, and collocations in pre- and post-tests, and the effect of the online reading tasks on word meaning knowledge, word family knowledge, and collocations knowledge was investigated.

5.2.1 The effect of the online reading tasks on word meaning knowledge

The first aspect of vocabulary knowledge which was investigated was word meaning knowledge. Descriptive as well as inferential findings were used for this investigation.

First, descriptive statistics were used to look at word meaning knowledge before and after the learners were exposed to the reading texts and the accompanying vocabulary-testing questions. This comparison showed 10 per cent (from 86.7 percent to 96.7 percent) increase in the number of participants who answered correctly in both the split-attention and the integrated groups.

After that, inferential statistics were used to investigate the effect of the treatment on word meaning knowledge among the learners. The effect of the treatment on participants' scores of the word meaning knowledge element in the vocabulary test was measured by the use of a non-parametric test (Wilcoxon Signed Rank Test), and the results showed a statistically significant increase in the word meaning test scores from pre-test to post-test.

All of the above indicates that although on one hand the maximum total score of correct answers is lower in the post-test than it is in the pre-test, on the other hand there is an increase in the numbers of participants who answered correctly. This increase in the number of participants who answered correctly, the higher mean and lower standard deviation in the post-test results, and the statistically significant increase in the word meaning test scores from pre- to post-test can lead to the belief that the online reading texts have the potential of helping increase the word meaning knowledge among the learners in this experiment. It could be argued that a larger number of participants might lead to better stronger conclusion in this case.

An important point to be taken into consideration when looking at the effect of the online reading texts on the word meaning knowledge is thinking of the connection between the meaning conveyed by the form of the word in the target language (English in this study) and the first language (Arabic and Chinese in this study) and considering whether the concept is well known in the first language of the participants; namely, Arabic or Chinese in this study.

5.2.2 The effect of the online reading tasks on word families knowledge

The second aspect of vocabulary knowledge which was investigated was word family knowledge. Descriptive as well as inferential findings were used for this investigation.

First, descriptive statistics were used to look at word family knowledge before and after the learners were exposed to the reading texts and the accompanying vocabulary-testing questions. This was achieved by comparing the output of the frequency of the recoded scores of word family knowledge in pre- and post- tests. Whereas the maximum total score of correct answers in the pre-test was 17 out of 20, the maximum total score of correct answers in the post-test was 18 out of 20. However, the minimum total score of correct answers in both tests was 5 out of 20. The mean of correct answers in the post-test was slightly higher than the mean of correct answers in the pre-test, and the standard deviation was slightly less in the post-test.

After that, inferential statistics were used to investigate the effect of the treatment on word family knowledge among the learners. The effect of the treatment on participants' scores of the word family knowledge element in the vocabulary test was measured by the use of a non-parametric test (Wilcoxon Signed Rank Test). Although the results showed an increase in the word family test scores from pre-test to post-test, this increase was not statistically significant.

All of the above descriptive results indicate that the online reading texts have helped increase the word family knowledge among the learners in this experiment; however, the inferential results are statistically insignificant. The important point here is that the inferential results have not found the opposite. They still support the finding that the reading texts have enhanced the word family knowledge.

5.2.3 The effect of the online reading tasks on collocations knowledge

The third aspect of vocabulary knowledge which was investigated was collocations knowledge. Descriptive as well as inferential findings were used for this investigation.

First, descriptive statistics were used to look at collocations knowledge before and after the learners were exposed to the reading texts and the accompanying vocabulary-testing questions. This comparison showed no difference between pre- and post-test results in the number of participants who answered correctly in both the split-attention and the integrated groups. Whereas the maximum total score of correct answers in both collocations pre-and post-tests was 20 out of 20, the minimum total score of correct answers in the collocations pre-test was 4 out of 20, and the minimum total score in the collocations post-test was 6 out of 20. The mean

of correct answers in the post-test was higher than the mean of correct answers in the pre-test, and the standard deviation was less in the post-test.

After that, inferential statistics were used to investigate the effect of the treatment on collocations knowledge among the learners. The effect of the treatment on participants' scores of the collocations knowledge element in the vocabulary test was measured by the use of a non-parametric test (Wilcoxon Signed Rank Test), and the results showed a statistically significant increase in the collocations test scores from pre-test to post-test.

All of the above indicates that although hardly any difference could be noticed between the pre- and post-tests aiming at testing collocations knowledge, the increase in total score of correct answers in the post-test, the statistically significant increase of the collocations test scores from pre- to post-tests, the increase in the mean of correct answers in the post-tests, in addition to the decrease in the standard deviation indicates that the reading text has enhanced the collocations knowledge among the participants in this study. It could be argued that a larger number of participants can lead to as stronger conclusion in this case.

5.2.4 Summary findings and discussion of the effect of the treatment on vocabulary knowledge building

The results reported so far support Horst et al. (1998)'s assumption that reading is essential for second language learning exactly as it is essential for first language development. The studies which claimed reading did not play an important role in vocabulary acquisition were believed to have methodological issues (Horst et al., 1998). For example, two out of thirty items was the mean score in Pitts et al. (1989)'s study where learners read *A Clockwork Orange* for an hour. Similarly, Day et al. (1991) and Hulstijn (1992) in their research which also uses the Clockwork Orange methodology, report learners identifying a comparatively small number of words. Dupuy and Krashen (1993) reported a higher score of seven words; however, reading was not the only variable since their research involved watching a video. As a whole, the findings of these studies of L2 report an average of one word identified correctly in every twelve words. Horst et al. (1998) believe that these findings are not startling because of the limitations in the experiments where the chances of reading and acquiring new vocabulary are very limited because the reading treatments were an hour long or even less than that. This study is believed to have led to better results because learners were given the reading texts over three sessions and were tested on vocabulary knowledge within each session. The questions in pre- and post-tests were taken from the within-experiment questions.

Although Saragi et al. (1978) gave out a 60000-word novel to be read over a few days, Pitts et al. (1989)'s study required the participants to read 6700 words, and Hulstijn (1992)'s study required the reading of 907 words, there was no way to double check whether the participants indeed read the whole assigned texts; thus, the actual amount of words read might be even less than stated in the research methodology of these studies. Pitts et al. (1989) reported that more than 50% of their participants did not finish the reading task; moreover, there is no evidence that the other studies have not faced the same issues of limitations in the chances of acquiring new words from reading texts, lack of enough words in the tests, and lack of experimental control. In order to avoid issues like these, the researcher in this study asked for permission for the tests and experiment to be taken at INTO and not at home so that the researcher could be present during the pre-test, experiment, and post-test.

As a result, the scores of this research support Schmitt (2010)'s statement that when studies deal with some or all of the above-mentioned issues, the results show more gains from reading than the gains showed by the results of earlier research.

A lot of studies focussed on online reading and on the methods which enhance text (Akyel and Erçetin, 2009, Huang et al., 2009, Anderson, 2003). A lot of research was also done on different formats for reading texts and different methods to deal with information which are provided by electronic documents but may be hard for readers who are taught to infer meaning from traditional paper-based reading texts (Coiro, 2003, Protopsaltis, 2008).

Although this study does not compare online reading with reading on paper, the descriptive findings of the gain in vocabulary knowledge after the online treatment can be added to the findings of researchers who believe that online reading in the integrated format is beneficial in second language learning. For example, in her empirical study, Chun (2001) tried to find out the ways in which L2 learners access information during reading L2 texts in an online learning environment. After comparing the learning of words which were hyperlinked online and learning the words which were not, she concluded that learners who were reading online performed better as long as they were hyperlinked with a website. Moreover, Tozcu and Coady (2004), tried to explore the influence of direct vocabulary learning using Computer-assisted Language Learning (CALL) on vocabulary knowledge, reading comprehension, and speed of word recognition. They concluded that a significantly larger amount of vocabulary was learned by students who used Tutorial CALL to learn highly frequent vocabulary than the amount learned by those who were not exposed to the CALL program, the reaction time for frequent word recognition was reduced amongst learners who were exposed to the CALL program, and reading comprehension was shown to be significantly better among those learners than those

who were not exposed to the Tutorial CALL; however, repetition and high-frequency of vocabulary use in multimedia-based online reading materials can have the potential of increasing cognitive load, especially for non-native speakers who are trying to translate while they are processing information (Tozcu and Coady, 2004). Online English reading materials, like any other online material, may include distractors such as pop-up windows which provide explanations, brightly coloured words, hyperlinks to different websites, explanatory pictures or videos, and many more. As opposed to studies like Chun (2001), some research has shown that these distractors can increase the amount of cognitive load, which is the amount of resources stored in the cognitive system which has limited capacity to hold information; namely, the working memory. While this increase in cognitive load can prevent effective reading (Akbulut, 2007), web pages with simple instructional design inspired by research results were found to improve ESL reading skills (Akbulut, 2007, Chun, 2001, Gerjets et al., 2004). The treatment in this study was chosen to be as simple as possible with no distractors to the text, so no distracting images, extremely bright colours, or hyperlinks were used in order for the researcher to be able to filter out the distractors and minimise the cognitive load.

Moreover, the reading text in the present study had the words repeated a few times within the text since some research has shown that repetition and frequency of occurrence may enhance the vocabulary learning process. For example, Nation (2022) states that frequency and range of occurrence are very beneficial to be used in order to differentiate various levels of vocabulary. In an earlier study, Nation (1990) argues that it is best to teach words which are difficult to define inductively by providing many examples of sentences including the words. This gives the students the opportunity to repeat the words and allows them to work harder towards learning the words. It can also be argued that dividing the reading texts into shorter paragraphs may have led to interrupting the thoughts of the readers, and although it was supposed to reduce the cognitive load, it might have increased it from another angle; more specifically, it might have added more pressure on the working memory to remember the content when moving from one paragraph to the other.

5.3 Is an Online Reading Task in the Integrated Format More Effective than in the Split-attention Format in Helping Adult EAP Learners Build Their Vocabulary Knowledge?

The second point investigated when the data was analysed was the effect of the online reading split-attention and integrated instructional formats on vocabulary knowledge among the learners. Descriptive as well as inferential findings were used for this investigation.

First, descriptive statistics were used to investigate whether there was an association between the different reading instructional formats and second language vocabulary learning or not. Although the descriptive results showed that participants who answered the post-test correctly were 13.3% more in the integrated format group than the split-attention format group, the results were statistically insignificant. This finding was confirmed in the inferential analysis where an independent-samples t-test was conducted to compare between the scores of the vocabulary knowledge 60-question post-test in the split-attention format group and the integrated format group. No significant difference in scores was statistically found although the integrated group had higher scores than the split-attention group.

When this study was designed, most of the following studies had been reviewed, and a few were published after this research was done. The studies were all reviewed in detail in Chapter two; however, a brief explanation of them will follow where they will be put in chronological order before making a conclusion to this section.

Schnotz and Kürschner (2007) stated that the split-attention effect occurred when the readers' visual focus was on various types of information, and when readers needed to split their attention between these different information types and receive extra cognitive load in order to gain linguistic knowledge. Moreover, Akbulut (2007) found that significantly higher vocabulary scores were achieved by learners who had access to definitions accompanied by visuals than by learners who only had annotations on their own; thus, a correlation between the effectiveness of instructional formats online and vocabulary learning, particularly for advanced L2 learners was found. Mikk (2008), on the other hand, found that a high correlation was found between long sentences and overloading working memory. Hung (2009) found that an integrated format (where questions were actually integrated into related paragraphs) successfully reduced the cognitive load, whereas a split-attention format (comprehension questions following reading passages) increased the cognitive load. Following the same model, Al-Shehri and Gitsaki (2010), who aimed at exploring the effect of split-attention and integrated

instructional formats on students' cognitive load and on the way these instructional formats could affect online reading and vocabulary learning in second language, found that the learners' reading comprehension was made easier when the integrated format was used than when the split-attention one was used. In addition to that, learners in the Split Attention with Online Dictionary and Integrated Format with Online Dictionary groups had better results in the vocabulary test; however, the reading task took them more time than it took the other two groups. Moreover, more vocabulary was looked up by learners in the split-attention groups than learners in the integrated format groups. Genç and Gülözer (2013) also followed the model of positioning comprehension questions after the reading text in split-attention format, and dividing the reading text into paragraphs and inserting comprehension questions between the suitable paragraphs in order to study the effect of different cognitive load caused by different instructional formats (split-attention and integrated formats) and text presentation methods (paper-based and online) on learners' second language reading comprehension performance. The results of a post-test revealed no statistically significant difference among the four groups with regards to the scores of learners' second language reading comprehension tests. However, a statistically significant difference in L2 comprehension between participants who read online text and those who read paper-based text was suggested by the results. Similarly, Mutlu et al. (2022) designed a study with the goal of evaluating the effect of the split-attention format in multi-media learning environments to compare subjective and objective cognitive load measurements. The results showed significant differences between the groups due to the split-attention effect. More cognitive load was involved in the situations where the text was not integrated into the image. To minimise or even prevent the split-attention effect, Mutlu et al. (2022) suggest integrating the text into the image when designing online learning materials.

It can be concluded that in order to study the effect of different instructional formats on the vocabulary knowledge building, the researcher in this study followed the model of exposing the split-attention group to the split-attention format represented by a reading text followed by vocabulary testing questions, and exposing the other group to the integrated format represented by dividing the reading texts into shorter paragraphs and inserting the vocabulary testing questions among the paragraphs aiming at reducing the cognitive load; however, the results of the present study do not support any of the results of the other studies which followed the same model. A potential reason is that the current study did not take into consideration the intrinsic cognitive load, which is the fundamental difficulty of a certain topic, no matter of the method of presenting this topic. Another difference is that most of the studies reviewed above focused on reading comprehension questions instead of vocabulary testing questions, and when they

focused on vocabulary, they followed different methodology. For example, some of the studies did not have pre-tests. One of the very important factors also is the number of the participants. The statistically non-significant results can be caused by the small number of questions testing each of the word meaning knowledge, the word family knowledge, and the collocations knowledge.

5.3.1 Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word meaning knowledge?

Descriptive statistics were used to investigate whether there was an association between the different reading instructional formats and second language word meaning knowledge or not. Although the descriptive results showed that participants who answered the post-test correctly were 6.7 % more in the integrated format group than the split-attention format group, the results were statistically insignificant. This finding was confirmed in the inferential analysis where a Mann-Whitney U Test was conducted to compare between the scores of the word meaning knowledge 20-question post-test in the split-attention format group and the integrated format group. It cannot be claimed that the two instructional formats had any effect on the word meaning knowledge in the experiment because the results found were not statistically significant.

5.3.2 Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word families knowledge?

Descriptive statistics were used to investigate whether there was an association between the different reading instructional formats and second language word family knowledge or not. Although the descriptive results showed that participants who answered the post-test correctly were 16.7 % more in the integrated format group than the split-attention format group, the results were statistically insignificant. This finding was confirmed in the inferential analysis where a Mann-Whitney U Test was conducted to compare between the scores of the word family knowledge 20-question post-test in the split-attention format group and the integrated format group. It cannot be claimed that the two instructional formats had any effect on the word family knowledge in the experiment because the results found were not statistically significant.

5.3.3 Is an online reading task in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word collocations knowledge?

Descriptive statistics were used to investigate whether there was an association between the different reading instructional formats and second language collocations knowledge or not. Although the descriptive results showed that participants who answered the post-test correctly were 6.7 % more in the integrated format group than the split-attention format group, the results were statistically insignificant. This finding was confirmed in the inferential analysis where a Mann-Whitney U Test was conducted to compare between the scores of the collocations knowledge 20-question post-test in the split-attention format group and the integrated format group. It cannot be claimed that the two instructional formats had any effect on the collocations knowledge in the experiment because the results found were not statistically significant.

5.3.4 Summary findings and discussion of the first research question

Descriptive and inferential statistics which were used to investigate the association between, and the influence of different reading instructional formats on the three aspects of vocabulary knowledge tested in this study; namely, word meaning knowledge, word families knowledge, and collocations knowledge, resulted in statistically insignificant findings. Although gains in the three aspects of vocabulary knowledge were found by comparing the score results, the statistically insignificant findings indicate that the different instructional formats of integrated and split-attention had no effect on the vocabulary gains although according to the Cognitive Load Theory, which has been used by second language acquisition (SLA) researchers to explain any differences in learners' performance when they are exposed to different instructional formats, various learning tasks were found to affect cognitive load and working memory differently, and this effect was shown to be emphasised in a multimedia environment where various interacting elements and tools can result in cognitive overload and decrease the learning results. Many researchers investigated the role that Cognitive Load Theory played in learning, and found that it is highly related to the mental effort needed to complete a task successfully, and as a result the ability of the working memory (Pillay, 1994), and (Cierniak et al., 2009).

Paas et al. (2003), Leahy and Sweller (2004), Schnotz and Kürschner (2007), and Cierniak et al. (2009); thus, Yeung et al. (1998), Al-Shehri and Gitsaki (2010), Hung (2009), Zumbach and Mohraz (2008), DeStefano and LeFevre (2007), and many others have tried to design instructional formats which can lead to less cognitive load. Whereas an integrated format can

allow for a smooth transfer of a new concept from working memory into a schema because it is immediate, a split-attention format does not allow this immediacy because of the delay. This can be reinforced by repetition.

The results of the present study do not support any of the above findings on the effect on cognitive load that different instructional formats may cause; however, Cognitive Load Theory divides cognitive load into three forms: intrinsic, extraneous, and germane. Whereas extraneous load is defined by Cierniak et al. (2009) as unneeded processing of information which is the result of instructional designs, Brunken et al. (2003) define intrinsic cognitive load as the cognitive load which is generated by the innate difficulty, instead of the instructional design of the subject. Because intrinsic and extraneous loads are extra burden on working memory (Paas et al., 2003), researchers recommend that the instructional format in any learning material should be designed cautiously, mainly when the innate level of difficulty in the content, or the intrinsic load, is high. Since the integrated format group in this study did not show statistically significant gains over the split-attention format group, one of the interpretations could be that the factor of intrinsic cognitive load was not taken into consideration when the experiment was designed. This point of learners' own innate ability is referred to by Yeung et al. (1998) who found the integrated format to be useful in decreasing the cognitive load and improving reading comprehension for both first language (L1) and second language (L2) learners. In their study, Yeung et al. (1998) administered five experiments to investigate the influence of cognitive load management using explanatory notes in reading paragraphs for readers with various linguistic levels. The third experiment showed that an integrated format reduced comprehension amongst adult readers but improved vocabulary learning. This result is different from the results of the present research; however, Yeung et al. (1998) did not find the same result for every group of learners they tested. Their argument was that the degree to which instructional formats impose an extraneous cognitive load affects its efficiency; thus, the learners' abilities play a very important role on whether the same format; namely split-attention or integrated, may enhance performance or decrease it among learners.

One of the factors which may have led to different results from the above-mentioned experiments is that the type of cognitive load which was mostly considered when designing the present experiment was extrinsic, which focused on the effect of the split-attention format on the working memory.

Another factor which may have affected the results of the experiment is the individual differences among the learners. One possibility is the difference in the learners' academic

abilities in general. In her empirical study which is reviewed in this research in 2.7.1., Chun (2001) tried to find out the ways in which L2 learners access information during reading L2 texts in an online learning environment. The results of Chun's study gave insights into the way L2 learners use multimedia to help with their comprehension during reading, but one of the aspects the study focused on was the different abilities of the learners where higher ability students spent less time on reading and looked up less words than lower ability students. Similarly, in their research, Ozono and Ito (2003) concluded that students' various abilities affected their reading comprehension and their cognitive load. Another example of individual differences among learners is their memory capacity. Chun and Payne (2004) investigated the effect of learners' working memory capacity on their L2 vocabulary acquisition and their reading comprehension. Data analysis showed more words were looked up by participants with lower verbal working memory capacity. According to Hyltenstam and Abrahamsson (2004), and to DeKeyser (2000), in order to achieve high levels of L2 proficiency which are native-like, L2 learners should have high levels of language aptitude, which is a talent for language learning and acquisition that includes various constituents and that can be measured by an aptitude battery. The nature of language aptitude can be affected by working memory (Miyake and Friedman, 1998, Ellis, 2005) and it is expected that part of the difference in the ultimate level of language attainment can be accounted for by the individual differences in working memory. Despite the little amount of research which has been conducted to study the relationship between working memory and processing of L2, most of the studies so far support the above-mentioned suggestion (Miyake and Friedman, 1998, Harrington and Sawyer, 1992, Harrington, 1991, Geva and Ryan, 1993). In their research, Harrington and Sawyer (1992) and Miyake and Friedman (1998) found a statistically significant correlation between L2 working memory capacity as measured by reading span tasks and L2 linguistic features comprehension. These differences in working memory among the learners were not measured in this study, but this could make a recommendation for further studies.

5.4 Do Adult EAP Learners at Different Proficiency Levels Acquire New Vocabulary Items from the Reading Text in Both Split-attention and Integrated Formats Differently?

The second point investigated when the data was analysed was the effect of different proficiency levels on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats. Inferential findings were used for this investigation.

A one-way between-groups analysis of variance was conducted to compare the scores of the vocabulary knowledge 60-question post-test among the four groups that the participants were divided into according to their proficiency levels and the treatment they were receiving. The difference in general vocabulary knowledge post-test scores among the four groups was statistically significant. There was a significantly different mean score of the split-attention low proficiency group from both the integrated high proficiency group and the split-attention high proficiency group where the mean score of the split-attention low proficiency group was 7.46 lower than the mean score of the integrated high proficiency group, and 7.33 lower than the mean score of the split-attention high proficiency group. These results indicate that the proficiency level affected the general vocabulary knowledge test scores of the split-attention group significantly but did not have significant effect on the general vocabulary knowledge post-test scores of the integrated group.

The above finding shows that the split-attention format low proficiency group had significantly lower scores in the post-test than both groups with high-proficiency level, whereas the integrated format low proficiency group was not affected. This could indicate that having less cognitive load on the working memory helped the low proficiency group. More details will follow in the conclusion of this section.

5.4.1 Do adult EAP learners at different proficiency levels acquire word meaning knowledge from the reading text in both split-attention and integrated formats differently?

In order to investigate the effect of different proficiency levels on word meaning acquisition in both split-attention and integrated instructional formats, a non-parametric test (Kruskal-Wallis Test) was conducted. The test compared the scores of the word meaning post-test scores among the four groups that the participants were divided into according to their proficiency levels and the treatment they were receiving. Whereas the highest median score was that of the split-attention group with high proficiency level, the lowest median score was that of the integrated low proficiency group. The result found was statistically significant.

This result which focuses on the word meaning aspect of the vocabulary knowledge contradicts the result above which shows, in contrast to this one, that the integrated format low proficiency level group was the least affected by the proficiency level of the learners when they were tested for vocabulary knowledge in general.

5.4.2 Do adult EAP learners at different proficiency levels acquire word family knowledge from the reading text in both split-attention and integrated formats differently?

In order to investigate the effect of different proficiency levels on word family acquisition in both split-attention and integrated instructional formats, a non-parametric test (Kruskal-Wallis Test) was conducted. The test compared the scores of the word family post-test scores among the four groups that the participants were divided into according to their proficiency levels and the treatment they were receiving. Whereas the highest median score was that of the integrated group with high proficiency level, the lowest median score was that of the split-attention low proficiency group; however, this result was not statistically significant.

Considering the Cognitive Load Theory, this result makes sense since the integrated format high proficiency group is expected to have been exposed to the least amount of cognitive load, whereas the split-attention low proficiency group is expected to have been exposed to the highest amount of extraneous cognitive load; hence, they are expected to have the lowest scores whereas the integrated high proficiency learners are expected to have the highest scores.

5.4.3 Do adult EAP learners at different proficiency levels acquire collocations knowledge from the reading text in both split-attention and integrated formats differently?

In order to investigate the effect of different proficiency levels on collocations acquisition in both split-attention and integrated instructional formats, a non-parametric test (Kruskal=Wallis Test) was conducted. The test compared the scores of the collocations post-test scores among the four groups that the participants were divided into according to their proficiency levels and the treatment they were receiving. Whereas the highest median score was that of the integrated group with high proficiency level, the lowest median score was that of the integrated low proficiency group; however, this result was not statistically significant.

This is another result which is not consistent with previous findings, and it can be due to the small numbers in the groups, or to the specific nature of the aspect of vocabulary knowledge which is being studied, which is collocations knowledge in this question. Collocations sometimes are learned as lexical chunks, and they can be a tricky aspect of learning vocabulary among EAP learners because not many learners are aware of their existence or of what they are.

5.4.4 Summary findings and discussion of the second research question

The inferential statistics used to investigate the effect of proficiency level on general vocabulary knowledge among the four groups showed that the split-attention format low proficiency group had significantly lower scores in the post-test than the both groups with high-proficiency level, whereas the integrated format low proficiency group did not have statistically significant differences in scores compared to the integrated format high proficiency group.

When only the word meaning element of the vocabulary knowledge was tested, the results showed that the split-attention group with high proficiency level had the highest scores whereas the integrated low proficiency group had the lowest scores.

When only the word family element of the vocabulary knowledge was tested, the results showed that whereas the integrated group with high proficiency level had the highest scores, the split-attention low proficiency group had the lowest scores.

When only the collocations knowledge element of the vocabulary knowledge was tested, the results showed that whereas the highest median score was that of the integrated group with high proficiency level, the lowest median score was that of the integrated low proficiency group; however, this result was not statistically significant.

It can be said that although higher proficiency groups had higher scores than lower proficiency groups, the inconsistency of which instructional format resulted in more gains, and the statistically insignificant results, especially within the groups of low numbers, could be interpreted in many different ways.

As mentioned in Ch.2, germane cognitive load helps learners be more aware during the learning process and is connected with effective schema construction and automatization. Automation of schemas is believed to reduce cognitive load. Burkes (2007) affirms that skills such as reading, which are habitually practised, can occur spontaneously without a lot of conscious effort by learners despite of the possible complexity of the accompanying tasks. Cierniak et al. (2009) advises that "… an instructional design should reduce extraneous load (i.e., information processing hindering learning) and increase germane load (i.e., information processing supporting learning)". Despite the unalterable nature of the intrinsic load due to it being an essential part of the subject matter, extraneous and germane cognitive load may be highly affected by instructional formats of learning materials. Consequently, various amounts of memory load may be generated by the very same learning material depending on the different

methods manipulated for presenting the material (Brunken et al., 2003). This may explain the inconsistency in the results. In other words, what was sometimes increased by extraneous cognitive load, could be decreased by germane cognitive load, and vice versa.

Building on the information above, a very important learning process to think about is automation. Operation of information can occur either consciously in limited working memory or automatically with no influence from the working memory. Automation happens by practice, and it minimises the cognitive load on working memory. According to Leahy and Sweller (2004) and Sweller et al. (1998), automation can be noticed , for example, among advanced readers who do not need to consciously pay attention to the letters while reading because the letters are automated, whereas less advanced or beginner readers need to notice each letter consciously in order to read the text. Input should be processed in working memory so that information can be built in the form of schemas in long-term memory, and this whole mass of mental activity forced on working memory is called cognitive load. It can be said that automation explains the higher scores among the groups of high proficiency level learners. This concept of automization of explicit knowledge is stressed by Ellis (2005) whose suggestions are consistent to a large degree with the idea of moving from declarative to procedural knowledge. It can be argued that working memory plays a more powerful role in the early stages of the language acquisition process because learners may rely more on explicit, declarative knowledge which is controlled by short-term memory, and they cannot yet use grammatical knowledge. Their working memory is likely to become constantly overloaded, resulting in slow and difficult analysis of information. When knowledge becomes proceduralized, information processing can become more automatic and possibly more independent from working memory, because controlled processing is more challenging for working memory than automatic processing. What has been said does not necessarily mean that working memory is not expected to play any role in the advanced stages of the learning process, since it has always been found that working memory can have an efficient role in L1 acquisition, which is potentially automatized and procedural, thus it can be suggested that it can be very effective at the early stages of the learning process to have a large working memory capacity, whereas later on during the learning process, working memory may start to be helpful the same way it is during first language acquisition. To conclude, although inconsistent, and sometimes statistically insignificant, the results of this section all have one thing in common, which is that higher proficiency level groups, no matter what instructional format they were exposed to, showed more gains in their post-tests.

5.5 Do Adult EAP Learners Whose First Language is Arabic Acquire New Vocabulary Items from the Reading Text in both Split-attention and Integrated Formats Differently from Adult EAP Learners Whose First Language is Chinese?

The fourth point investigated when the data was analysed was the effect of first language (L1) on acquiring new vocabulary items in both online reading split-attention and integrated instructional formats. Inferential findings were used for this investigation.

A one-way between-groups analysis of variance was conducted to compare the scores of the vocabulary knowledge 60-question post-test among the four groups that the participants were divided into according to their L1 and the treatment they were receiving. The difference in general vocabulary knowledge post-test scores among the four groups was statistically non-significant. The results of this test showed that L1 did not have any statistically significant effect on acquiring new vocabulary items in either online reading split-attention or integrated instructional formats.

Observers nowadays argue that specific properties of second language learners' first language (L1) are predicted to be noticed in their second language (L2) performance. For example, Arabic-speaking learners of English conventionally produce target-like and non-target-like English which is different from the English produced by Mandarin-speaking learners of English. In spite of how controversial this argument has been over decades, most contemporary researchers hypothesize that a learner's L1 definitely affects the acquisition of L2. The debate has always been about the nature of this effect. In order to find more about the role of L1 in different aspects of L2 vocabulary acquisition, further investigation about the effect of L1 on word meaning, word families, and collocations knowledge was done.

5.5.1 Do adult EAP learners whose first language is Arabic acquire word meaning knowledge from the reading text in both split-attention and integrated formats differently from adult EAP learners whose first language is Chinese?

In order to investigate the effect of first language (L1) on word meaning acquisition in both split-attention and integrated instructional formats, a non-parametric test (Kruskal-Wallis Test) was conducted. The test compared the scores of the word meaning post-test scores among the four groups that the participants were divided into according to their L1 and the treatment they

were receiving. Whereas the highest median score was that of the split-attention group whose L1 is Chinese, the lowest median score was that of the integrated group whose L1 was Arabic; however, the result found was statistically non-significant.

5.5.2 Do adult EAP learners whose first language is Arabic acquire word family knowledge from the reading text in both split-attention and integrated formats differently from adult EAP learners whose first language is Chinese?

In order to investigate the effect of different first language (L1) on word family acquisition in both split-attention and integrated instructional formats, a non-parametric test (Kruskal-Wallis Test) was conducted. The test compared the scores of the word family post-test scores among the four groups that the participants were divided into according to their L1 and the treatment they were receiving. Whereas the highest median score was that of the integrated format group whose L1 is Chinese, the lowest median score was that of the split-attention group whose L1 is Arabic. The result found was statistically significant.

5.5.3 Do adult EAP learners whose first language is Arabic acquire collocations knowledge from the reading text in both split-attention and integrated formats differently from adult EAP learners whose first language is Chinese?

In order to investigate the effect of different first language (L1) on collocations acquisition in both split-attention and integrated instructional formats, a non-parametric test (Kruskal-Wallis Test) was conducted. The test compared the scores of the collocations post-test scores among the four groups that the participants were divided into according to their L1 and the treatment they were receiving. Whereas the highest median score was that of the integrated group whose L1 is Chinese, the lowest median score was that of the integrated group whose L1 is Arabic; however, the result found was statistically non-significant.

5.5.4 Summary findings and discussion of the third research question

The results of the test which investigated the effect of L1 on vocabulary knowledge showed that L1 did not have any statistically significant effect on acquiring new vocabulary items in either online reading split-attention or integrated instructional formats.

When only the word meaning element of the vocabulary knowledge was investigated, the results showed that whereas the highest median score was that of the split-attention group whose L1 is Chinese, the lowest median score was that of the integrated group whose L1 was Arabic; however, the result was statistically non-significant.

The only statistically significant result was for the word family knowledge element of the vocabulary knowledge. Whereas the highest median score was that of the integrated format group whose L1 is Chinese, the lowest median score was that of the split-attention group whose L1 is Arabic. This result can be interpreted to be caused by the grammar elements which exists in the word family knowledge.

When only the collocations knowledge element of the vocabulary knowledge was investigated, the results showed that whereas the highest median score was that of the integrated group whose L1 is Chinese, the lowest median score was that of the integrated group whose L1 is Arabic; however, the result found was statistically non-significant.

Although the common factor among all the above-mentioned results is that the groups of L1 Chinese had higher scores than the groups of L1 Arabic, no matter what the instructional formats these groups were exposed to were, the results were inconsistent, and except for those investigating the word family element of the vocabulary knowledge, they were statistically non-significant.

One of the interpretations which can be possible here is Nation (2022)'s argument that knowing the form and meaning of a word separately is not enough because it is essential to be able to connect both the form and the meaning of the word. Nation (2022) gives the word *brunch* as an example where the learner may be aware of the form itself or aware of the concept of a meal which replaces both breakfast and lunch; however, learning the word involves relating the form *brunch* to the concept of the meal which replaces breakfast and lunch. Learners are more ready to learn the word when they can better relate its form to its meaning, especially if they are already familiar with the concept conveyed by the word. It can sometimes be valid to separate the word meaning from its form and measure the learners' knowledge of both aspects separately; however, vocabulary acquisition involves acquiring both word form and word meaning. This point was not taken into consideration in this research, and the researcher believes it is worth considering it when looking at the word meaning acquisition after being exposed to reading texts, especially when the learners have different first languages and cultural backgrounds. It can be argued that cultural backgrounds play a very important role in acquiring

second language word form and word meaning equally. In terms of form, cultural effect plays a role in learning word pronunciation, spelling, and grammatical rules; thus, words are formed and written differently because different languages have different phonetic structures or writing systems. Cultural traditions also involve some linguistic features like tone, register, and style, which affects word form within context. In terms of meaning, different cultural backgrounds may lead to semantic differences and various connotations. For example, the word "freedom" can evoke different political connotations in certain cultures. In addition, words can be strongly related to cultural associations. The Japanese word "sakura", for example, which means "cherry blossom", is strongly associated with celebrations and aesthetics.

5.6 Summary

In summary, the findings of this study are not fully in agreement with the majority of studies mentioned in the literature review in Chapter Two. The results are not fully against the results of other research; however, most of the results that are consistent with previous research are not statistically significant. This can be due to methodological limitations in the study, and to elements which have not been considered when designing the experiment, such as measuring the working memory span among participants or measuring collocations acquisition differently. It is worth considering that one of the common collocation acquisition problems for EAP learners is randomness. Collocations may be learned as lexical chunks which do not follow any rules and do not have reasons for existence. For example, if a learner is asked why they "do the homework' rather than "make the homework", the answer is that they do not know.

Another methodological limitation can be separating interdependent elements of lexical knowledge for measurement purposes. Although it is valid to separate elements of lexical knowledge in order to measure them, it should be considered that the different aspects of vocabulary knowledge such as vocabulary size, word meaning, word forms, and collocations are intricately connected in the realistic use of the language.

A mixed methods approach could have been more beneficial in this research. Using qualitative in addition to the quantitative methods could have given the researcher a comprehensive understanding of the research topic and a deeper insight into the learners' experiences and perspectives. Adding a qualitative approach to data collection and analysis could have helped the researcher generate hypotheses and explore underlying reasons for the results obtained.

These limitations will be explained in the next chapter.

Chapter 6. Conclusion

6.1 Introduction

This chapter presents, firstly, a summary of the main results of the research, followed by a consideration of pedagogical implications for teachers and institutions, as well as recommending implications for further research. The limitations of the study are assessed after that. The chapter concludes with a brief summary of the previous sections.

6.2 Research Procedure and Findings in Brief

The aim of the study was to investigate the effect of two online reading instructional formats; namely, split-attention and integrated instructional formats on vocabulary knowledge building among sixty participants (35 Arabic speakers, 25 Chinese speakers) who were a group of adult learners studying English for University Studies at INTO, Newcastle upon Tyne. More specifically, the study asked: are reading tasks in the integrated format more effective than in the split-attention format in helping adult EAP learners build their word meaning, word families and collocations knowledge? Do adult EAP learners at different proficiency levels, or with different first languages acquire new vocabulary items differently with the different formats?

Comparing the pre-test results and the post-test results of these learners showed a statistically significant increase of 3.3 in the test scores from pre-test to post-test in general vocabulary knowledge. It also showed statistically significant increase in their word meaning and their collocations knowledge scores, but a statistically non-significant increase in their word family knowledge scores. It is believed that online reading combined with tasks in certain instructional formats is supposed to improve the vocabulary knowledge among learners.

To answer the research questions, a quantitative approach was adopted. Collecting the quantitative data for the sake of this study was done by the use of a quasi-experimental intervention study.

For the sake of the experiment, the participants were assigned into low proficiency level learners and high proficiency level ones according to their IELTS scores. Participants were divided into a group exposed to split-attention reading instructional format and a group exposed to integrated reading instructional format. While the split-attention format involves vocabulary

testing questions following the online reading text, the integrated format involves vocabulary testing questions integrated into the reading text dividing it into shorter paragraphs (between three and five lines each); thus, the sixty participants were randomly assigned to four groups for the different levels of the treatment.

a. 15 low proficiency level learners exposed to split-attention format of online reading text.

b. 15 high proficiency level learners exposed to split-attention format of online reading text.

c. 15 low proficiency level learners exposed to integrated format of online reading text.

d. 15 high proficiency level learners exposed to integrated format of online reading text.

Before the administration of the post-test, all the groups were exposed to the website which included the bank of reading texts in both integrated and split-attention formats. Reading texts were chosen from Schmitt;Schmitt and Mann (2011). There was a selection of reading texts out of which the researcher chose a few ones for this study. The texts and the vocabulary exercises were written in a website designed especially for the sake of this study, and called The Vocabulary Learning Assistant. Link for the website is: http://vocabulary-language-assistant.com/

The data was analysed descriptively and with the use of inferential statistics.

6.3 Research Conclusion

An overview of the research findings suggest that there was an increase in the number of participants who answered the vocabulary questions correctly after the experiment in both split-attention and integrated groups. This increase was mainly in word meaning and word family questions. A paired-samples t-test showed a statistically significant increase in the test scores from pre-test to post-test, and the results showed that the increase was statistically significant in word meaning and collocation knowledge, but not in word family knowledge; however, there was no statistically significant difference between the results of the split-attention format and the integrated format group in the post-test although the scores of the integrated format group were higher than those of the split-attention format group. Moreover, a Chi-square test showed that the treatment group gave more correct answers in the post-test but the difference was not statistically significant in either word meaning, word family or collocation knowledge tests, so

it could be concluded that the different instructional formats did not significantly affect the vocabulary knowledge building in this study. While participants of higher proficiency level gave more correct answers than those of lower proficiency level in the split-attention format group, the difference in the integrated format group was not statistically significant, L1 did not have any statistically significant effect on acquiring new vocabulary items in either split-attention format or treatment format groups.

6.3.1 The effect of online reading tasks on vocabulary knowledge (including word meaning, word families, and collocations knowledge)

The online reading texts have helped increase the vocabulary knowledge among the learners in general in this experiment. They also had the potential of helping increase the word meaning knowledge among the learners, enhanced the word family knowledge, and the collocations knowledge among the participants in this study.

This study is believed to have led to better results because learners were given the reading texts over three sessions and were tested on vocabulary knowledge within each session. The questions in pre- and post-tests were taken from the within-experiment questions. Moreover, the treatment was chosen to be as simple as possible with no distractors to the text, so no images, bright colours, or hyperlinks were used in order for the researcher to be able to filter out the distractors and minimise the cognitive load. In addition, the reading text in the present study had the words repeated a few times within the text since some research has shown that repetition and frequency of occurrence may enhance the vocabulary learning process.

6.3.2 The effect of the online reading split-attention and integrated instructional formats on vocabulary knowledge (including word meaning, word families, and collocations knowledge)

No statistically significant difference in scores was found between the two groups although the integrated format group had higher scores than the split-attention format group.

Descriptive and inferential statistics which were used to investigate the association between, and the influence of different reading instructional formats on the three aspects of vocabulary knowledge tested in this study; namely, word meaning knowledge, word families knowledge, and collocations knowledge, resulted in statistically insignificant findings. Although gains in the three aspects of vocabulary knowledge were found by comparing the score results, the

statistically insignificant findings indicate that the different instructional formats of integrated and split-attention had no effect on the vocabulary gains.

6.3.3 The effect of different proficiency levels on vocabulary knowledge (including word meaning, word families, and collocations knowledge) in both split-attention and integrated instructional formats

The inferential statistics used to investigate the effect of proficiency level on general vocabulary knowledge among the four groups showed that the split-attention format low proficiency group had significantly lower scores in the post-test than both groups with high-proficiency level, whereas the integrated format low proficiency group was not affected.

When only the word meaning element of the vocabulary knowledge was tested, the results showed that the split-attention group with high proficiency level had the highest score, while the integrated low proficiency group had the lowest score.

When only the word family element of the vocabulary knowledge was tested, the results showed that whereas the integrated group with high proficiency level had the highest score, the split-attention low proficiency group had the lowest score.

When only the collocations knowledge element of the vocabulary knowledge was tested, the results show that while the integrated group with high proficiency level had the highest score, the integrated low proficiency group had the lowest score.

The results of the word family and the collocations components were not statistically significant though.

It can be said that although higher proficiency groups had higher scores than lower proficiency groups, the inconsistency of which instructional format resulted in more gains, and the statistically insignificant results, especially within the groups of low numbers, could be interpreted in different ways.

6.3.4 The effect of L1 on vocabulary knowledge (including word meaning, word families, and collocations knowledge) in both split-attention and integrated instructional formats

The results of the test which investigated the effect of L1 on vocabulary knowledge showed that L1 did not have any statistically significant effect on acquiring new vocabulary items in either online reading split-attention or integrated instructional formats.

When only the word meaning element of the vocabulary knowledge was investigated, the results showed that while the Chinese L1 split-attention group had the highest score, the Arabic L1 integrated group had the lowest score.

The only statistically significant result was the one which tested the word family knowledge element of the vocabulary knowledge. Whereas the Chinese L1 integrated format group had the highest score, the Arabic L1 split-attention group had the lowest score. This result can be interpreted to be caused by the grammar elements which exists in the word family knowledge.

When only the collocations knowledge element of the vocabulary knowledge was investigated, the results showed that the Chinese L1 integrated group had the highest score whereas the Arabic L1 integrated group had the lowest score.

Although the common factor among all the above-mentioned results is that the groups of L1 Chinese had higher scores than the groups of L1 Arabic, no matter what the instructional formats these groups were exposed to were, the results were inconsistent, and except for those investigating the word family element of the vocabulary knowledge, they were statistically non-significant.

All the above-mentioned techniques of using the L1 knowledge to learn a word in L2 may or may not have been used by the learners in this experiment. Maybe a questionnaire or interviews with the learners may have helped the researcher have an insight on how the participants used L1 knowledge to help them with the post-tests; especially in the word family knowledge test where the results were significant.

6.4 Pedagogical implications of the study

As mentioned above, the findings of this study suggested that whereas the online reading treatment had an effect on the vocabulary knowledge, whether this was in word meaning, word

families or collocations, the two instructional formats did not provide strong support for the vocabulary knowledge whether that was in word meaning, word families, or collocations. Moreover, there were no consistent results in the findings of the effect of proficiency levels or L1 on vocabulary knowledge whether that was in word meaning, word families, or collocations. In addition to that , most of the results were statistically insignificant. Taking all this into consideration, some implications for vocabulary teaching will follow.

First, since the findings of this study are in agreement with many research results investigating the role of reading in general, and online reading specifically on vocabulary knowledge, it is suggested that these results could be useful for learners studying English as a second language. The vocabulary knowledge of these students could be improved if they were exposed to online reading texts which contain the target vocabulary in context, which make sure the target vocabulary is repeated within the context to enhance the learning process, and which are designed with the consideration of minimising the cognitive load on the working memory of these learners; thus it is recommended that the online text should be as simple as possible with no distractors to the text, no distracting images, extremely bright colours, or unneeded hyperlinks should be used so that the teachers could help the learners focus on the target vocabulary within the reading text as much as possible.

Second, the findings of the effect of instructional formats on vocabulary knowledge do not significantly support the statement of the thesis which is written in the introduction chapter, and which states: “Using an integrated instructional format leads to higher vocabulary gains compared with a split-attention format”; however, although statistically non-significant, all the integrated format group scores are higher than the split-attention format group scores. Some implications for teaching in this respect are: It can be recommended when designing vocabulary teaching material to divide the teaching material into smaller sections and inserting the questions or the teaching tools between these shorter sections in order to reduce the cognitive load on the working memory of the learners; however, for this technique to work best, a few more points should be taken into consideration, such as the intrinsic cognitive load of the learners which is defined by Brunken et al. (2003) as the cognitive load which is generated by the innate difficulty, instead of the instructional design of the subject; thus, for some language learners certain aspects of the language could be more difficult than learning other aspects. Because intrinsic and extraneous loads are extra burden on working memory (Paas et al., 2003), researchers recommend, and this is supported in this study, that the instructional format in any learning material should be designed cautiously, mainly when the innate level of difficulty in

the content, or the intrinsic load, is high. This point of learners' own innate ability is referred to by Yeung et al. (1998) who found the integrated format to be useful in decreasing the cognitive load and improving reading comprehension for both first language (L1) and second language (L2) learners. In addition to the intrinsic cognitive load, some individual differences, such as learners' academic individual abilities and their working memory span should be considered when designing or choosing online teaching materials. For example, very complicated webpages or questions wording may not work with learners with lower individual abilities.

The third implication when designing or choosing vocabulary teaching material is taking into consideration the concept of germane cognitive load and the concept of automation. Germane cognitive load is the third type of cognitive load which can be useful and productive during the learning process. Germane cognitive load helps learners be more aware during the learning process and is connected with over-sufficient schema construction and automatization. Schema creation or development is believed to be able to lead to reduced cognitive load. Some inconsistency in the results of post-tests investigating the effect of the different proficiency levels on the three elements of vocabulary knowledge among groups of learners who were exposed to different instructional formats can be accounted for by the concept of automation; the stage which advanced learners reach where operation of information can occur either consciously in limited working memory or automatically with no influence from the working memory. Automatization happens by practice, and it minimises the cognitive load on working memory. In other words, what is sometimes increased by extraneous cognitive load, can be decreased by germane cognitive load, and vice versa. It can be argued in this context that an integrated format allows for a smooth transfer of a new concept from working memory into a schema because it is immediate. A split-attention format, because of the delay, does not have this immediacy.

To summarise, this study has given the researcher some implications for teaching or teacher training. Online reading combined with tasks in certain instructional formats has the potential to improve the vocabulary knowledge among EAP learners; thus, learners should be exposed to online reading texts taking into consideration that repetition of the elements which are being taught can enhance the learning process as long as the cognitive load on the learners' working memory is minimised. Teaching materials should be designed cautiously, especially when the innate level of difficulty in the content, or the intrinsic load, is high. The complexity of the subject matter itself determines the intrinsic cognitive load. Reducing intrinsic cognitive load

by dividing the reading texts into smaller sections, providing clear instructions, eliminating distractors or unnecessary information, focussing on essential content, and effectively presenting visual content can be beneficial in reducing the cognitive load on the working memory. Moreover, individual differences, such as the learners' academic individual abilities, their first language, or their working memory span should also be considered very carefully. Germane cognitive load and automation are also very important elements to think about when designing a teaching material. Finally, it is worth considering to generalise the above-mentioned implications to designing language teaching materials focussing on aspects other than reading and vocabulary. It would be interesting to apply these implications to designing writing, speaking, or listening teaching materials, or to explore the learning of grammar using the same implications.

6.5 Limitations of the study

Narrowing the scope and focus of this research has required some delimitations, such as time, geographical, population, variable, and statistical delimitations. This study is designed to explore the impact of integrated and split-attention online reading formats on building adult EAP learners' vocabulary knowledge. It has solely focused on the influence of the online split-attention and integrated instructional formats on EAP learners' vocabulary acquisition in particular by investigating whether there is a relation between cognitive load and building vocabulary knowledge. The research was conducted at a language centre in a UK university to ensure relevance to the intended population, which is EAP learners. The study targets a group of adult learners studying English for University Studies. The study considers only quantitative data related to word meaning, word family, and collocations acquisition among the population targeted. A quasi-experimental intervention study was sued to collect the quantitative data. Although the instructional design of the study is thoughtful and organised, it has some limitations.

This study is limited to the examination of the effect of extraneous cognitive load caused by different instructional formats, such as the split-attention and the integrated instructional formats, on the vocabulary knowledge building. The effect of the intrinsic cognitive load and the germane cognitive load was not thoroughly considered when the study was designed. Some strategies were used in order to reduce intrinsic and germane cognitive load, such as breaking down the reading text into smaller passages, providing the learners with guidance and clear instructions, eliminating redundant details and distractors from the teaching materials, and

concentrating on the necessary content without overwhelming the students; however, better reduction of intrinsic and germane cognitive load occurs by activating prior knowledge and taking into consideration what the learners already know by relating the material into their existing mental schema. This element of reducing intrinsic and germane cognitive load is missing in this research.

Although the effect of the different instructional formats on the vocabulary knowledge was meant to be investigated by looking at the cognitive load on the working memory caused by each of these instructional formats, the working memory span of the learners or their learning aptitude was not considered when dividing them into groups. In other words, the individual differences of the participants were not measured. The main reason why the working memory span was not measured is that it is demanding to run a working memory span test and exclude factors like "chunking" where learners can group items together during a digit span task. This can weaken the relationship between working memory span and other measures like IQ. Running such a test was felt to be complex and beyond this research. It requires a more longitudinal study focussing on proactive interference of old information with the new information which the students are trying to learn. While working memory span is very essential for more accurate results of the research, proficiency level was found by the researcher to be more applicable within the circumstances of this study.

Assigning the participants into low proficiency level learners and high proficiency level ones was according to their IELTS scores where learners with 5.5 IELTS score considered as the separating score between lower proficiency level learners and higher proficiency level learners because 5.5 is the score between 5 (which according to IELTS scoring method refers to the skill level "modest") and 6 (which according to IELTS scoring method refers to the skill level "competent"). A limitation of the study is not being able to divide the participants into groups building on the individual skills scores. Assigning the participants into different proficiency levels according to individual skills is believed by the researcher to have helped more in interpreting their results, especially with the word family knowledge results which are highly connected to the grammatical knowledge of the learners. Having copies of the learners' detailed scores was difficult at the time of data collection because of the data protection policies at INTO, Newcastle University.

Another limitation of this study is not having delayed post-tests after having the immediate post-tests. Having delayed post-tests would have been very efficient in testing the long-term effect of the treatment but this could not be achieved because of the participants' timelines.

It can be argued that the whole number of participants in this study is sixty participants. This means there was 30 participants in each of the split-attention format group and the integrated format group, and only 15 participants in each group when the participants were assigned to groups of higher and lower proficiency levels, and of different L1. These small numbers of participants in each group mean the study is not generalisable to wider population; however, this point may be controversial because the students are EAP learners. This may have external validity. From a trustworthiness/transferability point of view, very clear details on how the research was conducted was given by the researcher and future researchers will be able to decide how applicable the methods and findings in this study can be to their own contexts.

In addition to the small numbers of participants in each group, there were small numbers of questions focussing on each of the three aspects of vocabulary knowledge; namely, word meaning knowledge, word family knowledge, and collocations knowledge. This is believed by the researcher to be a very strong reason of having statistically insignificant results in most of the findings although most of the results made sense when compared with results of other literature in the field.

When analysing the data, most of the results considered were from the answers to the pre-test and the post-test in the study. Looking in more detail at the within-experiment answers may have given some further insight into other potential reasons of some results in the study.

The final limitation of the study to be mentioned is not collecting more data from the participants focussing on some techniques they may have used to relate their L1 vocabulary knowledge to their L2 vocabulary learning. A mixed method approach could have been more beneficial for the findings of this study. Designing a quantitative study required a lot of effort, and future researchers can base their work on the finding of the importance of using the integrated instructional format in teaching; however, a mixed method approach could have given more comprehensive understanding of the topic. Mixed methods can enhance research quality, but future researchers can use the methods which are more appropriate in their contexts.

6.6 Suggestions for further study

One of the recommendations for future studies is further investigating the influence of the intrinsic and cognate cognitive load on language teaching and learning, in addition to only investigating the effect of extraneous type of cognitive load.

Focussing more on the individual differences, such as individual proficiency levels and first language, in addition to the extraneous cognitive load is a suggestion for other studies

A lot of research has focussed on the effect of different instructional formats on reading comprehension. This study tried to only focus on second language vocabulary acquisition, but as mentioned before, small scale studies, do not lead to generalizable results. A suggestion for further research is investigating the effect of different instructional formats on second language vocabulary acquisition on a wider scale.

Another suggestion for further study is further investigating the influence of split-attention and integrated instructional formats on other skills like listening; or even for language production through writing or speaking.

Finally, a mixed methods approach can be suggested for more in depth and comprehensive understanding of the researched topic. A mixed method approach can provide a fuller picture of the findings by integrating both quantitative and qualitative data, it can balance the limitations of both approaches, and give stronger evidence in the findings.

References

1. 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**,** 499-517.

2. AKYEL, A. & ERÇETIN, G. 2009. Hypermedia reading strategies employed by advanced learners of English. *System,* 37**,** 136-152.

3. AL-AKLOBY, S. 2001. Teaching and learning English vocabulary in Saudi Arabian public schools: An exploratory study of some possible reasons behind students' failure to learn English vocabulary.

4. AL-HAZEMI, H. A. A.-G. 1993. *Low-level EFL vocabulary tests for Arabic speakers.* Swansea University.

5. 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**,** 383-402.

6. AL-SHEHRI, S. & GITSAKI, C. 2010. Online reading: a preliminary study of the impact of integrated and split-attention formats on L2 students' cognitive load. *ReCALL,* 22**,** 356-375.

7. ALAHIRSH, H. 2014. *Exploring the effectiveness of extensive reading on incidental vocabulary acquisition by EFL learners: an experimental case study in a Libyan University.* University of Nottingham.

8. ALSAHAFI, M. 2023. The Relationship Between Depth of Academic English Vocabulary Knowledge and Academic Success of Second Language University Students. *SAGE Open,* 13**,** 21582440231153342.

9. ALSAIF, A. 2011. *Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia.* Swansea University.

10. 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**,** 39-45.

11. 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**,** 21-33.

12. ANDERSON, J. R. 1983. *The Architecture of Cognition,* Cambridge MA, Harvard University Press.

13. ANDERSON, N. J. 2003. Scrolling, clicking, and reading English: Online reading strategies in a second/foreign language. *The Reading Matrix,* 3**,** 1-33.

14. ANTONENKO, P. D. & NIEDERHAUSER, D. S. 2010. The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior,* 26**,** 140-150.

15. ARDAC, D. & UNAL, S. 2008. Does the amount of on-screen text influence student learning from a multimedia-based instructional unit? *Instructional Science,* 36**,** 75-88.

16. BACHMAN, L. 2004a. Research Guidelines in TESOL: Alternative Perspectives Linking Observations to Interpretations and Uses in TESOL Research. *TESOL Quarterly,* 38**,** 723-728.

17. BACHMAN, L. F. 2004b. *Statistical analyses for language assessment,* Cambridge, Cambridge : Cambridge University Press.

18. BADDELEY, A. 2000. The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences,* 4**,** 417-423.

19. BADDELEY, A. 2013. *Essentials of human memory (classic edition),* Psychology Press.

20. BADDELEY, A. D. & HITCH, G. 1974. Working Memory. *In:* BOWER, G. H. (ed.) *Psychology of Learning and Motivation.* Academic Press.

21. BARCROFT, J. 2009. Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly,* 43**,** 79-103.

22. BASTURKMEN, H. & SHACKLEFORD, N. 2015. How content lecturers help students with language: An observational study of language-related episodes in interaction in first year accounting classrooms. *English for Specific Purposes,* 37**,** 87-97.

23. BECK, I. L., MCKEOWN, M. G. & KUCAN, L. 2013. *Bringing words to life: Robust vocabulary instruction,* Guilford Press.

24. BIBER, D., JOHANSSON, S., LEECH, G., CONRAD, S. & FINEGAN, E. 1999. Longman grammar of spoken and written English. Longman London.

25. BORNER, W. 1997. Implizites und explizites wissen im fremdsprachlichen Wortschatz. *Fremdsprachen Lehren und Lernen (FLuL),* 26**,** 44-67.

26. BOUMA, G. D., ATKINSON, G. B. J. & DIXON, B. R. 1995. *A handbook of social science research,* Oxford, Oxford : Oxford University Press.

27. BRITISH NATIONAL CORPUS (BNC).

28. BROWN, R., WARING, R. & DONKAEWBUA, S. 2008. Vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language,* 20**,** 136-163.

29. BRUNKEN, R., PLASS, J. L. & LEUTNER, D. 2003. Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist,* 38**,** 53-61.

30. BURKES, K. M. E. 2007. *Applying cognitive load theory to the design of online learning.* University of North Texas.

31. CAMPBELL, D. T. & STANLEY, J. C. 1963. Experimental and Quasi-experimental designs for research on teaching. *In:* GAGE, N. (ed.) *Handbook of Research on Teaching.* Chicago: Rnd McNally.

32. CARTER, R. & MCCARTHY, M. (eds.) 1988. *Vocabulary and Language Teaching,* London: Longman.

33. CHACON-BELTRAN, R., ABELLO-CONTESSE, C. & TORREBLANCA-LOPEZ, M. M. 2010. Vocabulary teaching and learning: introduction and overview. *In:* CHACON-BELTRAN, R., ABELLO-CONTESSE, C. & TORREBLANCA-LOPEZ, M. M. (eds.) *Insights into Non-native Vocabulary Teaching and Learning.* Bristol: Multilingual Matters.

34. CHALHOUB-DEVILLE, M. 2006. Drawing the line: the generalizability and limitations of research in applied linguistics. *In:* CHALHOUB-DEVILLE, M., CHAPELLE, C. A. & DUFF, P. (eds.) *Inference and Generalizability in Applied Linguistics: Multiple Perspectives.* Amsterdam: John Benjamins.

35. CHANG, L. & LI, J. 2009. A correlational study of metacognitive strategies, listening comprehension and incidental vocabulary acquisition. *Journal of the Foreign Language World,* 6**,** 50-57.

36. CHANNELL, J. 1981. Applying semantic theory to vocabulary teaching. *ELT Journal,* 35**,** 115-122.

37. CHAPELLE, C. 1999. Validity in Language Assessment. *Annual Review of Applied Linguistics,* 19**,** 254-272.

38. CHUN, D. M. 2001. L2 Reading on the Web: Strategies for Accessing Information in Hypermedia. *Computer Assisted Language Learning,* 14**,** 367-403.

39. CIERNIAK, G., SCHEITER, K. & GERJETS, P. 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior,* 25**,** 315-324.

40. CLARKE, D. D. 2004. "Structured judgement methods": The best of both worlds? *In:* TODD, Z., NERLICH, B., MCKEOWN, S. & CLARKE, D. D. (eds.) *Mixing Methods*

in Psychology: The Integration of Qualitative and Quantitative Methods in Theory and Practice. Hove, East Sussex: Psychology Press.

41. COADY, J. & HUCKIN, T. 1997. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy,* New York, Cambridge University Press.

42. COIRO, J. 2003. Reading Comprehension on the Internet: Expanding Our Understanding of Reading Comprehension to Encompass New Literacies. *The Reading Teacher,* 56**,** 458-464.

43. CONSTANTINESCU, A. I. 2007. Using technology to assist in vocabulary acquisition and reading comprehension. *The Internet TESL Journal,* 13**,** 122-133.

44. COOPER, G. 1998. Research into Cognitive Load Theory and Instructional Design at UNSW.

45. COXHEAD, A. 2000. A New Academic Word List. *TESOL Quarterly,* 34**,** 213-238.

46. COXHEAD, A. 2020. Academic Vocabulary. *In:* WEBB, S. (ed.) *The Routledge handbook of vocabulary studies.* Routledge.

47. CROSSMAN, K. & PINCHBECK, G. 2012. An Intensive Academic English Course for Generation1.5 ELLs Bound for Postsecondary Studies: Curriculum Design, Development, and Implementation. *TESL Canada Journal,* 29**,** 231.

48. DALLER, H., MILTON, J. & TREFFERS-DALLER, J. (eds.) 2007. *Modelling and Assessing Vocabulary Knowledge,* Cambridge: Cambridge University Press.

49. DANEMAN, M. & CARPENTER, P. A. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior,* 19**,** 450-466.

50. DANEMAN, M. & MERIKLE, P. M. 1996. Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review,* 3**,** 422-433.

51. DANG, T. N. Y., COXHEAD, A. & WEBB, S. 2017. The Academic Spoken Word List. *Language Learning,* 67**,** 959-997.

52. DANG, T. N. Y., LU, C. & WEBB, S. 2022. Incidental Learning of Collocations in an Academic Lecture Through Different Input Modes. *Language Learning,* 72**,** 728-764.

53. DANG, T. N. Y. & WEBB, S. 2014. The lexical profile of academic spoken English. *English for Specific Purposes,* 33**,** 66-76.

54. DAULTON, F. E. 1998. Japanese loanword cognates and the acquisition of English vocabulary. *The Language Teacher-Kyoto-Jalt-,* 22**,** 17-25.

55. DAY, R. R. & BAMFORD, J. 1998. *Extensive Reading in the Second Language Classroom,* Cambridge Cambridge University Press.

56. DAY, R. R., OMURA, C. & HIRAMATSU, M. 1991. Incidental EFL Vocabulary Learning and Reading. *Reading in a foreign language,* 7**,** 541-551.

57. DEKEYSER, R. M. 2000. The Robustness ff Critical Period Effects in Second Language Acquisition. *Studies in Second Language Acquisition,* 22**,** 499-533.

58. DESTEFANO, D. & LEFEVRE, J.-A. 2007. Cognitive load in hypertext reading: A review. *Computers in Human Behavior,* 23**,** 1616-1641.

59. DÖRNYEI, Z. 2010. Research Methods in Applied Linguistics, de Z. Dörnyei. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*.

60. DUPUY, B. & KRASHEN, S. 1993. Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning,* 4**,** 55-63.

61. ELLIS, N. 1994a. Consciousness in second languae learning: psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review,* 11**,** 37-56.

62. ELLIS, N. 1994b. Introduction: implicit and dexplicit language learning-an overview. *In:* ELLIS, N. (ed.) *Implicit and Explicit Learning of Languages.* London etc: Academic Press.

63. ELLIS, N. 1994c. Vocabulary acquisition: The implicits ins and outs of explicit cognitive mediation. *In:* ELLIS, N. (ed.) *Implicit and Explicit Learnign of Languages.* London etc: Academic Press.

64. ELLIS, N. C. 2005. At The Interface: Dynamic Interactions of Explicit and Implicit Language Knowledge. *Studies in Second Language Acquisition,* 27**,** 305-352.

65. ENGLE, R. W., TUHOLSKI, S. W., LAUGHLIN, J. E. & CONWAY, A. R. A. 1999. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General,* 128**,** 309-331.

66. FAN, M. 2000. How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal,* 31**,** 105-119.

67. FAN, Y. 2021. A Study on the Effects of Chinese EFL Learners' English Proficiency and Involvement Load on Incidental Vocabulary Acquisition. *International Journal of English Linguistics,* 11**,** 76-82.

68. FIRDA, I. N., AZKIYAH, I. & ALBIANSYAH 2021. Testing Breadth and Depth of Vocabulary Knowledge and Their Relationship with Vocabulary Size of EFL Students. *Journal of English Teaching,* 7**,** 89-100.

69. FREDERIKSEN, J. G., SØRENSEN, S. M. D., KONGE, L., SVENDSEN, M. B. S., NOBEL-JØRGENSEN, M., BJERRUM, F. & ANDERSEN, S. A. W. 2020. Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical Endoscopy,* 34**,** 1244-1252.

70. FRIEDERICI, A. D., STEINHAUER, K., MECKLINGER, A. & MEYER, M. 1998. Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology,* 47**,** 193-221.

71. GAI, S. & WEN, Q. 2013. Learning without awareness by Chinese EFL learners. *Foreign Language Teaching and Research,* 4**,** 557-567.

72. GAIRNS, R. & REDMAN, S. 1986. *Working with Words: A Guide to Teaching and Learning Vocabulary,* Cambridge University Press.

73. GARDNER, D. & DAVIES, M. 2014. A New Academic Vocabulary List. *Applied Linguistics,* 35**,** 305-327.

74. GASS, S. 1999. Discussion: Incidental vocabualry acquisition. *Studies in Second Language Acquisition,* 12**,** 319-333.

75. GENÇ, H. & GÜLÖZER, K. 2013. The Effect of Cognitive Load Associated With Instructional Formats and Types of Presentation on Second Language Reading Comprehension Performance. *Turkish Online Journal of Educational Technology,* 12.

76. GERJETS, P., SCHEITER, K. & CATRAMBONE, R. 2004. Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. *Instructional Science,* 32 (1-2)**,** 33-58.

77. GEVA, E. & RYAN, E. B. 1993. Linguistic and Cognitive Correlates of Academic Skills in First and Second Languages. *Language Learning,* 43**,** 5-42.

78. GOULDEN, R., NATION, I. S. P. & READ, J. 1990. How large can a receptive vocabulary be? *Applied Linguistics,* 11**,** 341-363.

79. GRAFF, M. 2003. Assessing Learning from Hypertext: An Individual Differences Perspective. *Journal of Interactive Learning Research,* 14**,** 425-438.

80. GYLLSTAD, H. 2007. *Testing English Collocations - Developing Receptive Tests for Use with Advanced Swedish Learners,* Lund University, Media-Tryck.

81. HALL, C. J. 2002. The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL, International Review of Applied Linguistics in Language Teaching,* 40**,** 69-87.

82. HAMMERSLEY, M. 2018. *What's wrong with ethnography? : methodological explorations,* London : Routledge.

83. HARRINGTON, M. 1991. Individual differences in L2 reading: Processing capacity versus linguistic knowledge. . Paper presented at the Annual Meeting of the American Association of Applied Linguistics., New York.

84. HARRINGTON, M. & SAWYER, M. 1992. L2 Working Memory Capacity and L2 Reading Skill. *Studies in Second Language Acquisition,* 14**,** 25-38.

85. HARRIS, V. & SNOW, D. 2004. *Classic Pathfinder: Doing it for themselves: focus on learning strategies and vocabulary building,* London, CILT.

86. HAZENBERG, S. & HULSTUN, J. H. 1996. Defining a Minimal Receptive Second-Language Vocabulary for Non-native University Students: An Empirical Investigation. *Applied Linguistics,* 17**,** 145-163.

87. HENNING, G. 1987. A guide to language testing: Development, evaluation, research.

88. HENRIKSEN, B. & DANELUND, L. 2015. Chapter two studies of Danish l2 learners' vocabulary knowledge and the lexical richness of their written production in English. *Lexical issues in L2 writing,* 29.

89. HORST, M. 2005. Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review,* 61**,** 355-382.

90. HORST, M., COBB, T. & MEARA, P. 1998. Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language,* 11**,** 207-223.

91. HOU, D. 2004. Involvement load hypothesis and senior middle school students' vocabulary retention. *Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition),* 5**,** 394-398.

92. HU, M. & NATION, P. 2000. Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language,* 13.

93. HUANG, H.-C., CHERN, C.-L. & LIN, C.-C. 2009. EFL learners' use of online reading strategies and comprehension of texts: An exploratory study. *Computers & Education,* 52**,** 13-26.

94. HUCKIN, T. & COADY, J. 1999. Incidental vocabulary acquisition in a second language: a review. *Studies in Second Language Acquisition,* 21**,** 181-193.

95. HUCKIN, T., HAYNES, M. & COADY, J. 1993. *Second language reading and vocabulary learning,* Norwood, NJ, Ablex.

96. HUGHES, A. 2003. *Testing for Language Teachers,* Cambridge, Cambridge University Press.

97. HULSTIJN, J., H. 2001. Implicit and incidental second language learning: experiments in processing of natural and partly artificial input. *In:* DECHERT, H. W. & RAUPACH, M. (eds.) *Interlingual processes.* Tubingen: Narr.

98. HULSTIJN, J. H. 1992. Retention of Inferred and Given Word Meanings: Experiments in Incidental Vocabulary Learning. *In:* ARNAUD, P. J. L. & BÉJOINT, H. (eds.) *Vocabulary and Applied Linguistics.* London: Palgrave Macmillan UK.

99. HUNG, H. C. M. Applying cognitive load theory in reading comprehension. Proceedings of the 5th CamTESOL Conference on English Language "The Globalisation of ELT: Emerging Directions, 2009. 184-196.

100. HYLTENSTAM, K. & ABRAHAMSSON, N. 2004. Language aptitude and near-native L2 proficiency. *Attaining the Ultimate.* Radbound University Nijmegen.

101. JOPPE, M. 2000. The Research Process. [Accessed 23 October 2008].

102. JUFFS, A. 2004. Representation, Processing and Working Memory in a Second Language. *Transactions of the Philological Society,* 102**,** 199-225.

103. JUST, M. A. & CARPENTER, P. A. 1992. A capacity theory of comprehension: individual differences in working memory. *Psychol Rev,* 99**,** 122-49.

104. KALYUGA, S., CHANDLER, P. & SWELLER, J. Levels of expertise and user-adapted formats of instructional presentations: A cognitive load approach. User Modeling, 1997. Springer, 261-272.

105. KALYUGA, S., CHANDLER, P. & SWELLER, J. 1998. Levels of Expertise and Instructional Design. *Human Factors,* 40**,** 1-17.

106. KELLER, T., GERJETS, P., SCHEITER, K. & GARSOFFKY, B. 2006. Information visualizations for knowledge acquisition: The impact of dimensionality and color coding. *Computers in Human Behavior,* 22**,** 43-65.

107. KING, J. & JUST, M. A. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language,* 30**,** 580-602.

108. KRASHEN, S. 1981. *Second Language Acquisition and Second Language Learning,* Oxford, Pergamon Press.

109. KRASHEN, S. D. & TERRELL, T. D. 1983. *The Natural Approach: Language Acquisition in the Classroom,* San Francisco, Alemany Press.

110. LAUFER, B. 1992. How Much Lexis is Necessary for Reading Comprehension? *In:* ARNAUD, P. J. L. & BÉJOINT, H. (eds.) *Vocabulary and Applied Linguistics.* London: Palgrave Macmillan UK.

111. LAUFER, B. 2000. Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. Selected Papers from AILA'99 Tokyo (pp. 47-62). Tokyo: Waseda University Press.

112. LAUFER, B. 2003. Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *Canadian Modern Language Review-revue Canadienne Des Langues Vivantes,* 59**,** 567-587.

113. LAUFER, B. 2005a. Focus on form in second language vocabulary learning. *EUROSLA Yearbook,* 5**,** 223-250.

114. LAUFER, B. & GOLDSTEIN, Z. 2004. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning,* 54**,** 399-436.

115. LAUFER, B. & HULSTIJN, J. 2001. Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics,* 22**,** 1-26.

116. LEAHY, W. & SWELLER, J. 2004. Cognitive load and the imagination effect. *Applied Cognitive Psychology,* 18**,** 857-875.

117. LEVY, M. & STOCKWELL, G. 2006. *CALL dimensions: options and issues in computer assisted language learning,* New Jersey, Lawrence Erlbaum Associates.

118. LEWIS, M. 1993. *The Lexical Approach: The state of ELT and a way forward,* Hove, UK, Language Teaching Publications.

119. LI, H. & TIAN, Q. 2005. On the incidental vocabulary acquisition of second language. *Foreign Language Educational,* 3**,** 52-55.

120. LI, J. & SCHMITT, N. 2009. The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing,* 18**,** 85-102.

121. LI, Y. & ZHENG, S. 2008. An empirical study of the incidental acquisition of idioms. *Journal of Sichuan College of Education,* 11**,** 83-85.

122. LIN, C. C. & HIRSH, D. 2012. Manipulating instructional method: The effect on productive vocabulary use. *Linguistic Insights,* 155**,** 117-148.

123. LUXTON, J., FRY, J. & COXHEAD, A. 2017. Exploring the Knowledge and Development of Academic English Vocabulary of Students in New Zealand Secondary Schools. *set: Research Information for Teachers,* 1**,** 12-22.

124. LYNCH, B. K. 2001. Language Assessment and Program Evaluation. *TESOL Quarterly,* 35**,** 603-605.

125. MACKEY, A. & GASS, S. M. 2005. *Second language research : methodology and design,* Mahwah, NJ

126. Lawrence Erlbaum.

127. MARSHALL, C. & ROSSMAN, G. B. 2016. *Designing qualitative research,* Los Angeles, California : SAGE.

128. MASRAI, A. 2015. Investigating and explaining the relationship between L1 mental lexicon size and organisation and L2 vocabulary development. *Unpublished PhD thesis, Swansea University, UK*.

129. MAXWELL, L. A. 2013. *Common Core Ratchets Up Language Demands for English-Learners* [Online]. Education Week. Available: https://www.edweek.org/teaching-learning/common-core-ratchets-up-language-demands-for-english-learners/2013/10 [Accessed].

130. MAYER, R. E. 1997. Multimedia learning: Are we asking the right questions? *Educational Psychologist,* 32**,** 1-19.

131. MCCARTHY, M. 1990. *Vocabulary,* Oxford Oxford University Press.

132. MCKEOWN, M., BECK, I. L., OMANSON, R. C. & POPLE, M. T. 1985. Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words'. *Reading Research Quarterly,* 20**,** 522-535.

133. MEARA, P. 1990. A note on passive vocabulary. *Second Language Research,* 6**,** 150-154.

134. MEARA, P. & WOLTER, B. 2004. V_Links, beyond vocabulary depth. *Angles on the English Speaking World,* 4**,** 85-96.

135. MELKA, F. 1997. Receptive vs. productive aspects of vocabulary. *In:* SCHMITT, N. & MCCARTHY, M. (eds.) *Vocabulary: Description, Acquisition, and Pedagogy.* Cambridge: Cambridge University Press.

136. MIKK, J. 2008. Sentence length for revealing the cognitive load reversal effect in text comprehension. *Educational Studies,* 34**,** 119-127.

137. MILLER, G. A. 1999. On knowing a word. *Annual Review of Psychology,* 50**,** 1-19.

138. MILTON, J. 2008. Vocabulary uptake from informal learning tasks. *Language Learning Journal,* 36**,** 227-238.

139. MILTON, J. 2009. *Measuring Second Language Vocabulary Acquisition,* Clevedon, Multilingual Matters.

140. MILTON, J. & DONZELLI, G. 2013. The acquisition of the lexicon. *In:* HERCHENSOHN, J. & YOUNG-SCHOLTEN, M. (eds.) *Handbok of Second Language Acquisition.* Cambridge: Cambridge University Press.

141. MIYAKE, A. & FRIEDMAN, N. P. 1998. Individual differences in second language proficiency: Working memory as language aptitude. *Foreign language learning: Psycholinguistic studies on training and retention***,** 339-364.

142. MORENO, R. & VALDEZ, A. 2005. Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development,* 53**,** 35-45.

143. MUTLU, D., OZEL, P., ALTINDIS, F. & YILMAZ, B. 2022. Split-attention effects in multimedia learning environments: eye-tracking and EEG analysis. *Multimedia Tools and Applications,* 81**,** 8259-8282.

144. NATION, I. 2006. How large a vocabulary is needed for reading and listening? *Canadian modern language review,* 63**,** 59-82.

145. NATION, I. S. P. 1990. *Teaching and Learning Vocabulary* New York, Newbury House.

146. NATION, I. S. P. 2001. *Learnig Vocabulary in Another Language,* Cambridge, Cambridge University Press.

147. NATION, I. S. P. 2016. Making and using word lists for language learning and testing. *Making and Using Word Lists for Language Learning and Testing***,** 1-224.

148. NATION, I. S. P. 2022. *Learning Vocabulary in Another Language,* Cambridge, Cambridge University Press.

149. NATION, I. S. P. & CARTER, R. 1989. *Vocabulary acquisition, AILA Review, 6,* Amsterdam, Association Internationale de Linguistique Appliquee.

150. NATION, I. S. P. & WANG, K. 1999. Graded readers and vocabulary. *Reading in a Foreign Language,* 12.

151. OOI, D. & LEE KIM-SEOH, J. 1996. Vocabulary Teaching: Looking behind the word. *ELT Journal,* 50**,** 52-59.

152. OSAKA, M. & OSAKA, N. 1992. Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society,* 30**,** 287-289.

153. OSAKA, M., OSAKA, N. & GRONER, R. 1993. Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society,* 31**,** 117-118.

154. OZONO, S. & ITO, H. 2003. Logical connectives as catalysts for interactive L2 reading. *System,* 31**,** 283-297.

155. PAAS, F., RENKL, A. & SWELLER, J. 2003. Cognitive load theory and instructional design: Recent developments. *Educational psychologist,* 38**,** 1-4.

156. PAAS, F., VAN GOG, T. & SWELLER, J. 2010. Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. *Educational Psychology Review,* 22**,** 115-121.

157. PATTON, M. Q. 1990. *Qualitative evaluation and research methods,* Newbury Park, Calif., Newbury Park, Calif. : Sage Publications.

158. PAWLEY, A. & SYDER, F. H. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *In:* RICHARDS, J. C. & SCHMIDT, R. W. (eds.) *Language and communication.*

159. PETERS, E. 2007. Manipulating L2 learners' online dictionary use and its effect on L2 word retention.

160. PIENEMANN, M. 1980. The second language acquisition of immigrant children. *In:* FELIX, S. W. (ed.) *Second Language Development. Trends and Issues.* Tubingen: Narr.

161. PIENEMANN, M. 1984. Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition,* 6**,** 186-214.

162. PIENEMANN, M. 1985. Leanability and syllabus construction. *In:* HYLTENSTAM, K. & PIENEMANN, M. (eds.) *Modelling and Assessing Second Language Development.* Clevedon, Avon: Multilingual Matters.

163. PIGADA, M. & SCHMITT, N. 2006. Vocabulary acquisition from extensive reading: Acase study. *Reading in a Foreign Language,* 18**,** 1-28.

164. PILLAY, H. K. 1994. Cognitive load and mental rotation: structuring orthographic projection for learning and problem solving. *Instructional Science,* 22**,** 91-113.

165. PITTS, M., WHITE, H. & KRASHEN, S. 1989. Acquiring the second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language,* 5**,** 271-275.

166. PROTOPSALTIS, A. 2008. Reading strategies in hypertexts and factors influencing hyperlink selection. *Journal of Educational Multimedia and Hypermedia,* 17**,** 191-213.

167. QIAN, D. D. 2004. Second language lexical inferencing: Preferences, perceptions and practices. *In:* BOGAARDS, P. & LAUFER, B. (eds.) *Vocabulary in a Second Language.* Amsterdam: John Benjamin Publishing Company.

168. REIDER, A. 2003. Implicit and explicit leanring in incidental vocabulary acquisition. *VIEWS,* 12**,** 24-39.

169. RENANDYA, W. A., RAJAN, B. R. S. & JACOBS, G. M. 1999. Extensive reading with adult learners of English as a second language. *RELC Journal,* 30**,** 39-60.

170. RICHARDS, K., ROSS, S. & SEEDHOUSE, P. 2011. *Research methods for applied language studies,* Milton Park, Abingdon, Oxon

171. New York, NY, Milton Park, Abingdon, Oxon

172. New York, NY : Routledge.

173. RIDING, R. & CHEEMA, I. 1991. Cognitive Styles—an overview and integration. *Educational Psychology,* 11**,** 193-215.

174. RIVERS, W. M. 1983. *Speaking in Many Tongues,* Cambridge Cambridge University Press.

175. SARAGI, T., NATION, I. S. P. & MEISTER, G. F. 1978. Vocabulary learning and reading. *System,* 6**,** 72-78.

176. SCHMITT, D. & SCHMITT, N. 2011. *Focus on vocabulary 2 : mastering the academic word list,* White Plains, N.Y., White Plains, N.Y. : Longman.

177. SCHMITT, D., SCHMITT, N. & MANN, D. 2011a. *Focus on vocabulary 1: Bridging Vocabulary,* White Plains, N.Y., White Plains, N.Y. : Pearson Longman.

178. SCHMITT, N. 2000. *Vocabulary in Langauge Teaching,* Cambridge, Cambridge University Press.

179. SCHMITT, N. 2008. Instructed second language vocabulary learning. *Language teaching research,* 12**,** 329-363.

180. SCHMITT, N. 2010. *Researching Vocabulary: A Vocabulary Research Manual,* UK, Palgrave Macmillan.

181. SCHMITT, N., JIANG, X. & GRABE, W. 2011b. The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal,* 95**,** 26-43.

182. SCHMITT, N. & MCCARTHY, M. 1997. *Vocabulary: Description, acquisition and pedagogy,* New York, Cambridge University Press.

183. SCHMITT, N., SCHMITT, D. & CLAPHAM, C. 2001. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing,* 18**,** 55-88.

184. SCHMITT, N. & ZIMMERMAN, C. B. 2002. Derivative word forms: What do learners know? *Tesol Quarterly,* 36**,** 145-171.

185. SCHNOTZ, W. & KÜRSCHNER, C. 2007. A Reconsideration of Cognitive Load Theory. *Educational Psychology Review,* 19**,** 469-508.

186. SEASHORE, R. H. & ECKERSON, L. D. 1940. The measurement of individual differences in general English vocabulary. *Journal of Educational Psychology,* 31**,** 14-38.

187. SINGLETON, D. 1999. *Exploring the Second Language Mental Lexicon,* Cambridge, C.U.P.

188. SKULMOWSKI, A. & XU, K. M. 2022. Understanding Cognitive Load in Digital and Online Learning: a New Perspective on Extraneous Cognitive Load. *Educational Psychology Review,* 34**,** 171-196.

189. SMITH, G. T. 2005. On Construct Validity: Issues of Method and Measurement. *Psychological Assessment,* 17**,** 396-408.

190. STAHL, S. 1983. Differential word knowledge and reading comprehension. *Journal of Reading Behaviour,* 15**,** 33-50.

191. STERNBERG, R. J. 1987. Most vocabulary is learned from context. *The nature of vocabulary acquisition.* Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

192. STORCH, N. & TAPPER, J. 2009. The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes,* 8**,** 207-223.

193. SWEET, H. 1899, 1964. *The Practical Study of Languages,* Oxford, Oxford University Press.

194. SWELLER, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science,* 12**,** 257-285.

195. SWELLER, J., AYRES, P. & KALYUGA, S. 2011. *Cognitive load theory,* New York, New York : Springer.

196. SWELLER, J. & CHANDLER, P. 1991. Evidence for Cognitive Load Theory. *Cognition and Instruction,* 8**,** 351-362.

197. SWELLER, J., VAN MERRIËNBOER, J. J. G. & PAAS, F. 2019. Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review,* 31**,** 261-292.

198. SWELLER, J., VAN MERRIENBOER, J. J. G. & PAAS, F. G. W. C. 1998. Cognitive Architecture and Instructional Design. *Educational Psychology Review,* 10**,** 251-296.

199. TENG, F. 2019. The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *The Language Learning Journal,* 47**,** 145-158.

200. TONZAR, C., LOTTO, L. & JOB, R. 2009. L2 Vocabulary Acquisition in Children: Effects of Learning Method and Cognate Status. *Language Learning,* 59**,** 623-646.

201. TOWELL, R. & HAWKINS, R. D. 1994. *Approaches to Second Language Acquisition,* Longdunn Press, Bristol, Multilingual Matters.

202. TOZCU, A. & COADY, J. 2004. Successful Learning of Frequent Vocabulary through CALL also Benefits Reading Comprehension and Speed. *Computer Assisted Language Learning,* 17**,** 473-495.

203. TURNER, M. L. & ENGLE, R. W. 1989. Is working memory capacity task dependent? *Journal of Memory and Language,* 28**,** 127-154.

204. VAN MERRIËNBOER, J. J. G. & SWELLER, J. 2005. Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review,* 17**,** 147-177.

205. VAN SCHAIK, P. & LING, J. 2003. The effect of link colour on information retrieval in educational intranet use. *Computers in Human Behavior,* 19**,** 553-564.

206. VAN ZEELAND, H. & SCHMITT, N. 2013. Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics,* 34**,** 457-479.

207. VASSILIU, P. 2001. *Lexical Input and Uptake in the Low Level EFL Classroom.* PhD Unpublished PhD, University of Wales Swansea.

208. VERMEER, A. 2001. Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics,* 22**,** 217-234.

209. VOS, S. H. 1999. *Verbal working memory and sentence processing: An electrophysiological investigation.* Katholieke Universiteit Nijmegen, The Netherlands.

210. WANG, Q. 2020. Incidental L2 vocabulary acquisition and reading: Concerns, progresses and future directions. *Chinese Journal of Applied Linguistics,* 43**,** 469-488.

211. WARING, R. 1999. *Tasks for assessing second language receptive and productive vocabulary.* Unpublished PhD, University of Wales, Swansea.

212. WARING, R. & TAKAKI, M. 2003. At what rate do learnerslearn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language,* 15**,** 130-163.

213. WATERS, G. S. & CAPLAN, D. 1996. The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review,* 103**,** 761-772.

214. WATERS, G. S. & CAPLAN, D. 2004. Verbal Working Memory and on-Line Syntactic Processing: Evidence from self-Paced Listening. *The Quarterly Journal of Experimental Psychology Section A,* 57**,** 129-163.

215. WEBB, S. 2007. The effects of repetition on vocabulary knowledge. *Applied linguistics,* 28**,** 46-65.

216. WEBB, S. A. & CHANG, A. C.-S. 2012. Second language vocabulary growth. *RELC journal,* 43**,** 113-126.

217. WEIR, C. J. 2005. Language testing and validation. *Hampshire: Palgrave McMillan*.

218. WESCHE, M. & PARIBAKHT, T. S. 1996. Assessing second language vocabulary knowledge: depth versus breadth. *The Canadian Modern Language Review/ La Canadienne des Langue Vivantes,* 53**,** 13-40.

219. WEST, R. F. & STANOVICH, K. E. 1991. The Incidental Acquisition of Information from Reading. *Psychological Science,* 2**,** 325-330.

220. WILKINS, D. A. 1972. *Linguistics in Language Teaching,* London, Arnold.

221. WU, J., LANG, J. & DANG, Q. 2007. Incidental vocabulary acquisition and involvement load hypothesis. *Foreign Language Teaching and Research,* 5**,** 360-366.

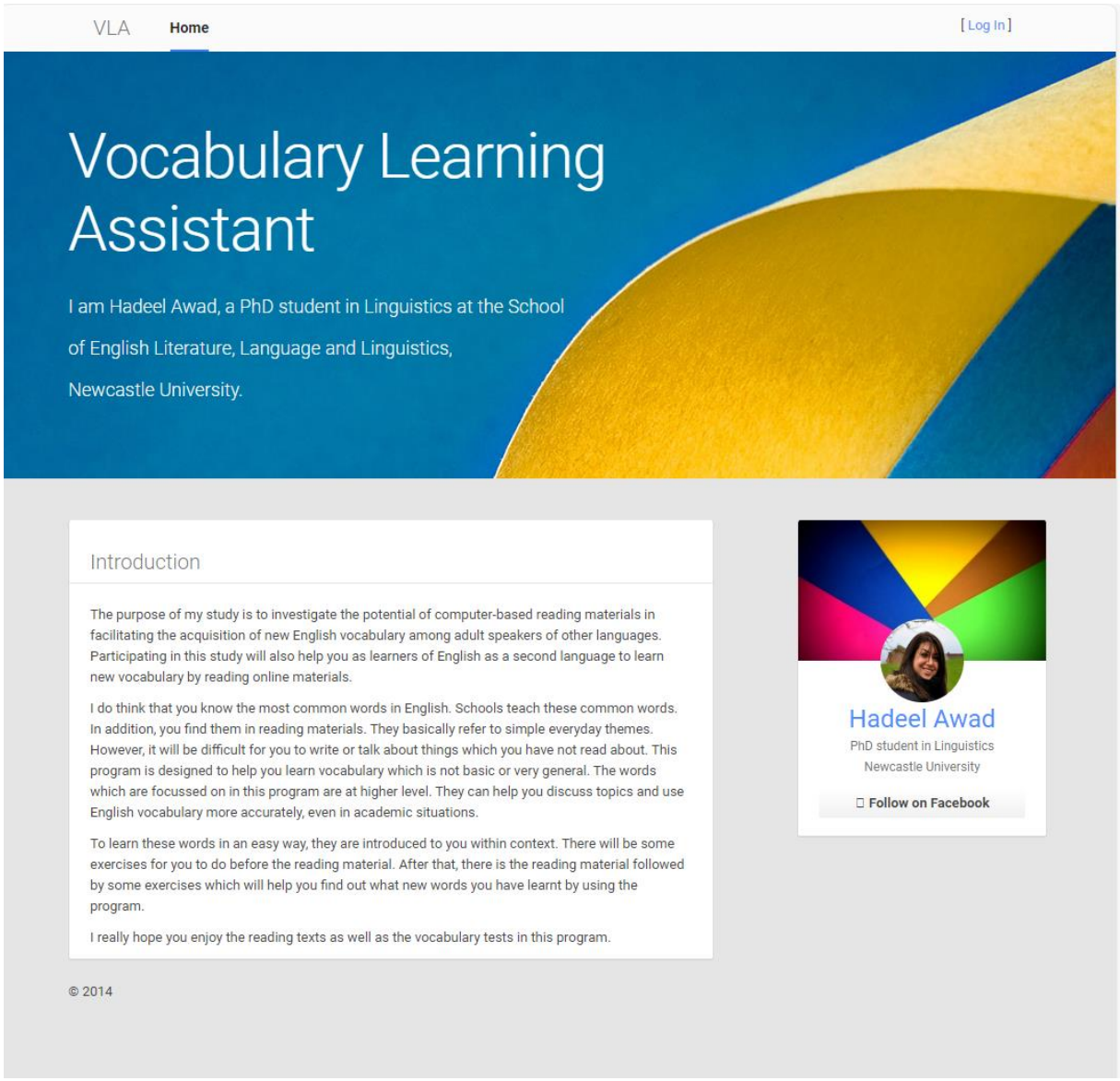
222. YEUNG, A. S., JIN, P. & SWELLER, J. 1998. Cognitive Load and Learner Expertise: Split-Attention and Redundancy Effects in Reading with Explanatory Notes. *Contemporary Educational Psychology,* 23**,** 1-21.

223. ZIMMERMAN, C. B. 1997. Historical trends in second language vocabulary instruction. *In:* COADY, J. & HUCKIN, T. (eds.) *Second Language Vocabulary Acquisition.* Cambridge: Cambridge University Press.

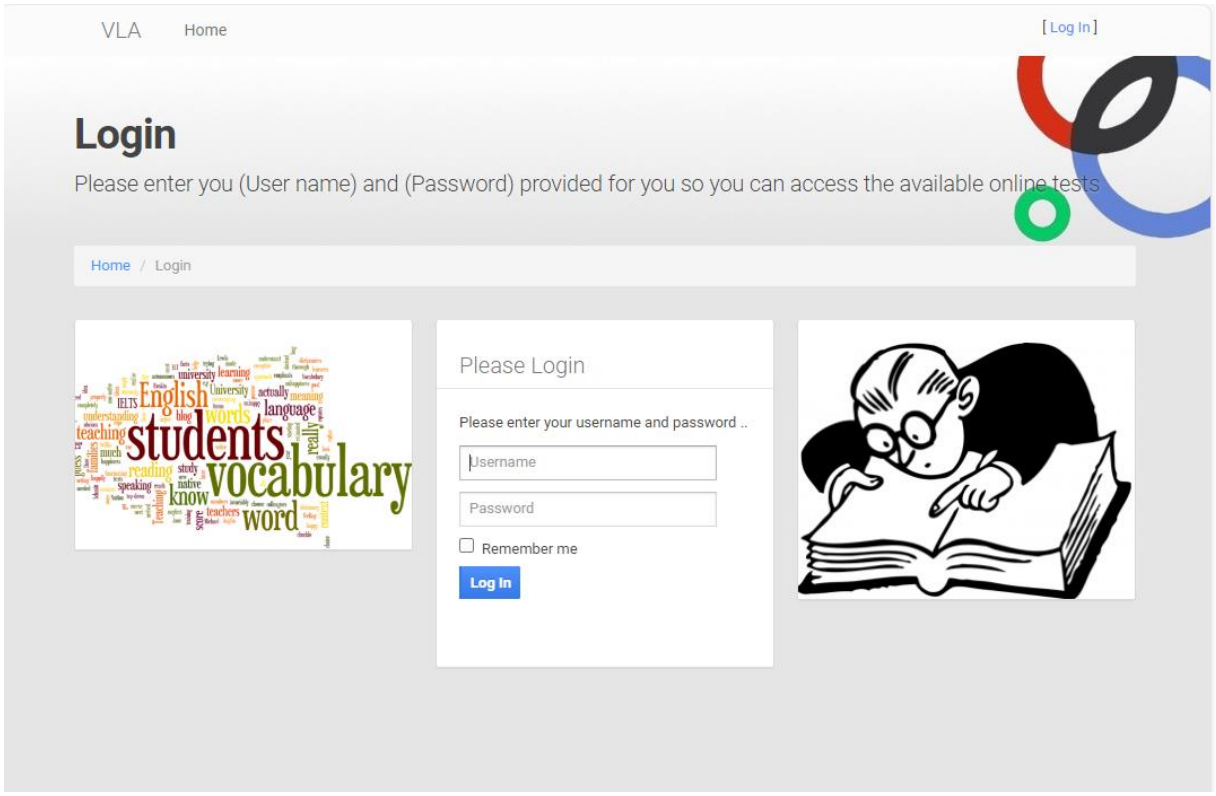
224. ZUMBACH, J. & MOHRAZ, M. 2008. Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior,* 24**,** 875-887.

Appendices

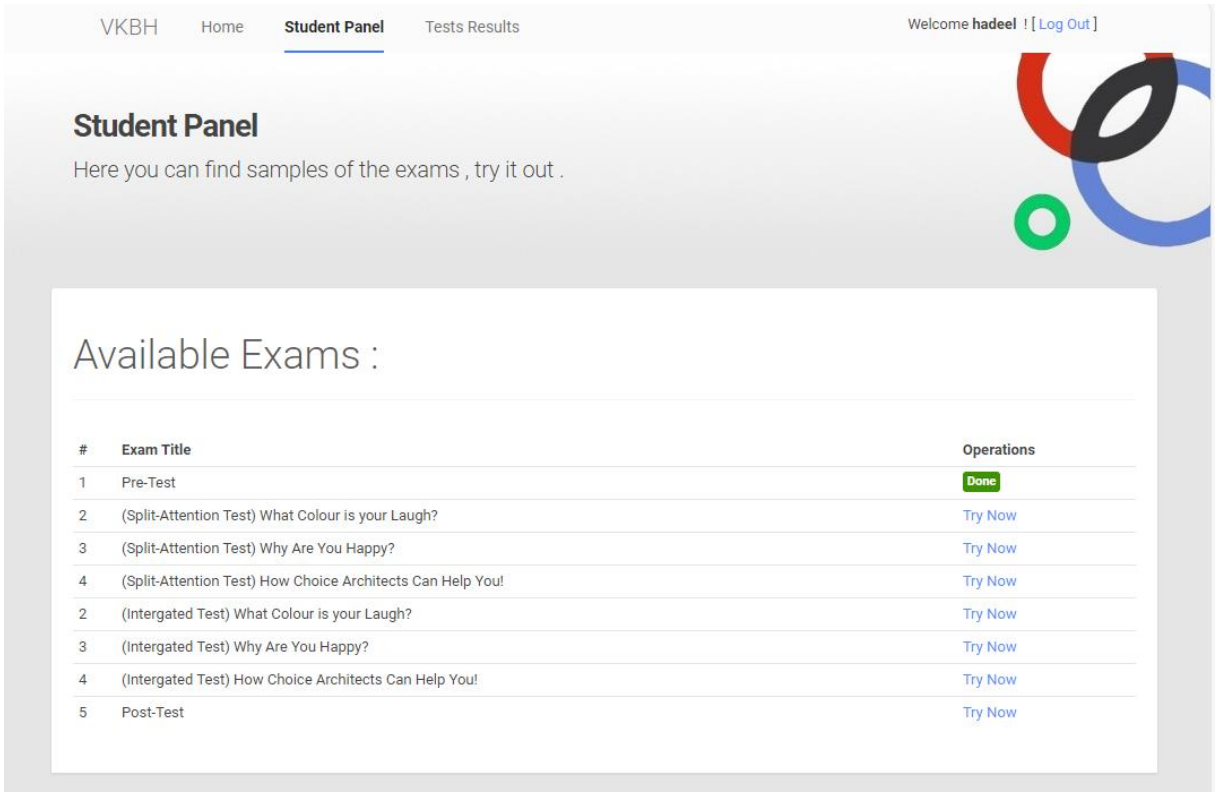
Appendix A : Website



Homepage



Login Page



Student Panel

Appendix B : Pre- and Post-Tests

1. Something occurring because of a natural tendency to behave in a particular way or a natural ability to know something that is not learned is something occurring _____

a. instinctively

b. happily

c. obscurely

2. **Obscure** is **NOT** the synonym of _____

a. unclear

b. vague

c. precise

d. difficult to understand

3. The study of the nature and meaning of existence, truth, good, and evil is _____

a. science

b. philosophy

c. high-tech

4. The speed at which something happens or is done is _____

a. thrill

b. pace

c. happiness

5. **thrill** is **NOT** the synonym of _____

a. boredom

b. excitement

c. pleasure

d. adventure

6. A place where you can buy and eat a meal is a _____

a. grocery
b. work
c. restaurant

7. Something a person does, especially something that is very good or bad is _____
a. joy
b. relationships
c. deed

8. Choose the word which best matches the following word to form a collocation.

_____ steak

a. rapid
b. rare
c. leisurely

9. **Discipline** is **NOT** the synonym of _____
a. control
b. regulation
c. restraint
d. indulgence

10. The new clothes, cars, and other things we _____ for don't necessarily make us happier.
a. longing
b. length
c. long

11. Choose the word which best matches the following word to form a collocation.
inherit _____

a. pace
b. property

c. monk

12. A member of an all-male religious group that lives apart from other people is a _____

a. philosopher

b. scientist

c. monk

13. Many parents are very _____ to their children.

a. devoted

b. devote

c. devotion

14. An event, situation, etc., that people do not understand or cannot explain because they do not know enough about it is _____

a. depression

b. mystery

c. understanding

15. Cheerful is **NOT** the synonym of _____

a. serious

b. happy

c. positive

d. joyful

16. Treatment that helps someone feel better or grow stronger is _____

a. shopping

b. therapy

c. benefit

17. Astonishing is **NOT** the synonym of _____

a. amazing

b. predictable

c. surprising

d. shocking

18. The government's **_____** that the recession was over did little to reassure the small businesses that were still struggling.

a. assertive

b. asserted

c. assertion

19. The way in which something is done or happens is _____

a. manner

b. review

c. esteem

20. A feeling of respect for someone, or a good opinion of someone is _____

a. manner

b. esteem

c. review

21. Review is **NOT** the synonym of _____

a. evaluation

b. assessment

c. description

d. proposal

22. To add something very different to something, so that it becomes part of it is to _____

a. boost

b. graft

c. suggest

23. **Boost** is **NOT** the synonym of _____

a. increase

b. enhance

c. improve

d. limit

24. Choose the word which best matches the following word to form a collocation.
_____ intervention

a. rapid

b. rare

c. leisurely

25. Cinnamon Orange Instant Tea is a _____ blend of black tea and oils based on an original recipe.

a. flavour

b. flavoured

c. flavouring

26. A _____ is a garden planted with flowers.

a. rose garden

b. bouquet garden

c. orange garden

27. _____ is a liquid for making permanent marks.

a. Indelible black

b. Indelible ink

c. Indelible flavour

28. The student's _____ comment showed that she had real insight into the subject.

a. perceive

b. perceptible

c. perception

29. The _____ from the pipe into the river carried poisonous waste.

a. seep

b. seepage

c. seepy

30. The indie band has launched a new _____-sounding album. It's unlike anything they have released before.

a. orchestra

b. orchestral

c. orchestrate

31. A _____ is a shaky piece of furniture.

a. wobbly table

b. simultaneous table

c. orchestral table

32. The _____ is the immediate conversion from one language to another.

a. wobbly translation

b. simultaneous translation

c. orchestral translation

33. Canada suffered from an _____ cold and wet weather last year.

a. abnormally
b. abnormality
c. abnormal

34. Image editing software allows photographers to _____ the quality of their photos.
a. refined
b. refine
c. refinement

35. A _____ is an insect bite.
a. bee sting
b. buzz sting
c. purple sting

36. My New Year's resolution is to leave more room in my life for _____.
a. spontaneously
b. spontaneity
c. spontaneous

37. A _____ happens when two people don't like each other.
a. spontaneous clash
b. personality clash
c. bubbly clash

38. Many believe the low participation in the election was due to voter _____ with politics.
a. disillusion
b. disillusioned
c. disillusive

39. Many successful athletes state that they were _____ by watching the Olympics when they were young
a. inspiriation

b. inspiring

c. inspired

40. A _____ is an important member of the community.

a. prominent citizen

b. musical citizen

c. abstract citizen

41. _____ of a person or event is realistic representation of them.

a. Accurate inspiration

b. Accurate perception

c. Accurate portrayal

42. The airline has announced that it plans to increase the _____ of its services to Turkey during peak summer months.

a. frequency

b. frequent

c. frequently

43. Much of the last century was defined by the _____ between communism and capitalism

a. clash

b. clashed

c. clashing

44. A _____ is a list of food available for the evening meal

a. dinner menu

b. dinner choice

c. dinner architect

45. To _____ is to make something hurt less

a. help pain

b. relieve pain

c. want pain

46. _____ are luxury items commonly named after a designer

a. Designer labels

b. Designer blogs

c. Designer menus

47. The _____ plan had the support of all the members of the committee

a. adopt

b. adoptive

c. adopted

48. _____ are things that people buy for their own use

a. Consumer goods

b. Choice goods

c. Layout goods

49. She spoke to the dog with a _____ voice in order to calm him down

a. gentleness

b. gentle

c. gently

50. The _____ of taxes by England was the main cause of the American War for Independence

a. impose

b. imposition

c. imposed

51. A _____ is an indirect suggestion that is not very easy to understand

a. subtle government

b. subtle hint

c. subtle lifestyle

52. A _____ is the main person responsible for a plan or idea

a. comfortable architect
b. plan architect
c. chief architect

53. A _____ is a life in which you have enough money to buy the things you want

a. plan lifestyle
b. chief lifestyle
c. comfortable lifestyle

54. The _____ of smallpox as an "eradicated disease" occurred in 1979; it is the only human disease to be completely eliminated.

a. classify
b. classification
c. classified

55. The spy _____ around the corner to see if anyone was following him.

a. glanced
b. glancing
c. glance

56. The athlete ate very carefully, with an occasional chocolate bar being his only _____.

a. indulgence
b. indulgent
c. indulgently

57. To _____ is to quickly get something to eat.

a. classify a bite
b. indulge a bite
c. grab a bite

58. The _____ with the bear in the forest left him shaken and scared

a. encountering
b. encounter
c. encountered

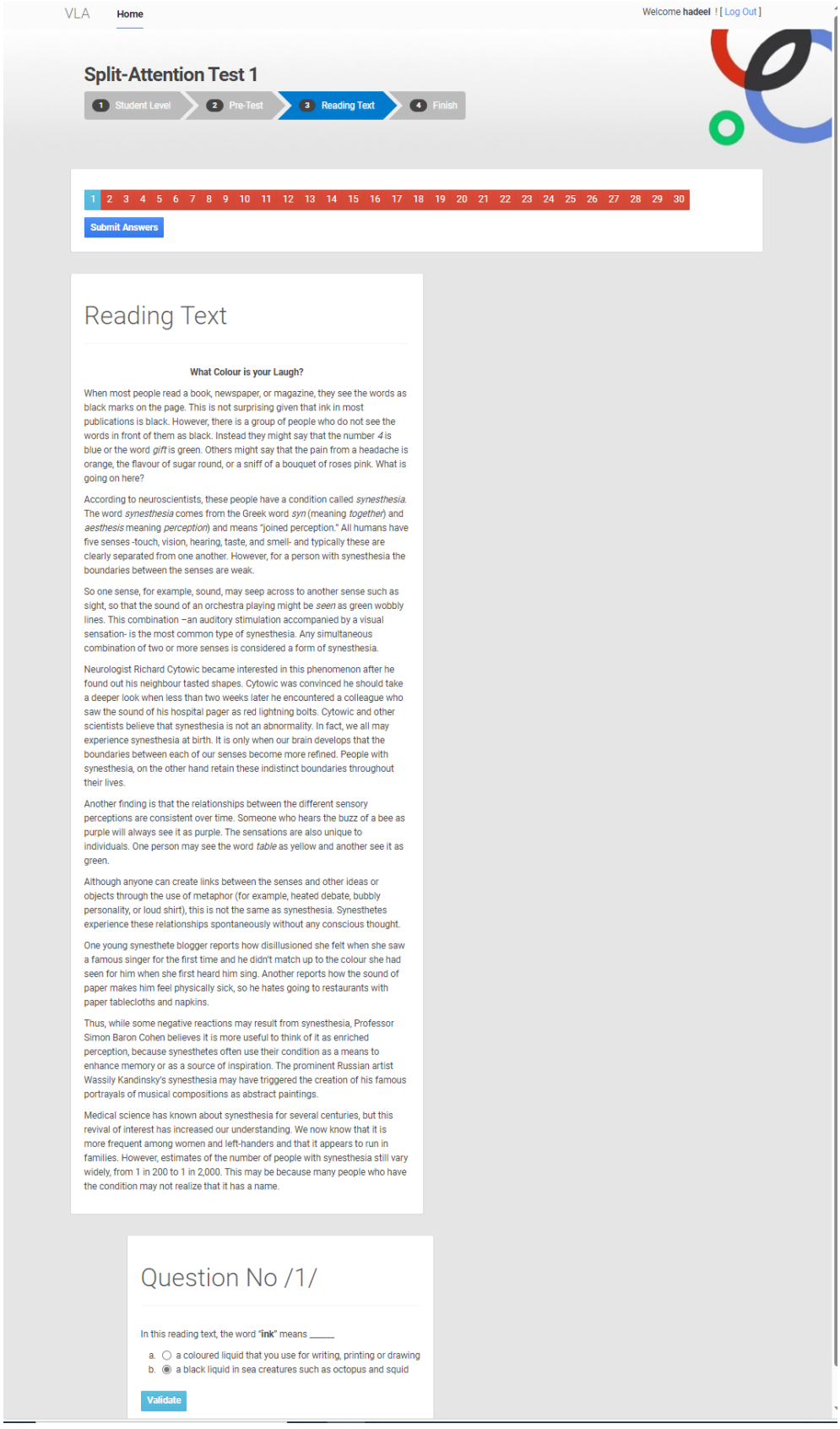
59. After moving to Australia from America, she was only able to make _____ visits home to see her parents.

a. infrequency
b. infrequent
c. infrequently

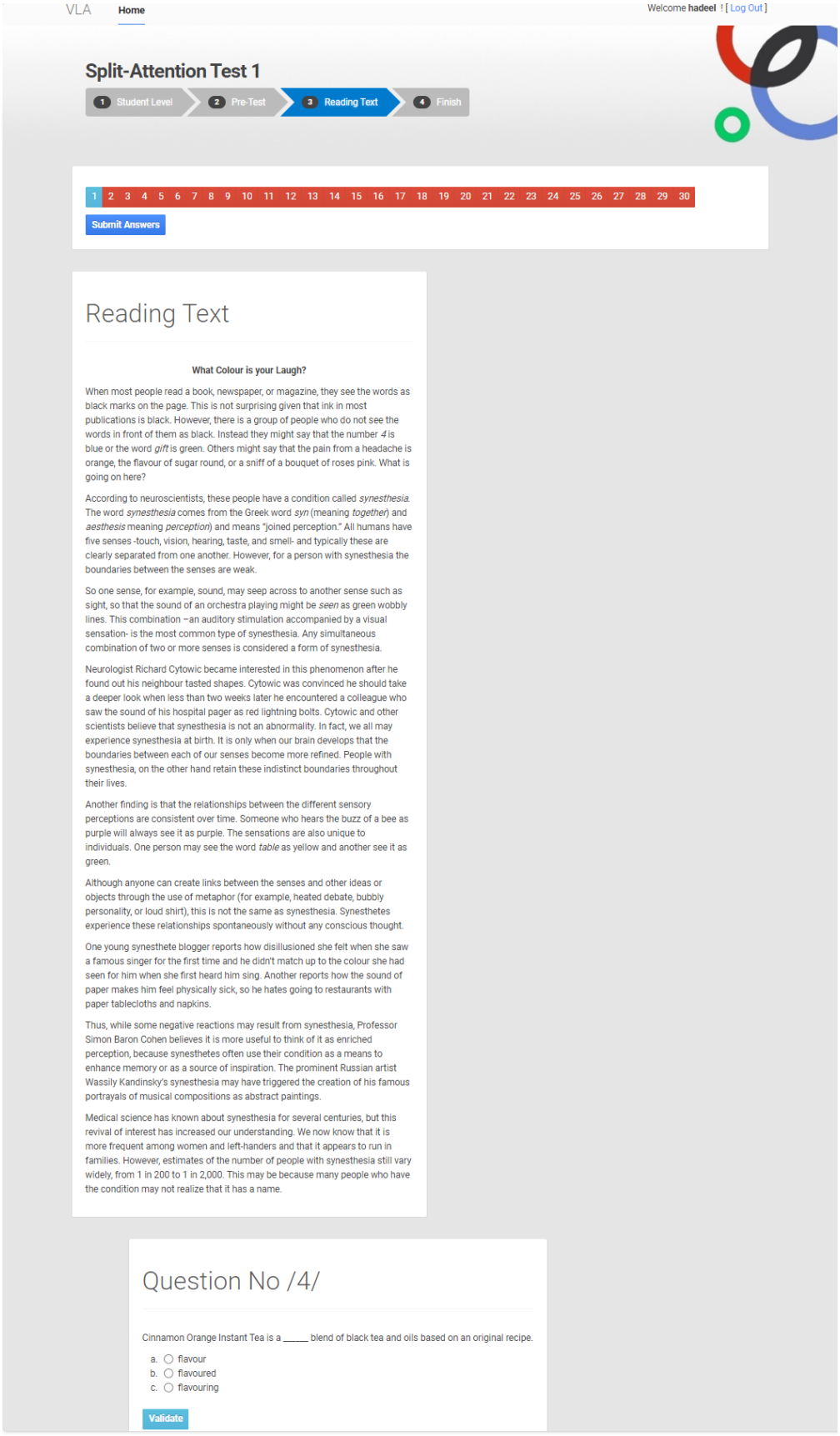
60. The _____ of blood is one of the most valuable things anybody can do.

a. donate
b. donation
c. donated

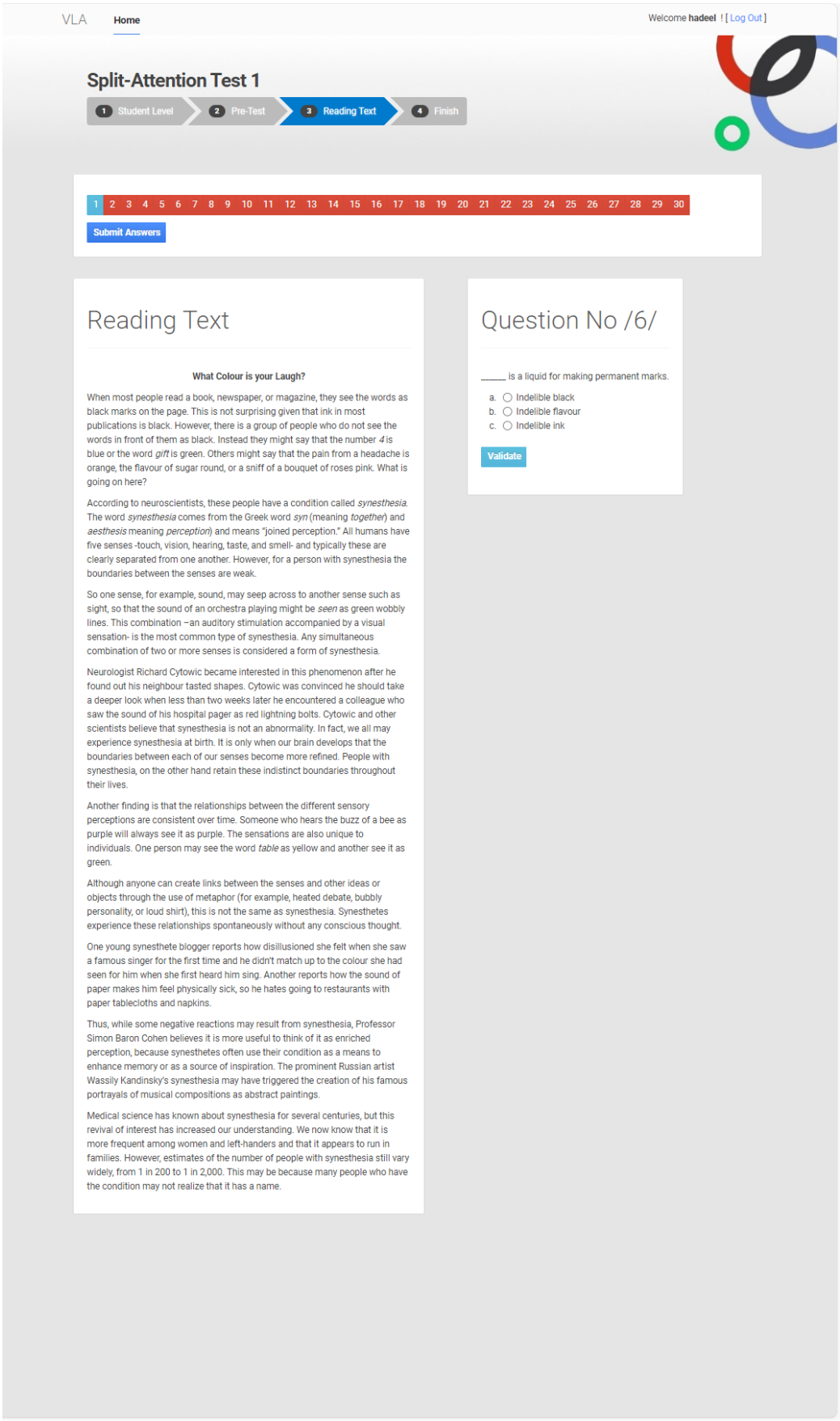
Appendix C : Experiment



A question about word meaning from the first reading text in the split-attention format.



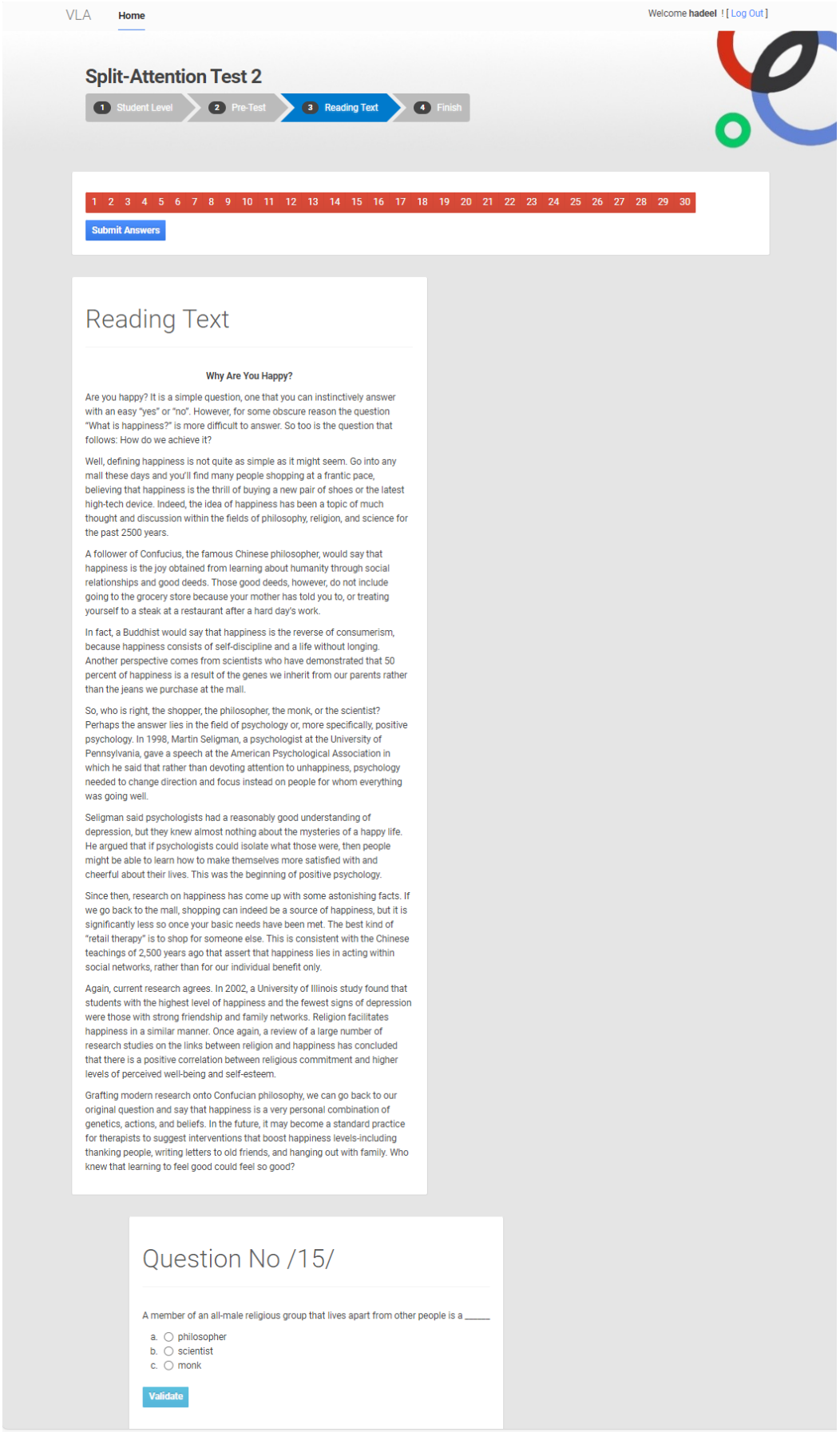
A question about word form from the first reading text in the split-attention format.



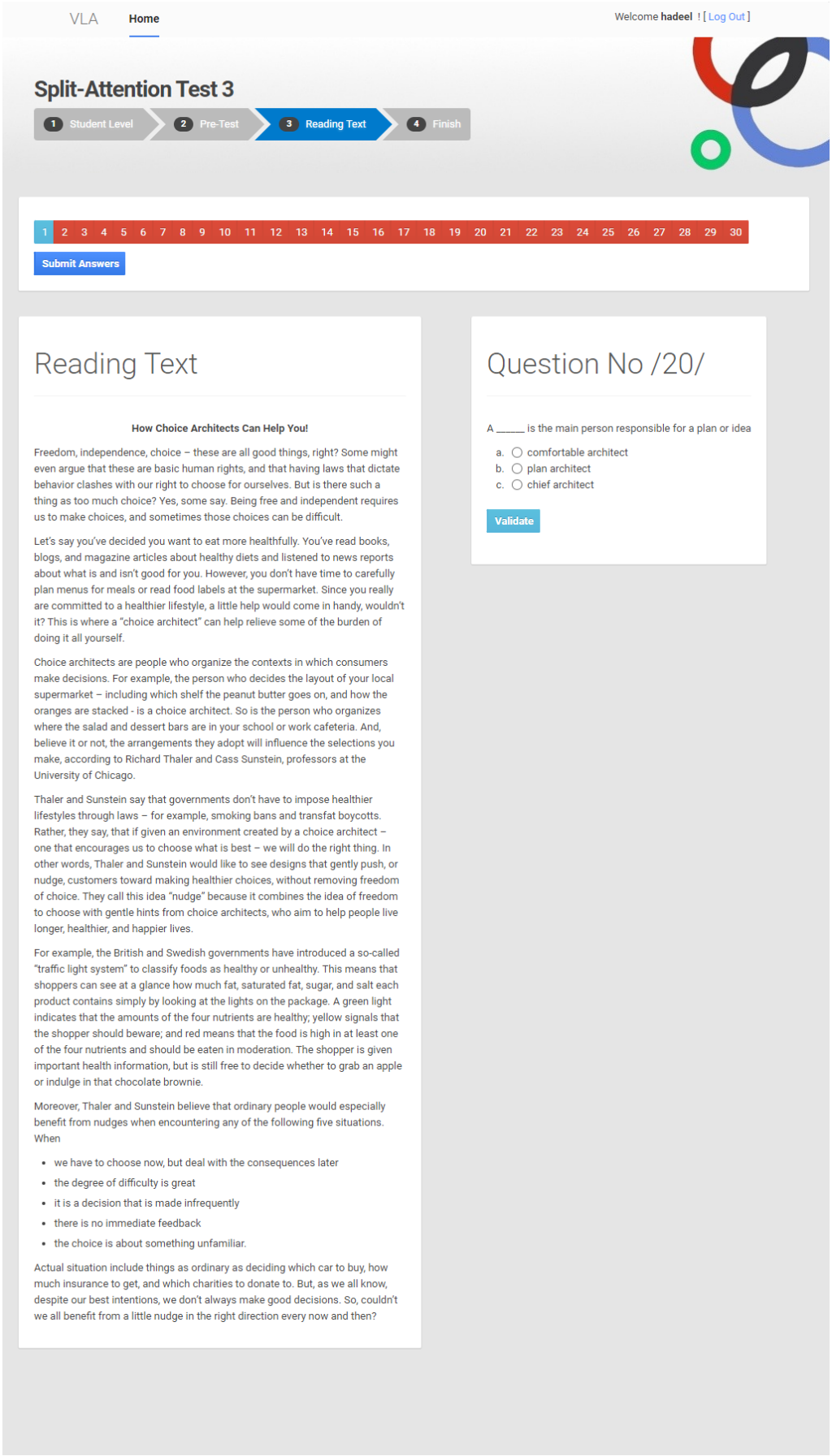
A question about collocations from the first reading text in the split-attention format.

A question about collocations from the second reading text in the split-attention format.

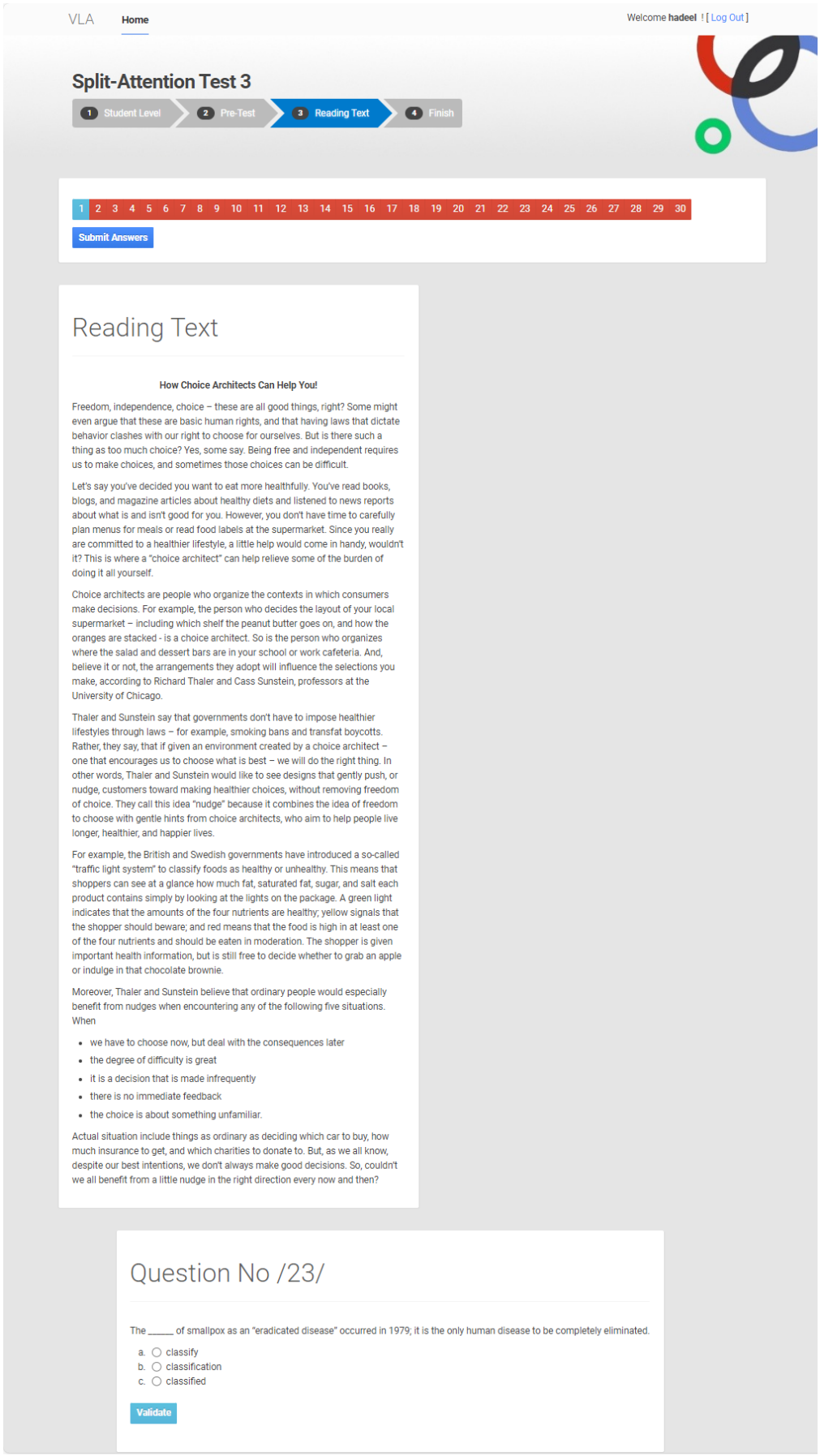
A question about word form from the second reading text in the split-attention format.



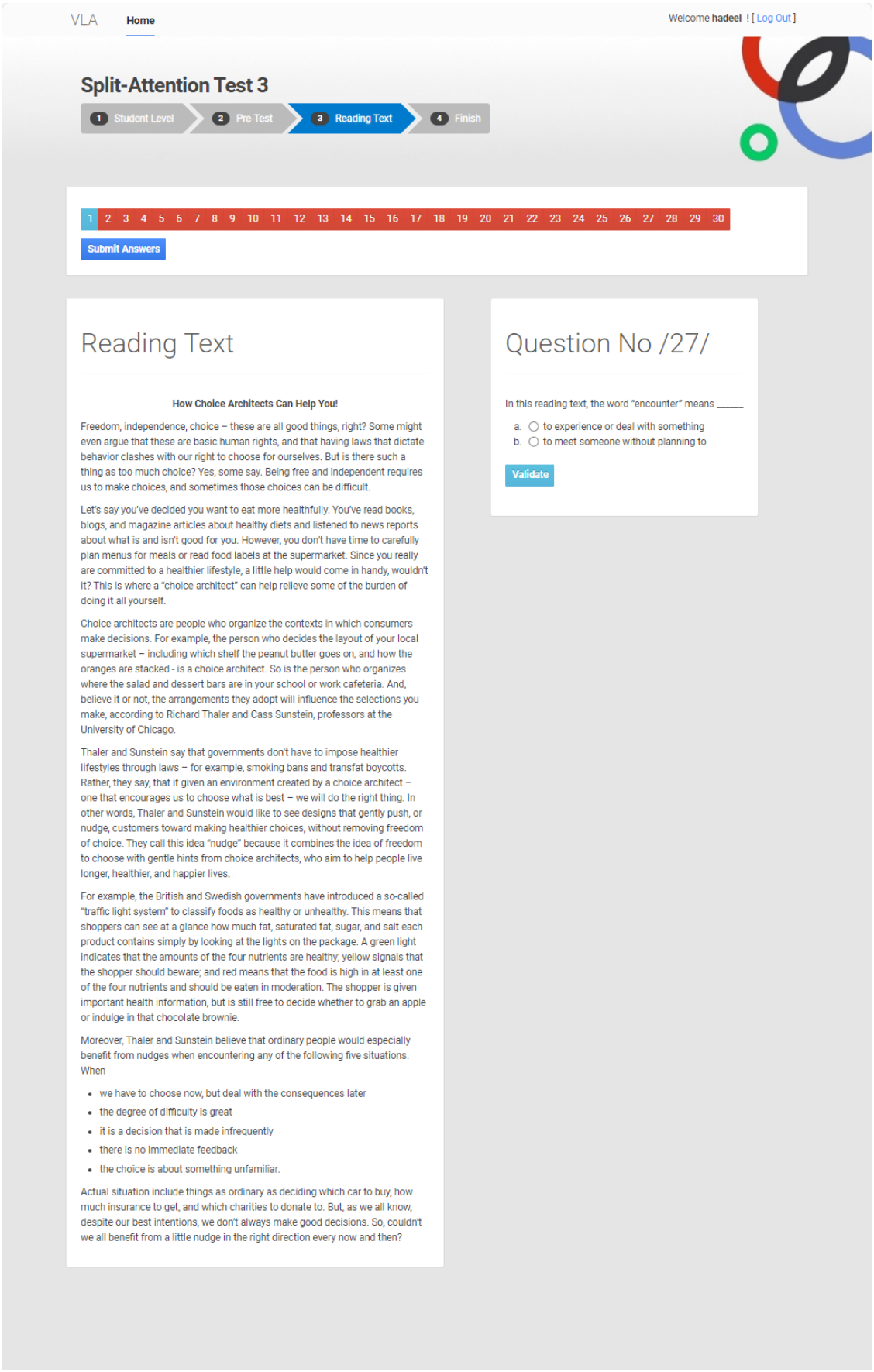
A question about word meaning from the second reading text in the split-attention format.



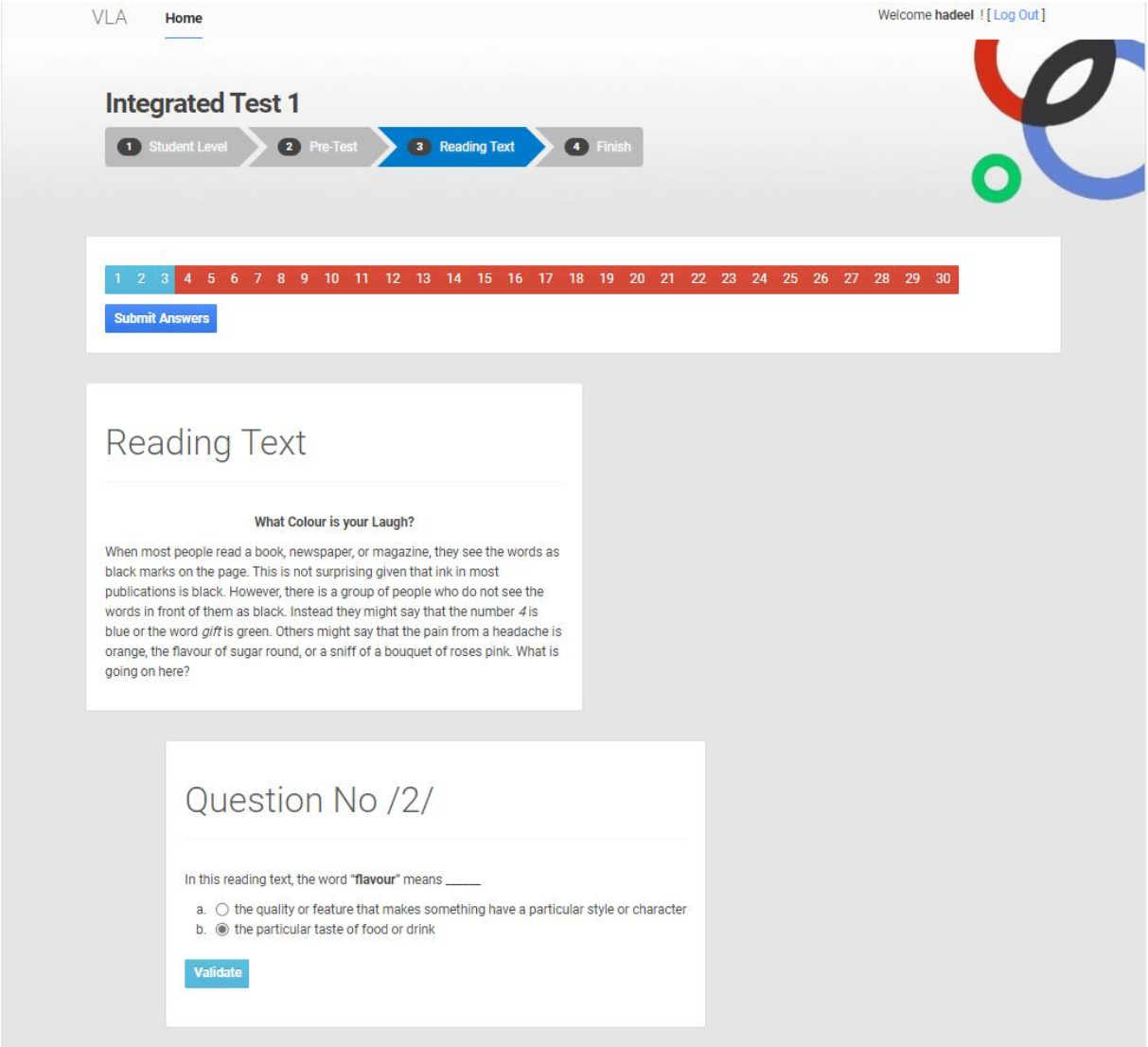
A question about collocations from the third reading text in the split-attention format.



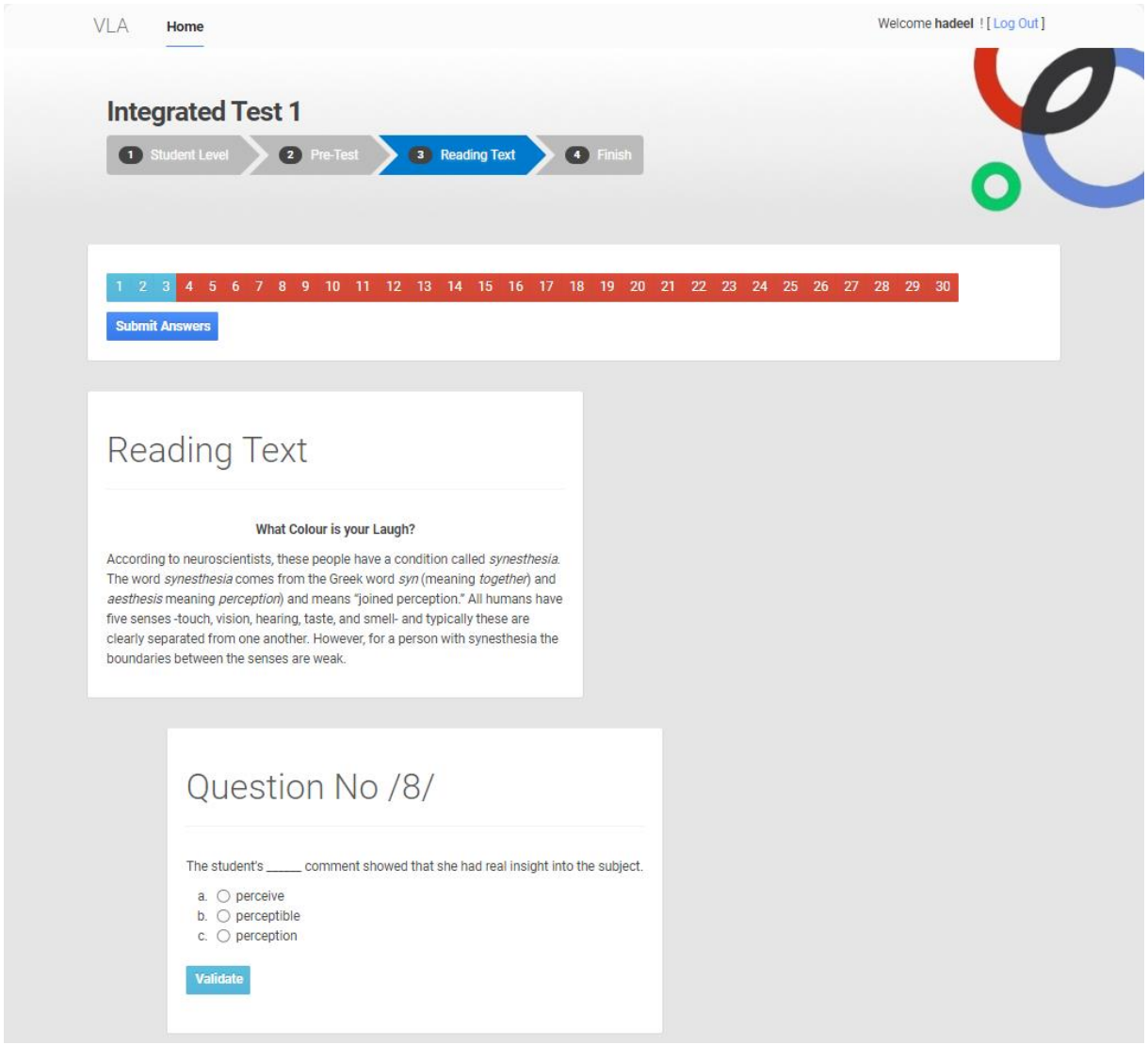
A question about word form from the third reading text in the split-attention format.



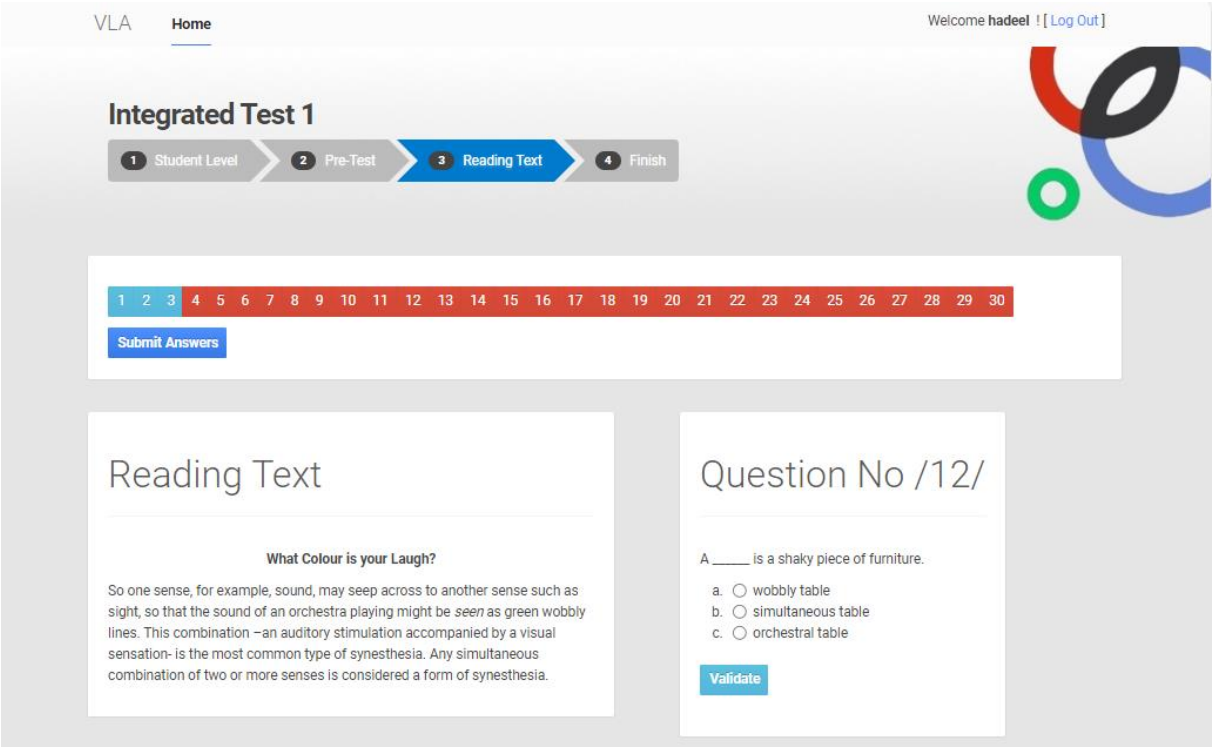
A question about word meaning from the third reading text in the split-attention format.



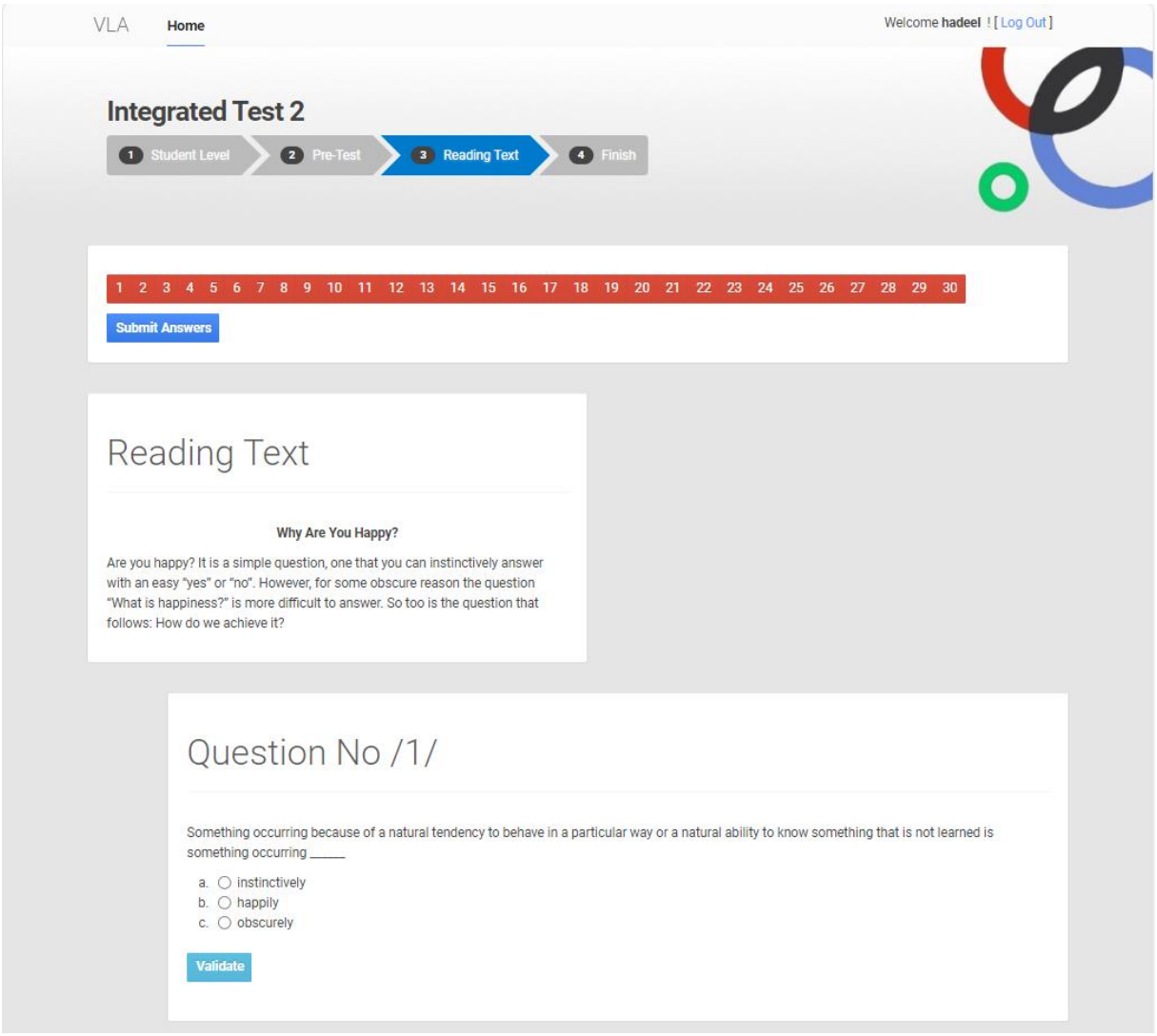
A question about word meaning from the first reading text in the integrated format.



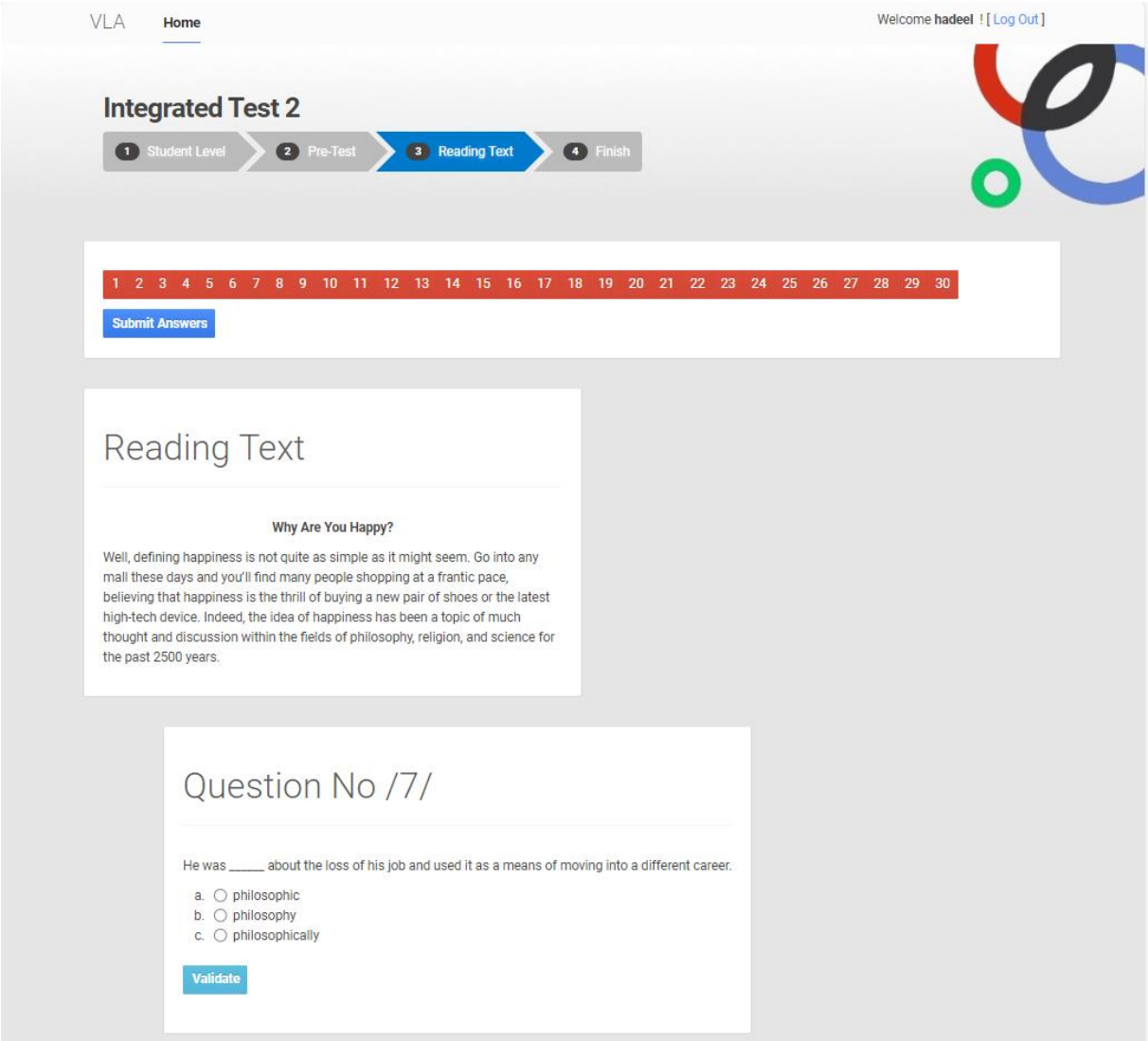
A question about word form from the first reading text in the integrated format.



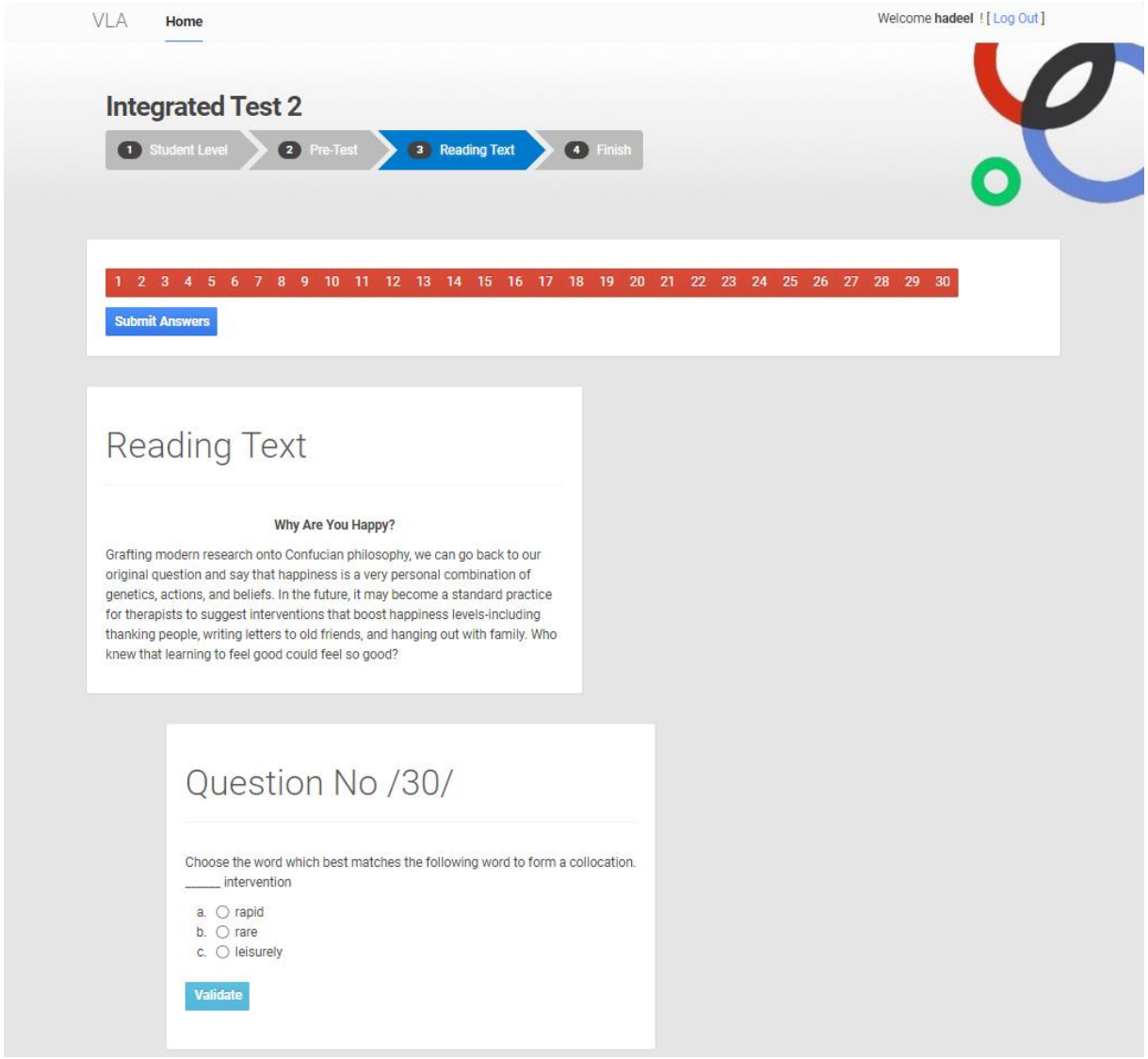
A question about collocations from the first reading text in the integrated format.



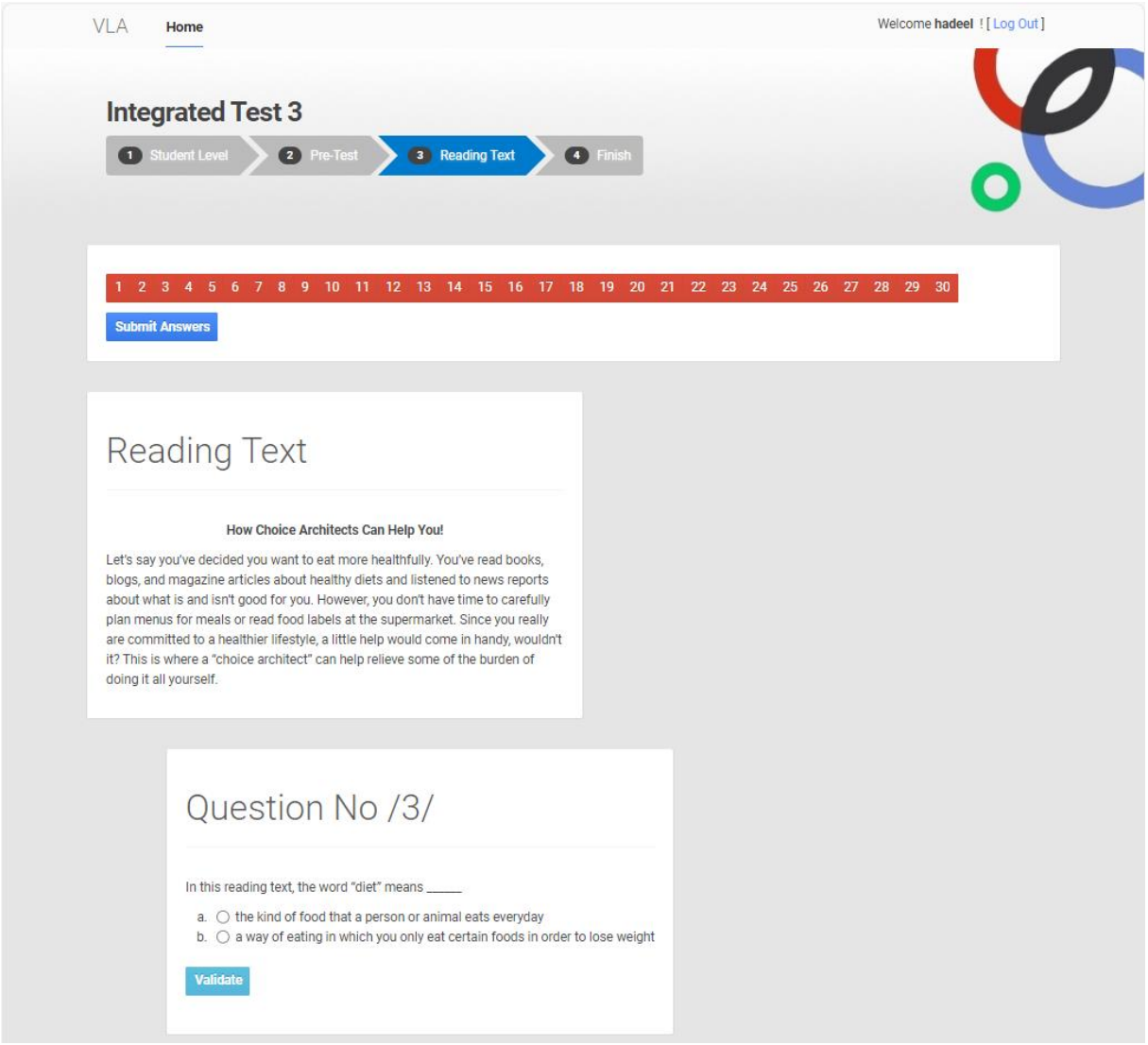
A question about word meaning from the second reading text in the integrated format.



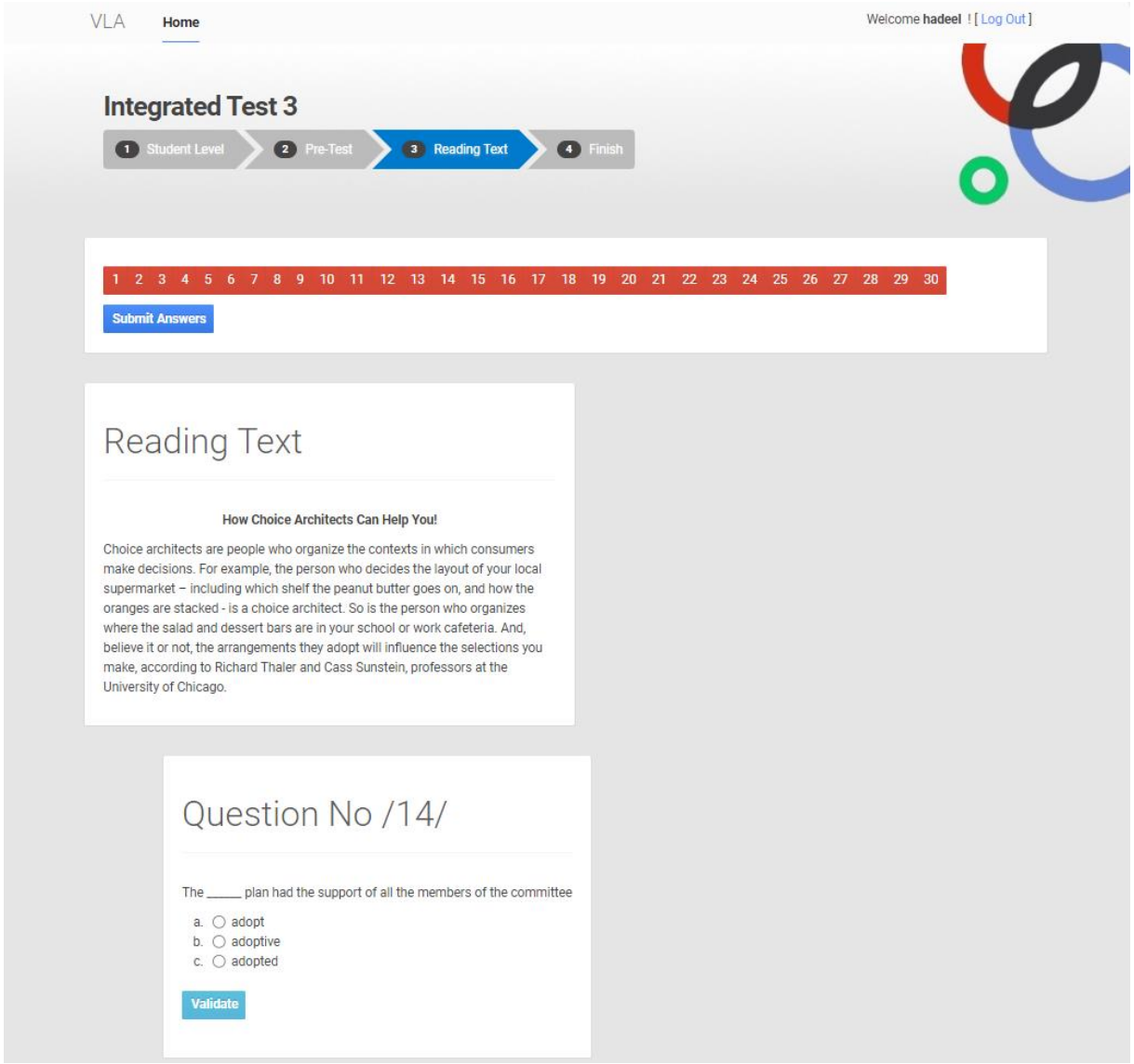
A question about word form from the second reading text in the integrated format.



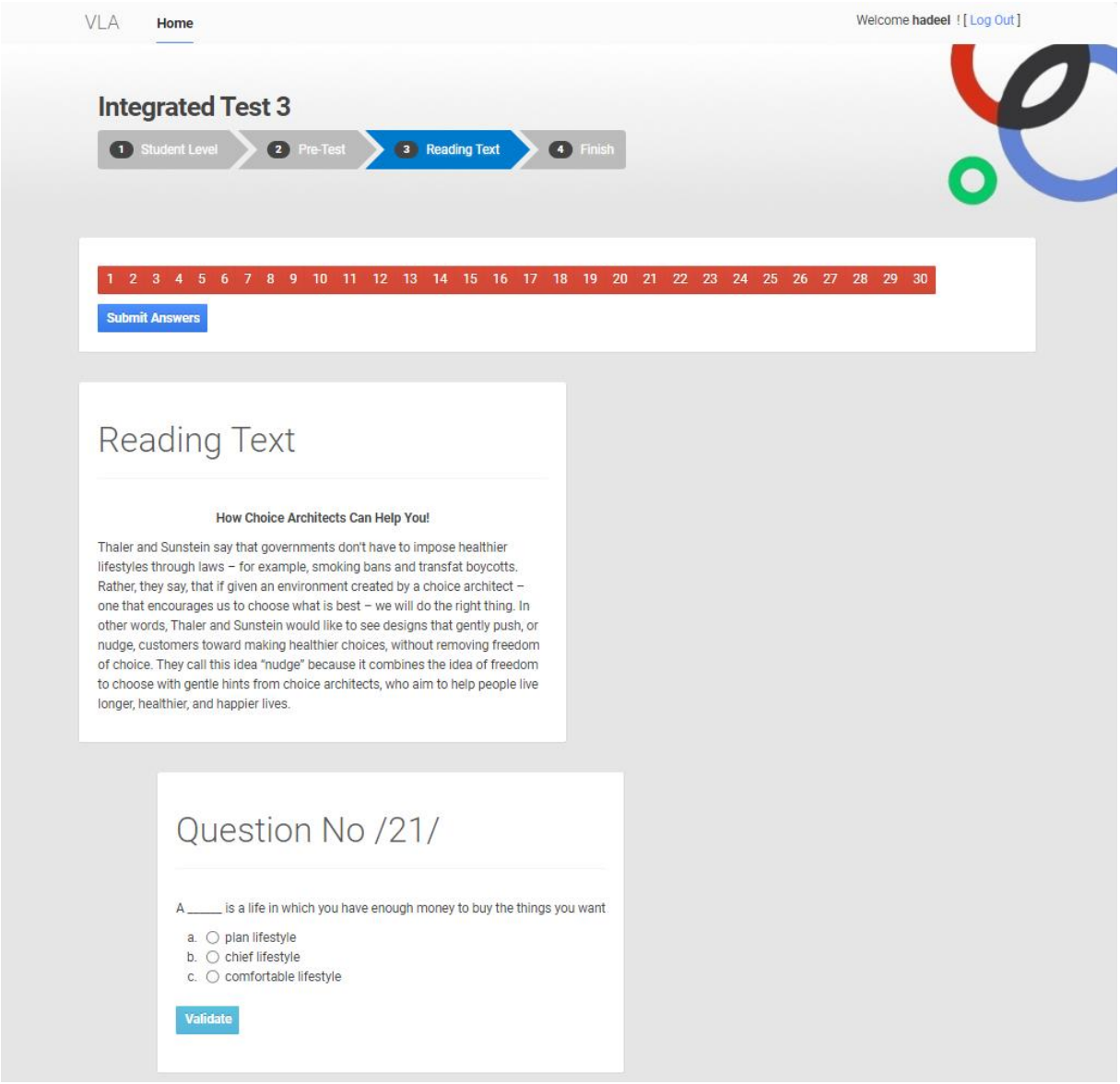
A question about collocations from the second reading text in the integrated format.



A question about word meaning from the third reading text in the integrated format.



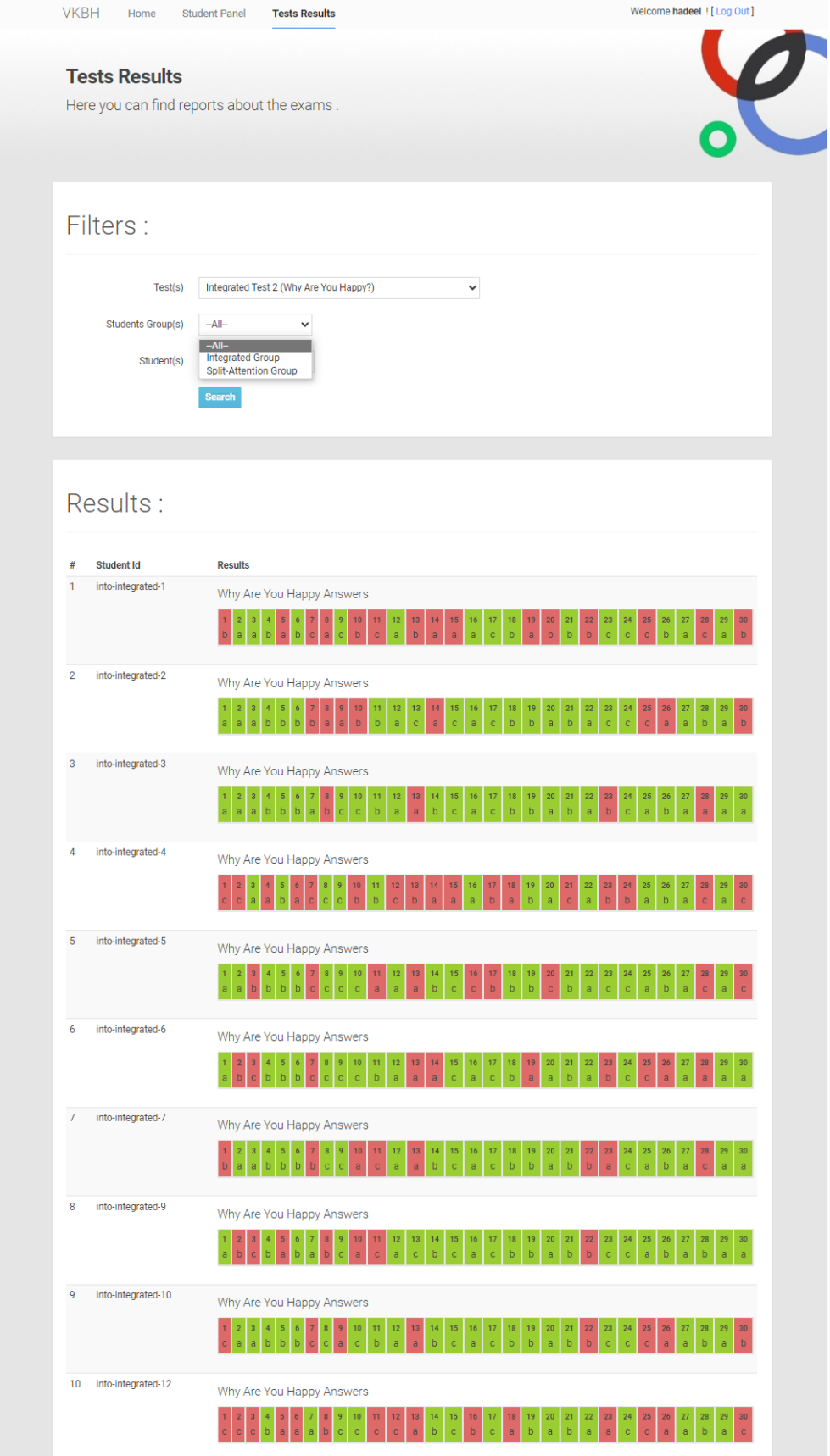
A question about word form from the third reading text in the integrated format.



A question about collocations from the third reading text in the integrated format.

Appendix D : Test Results

First example of some test results showing the display and some filtering options in the administrator panel



Second example of some test results showing the display and some filtering options in the administrator panel

Third example of some test results showing the display and some filtering options in the administrator panel