EFFECTS OF MULTIMEDIA MODALITY AND L2 WORKING MEMORY CAPACITY ON L2 COMPREHENSION

KADİR KOZAN

BOĞAZİÇİ UNIVERSITY

2009

EFFECTS OF MULTIMEDIA MODALITY AND L2 WORKING MEMORY

CAPACITY ON L2 COMPREHENSION

Thesis submitted to the

Institute for Graduate Studies in the Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

English Language Education

by

Kadir Kozan

Boğaziçi University

2009

Thesis Abstract

Kadir Kozan, "Effects of Multimedia Modality and L2 Working Memory Capacity on L2 Comprehension"

This investigation aims at exploring immediate and delayed effects of extraneous cognitive load caused by the presentation mode of an English text and of English verbal working memory capacity of advanced learners of English on second/foreign language (henceforth L2) comprehension under the conditions of high intrinsic cognitive load and low prior knowledge.

English text was presented in two different presentation modes on a website on the computer environment: 1) text with pictures; and 2) narration with pictures. 29 advanced ELT students were randomly assigned to the experimental groups and were asked to read or listen to the text with the intention of comprehension. Right after and three weeks after the treatment, the participants were given retention and transfer comprehension tests. They were also given a subjective cognitive load rating scale to indicate the level of cognitive load they thought the multimedia environment involved.

Results indicated that it was the combined effect of time, extraneous cognitive load and L2 verbal working memory capacity that had a significant effect on retention of information from the treatment text. In other words, the results of the study indicated that presenting verbal information aurally might affect retention of information over time depending on working memory capacity. As for transfer of knowledge, a significant main effect of time was found. This significant main effect of time means that immediate transfer scores of the participants were significantly higher than their delayed transfer scores on average regardless of extraneous cognitive load and working memory. Moreover, it was found that participants in the narration with pictures condition that is assumed to expose less extraneous cognitive load than the text with pictures condition reported significantly higher cognitive load ratings than their counterparts in the text with pictures condition.

Results are interpreted in the light of cognitive theory of multimedia learning (Mayer, 2001) and cognitive load theory (Sweller, 1988). The results provided additive information on the modality principle of cognitive theory of multimedia learning under the conditions of high intrinsic cognitive load and low prior knowledge in an L2 multimedia learning environment. Finally, it is claimed that the assumptions of modality principle may change for L2 learners, which depends on time of testing, working memory capacity, and type of comprehension.

Tez Özeti

Kadir Kozan, "Multimedya Modalite Efektinin ve İkinci Dil İşlek Bellek Kapasitesinin İkinci Dilde Okuduğunu ve Dinlediğini Anlama Başarısına Etkileri"

Bu çalışma sunu modunun yol açtığı harici bilişsel yükün ve İngilizce sözel işlek bellek kapasitesinin öğrenici tarafından yönetilen çoklu bir öğrenme ortamında sunulan, yoğun içsel-bilişsel yük taşıyan bir öğrenme materyali hakkında düşük düzeyde bilgi sahibi olan ileri düzey İngilizce bilenlerin ikinci dilde, yani İngilizce'de, okuduğunu ve dinlediğini anlama başarısı üzerine etkilerini araştırmayı amaçlamaktadır.

Öğrenme materyali katılımcılara bilgisayar ortamında web sitesi olarak hazırlanmış iki farklı sunu modunda sunulmuştur: 1) okuma ve resimler; 2) dinleme ve resimler. 29 ileri düzey İngilizce öğretmenliği öğrencisi 4 deney grubuna rasgele örneklem yoluyla atanmış ve kendilerinden okuma ve dinleme parçasını anlama amacıyla okumaları ya da dinlemeleri istenmiştir. Uygulamadan hemen sonra katılımcılara okuduğunu veya dinlediğini akılda tutma ve okuduğunu veya dinlediğini aktarma testleri uygulanmıştır. Bu testlerle birlikte katılımcılara ayrıca parçayı okudukları ya da dinledikleri çoklu öğrenme ortamının ne kadar bilişsel yük taşıdığını puanlayacakları bir kişisel bilişsel yük ölçeği verilmiştir. Aynı testler, testlerin ilk uygulamasından iki hafta sonra deneklere tekrar verilmiştir.

Çoklu bir öğrenme ortamında ikinci dilde okunan ya da dinlenen metindeki bilgi içeriğini akılda tutmanın zaman, İngilizce sözel işlek bellek kapasitesi ve harici bilişsel yükün ortak etkisine bağlı olduğu bulunmuştur. Başka bir deyişle, çalışmanın sonuçları sözel bilginin dinleme modunda sunulmasının işlek bellek kapasitesine bağlı olarak bilgiyi akılda tutma başarısını uzun vadede etkileyebileceğini göstermiştir. Bilginin aktarımına gelince, katılımcıların ön testte geç teste oranla istatistiksel olarak daha iyi performans gösterdiği görülmüştür. Yani çoklu bir öğrenme ortamında ikinci dilde okunan ya da dinlenen bir metinden bilgi aktarımının zamana bağlı olduğu ve zamanla aktarım performansının düştüğü bulunmuştur. Ayrıca teorik olarak okuma ve resimler sunumuna göre daha az harici bilişsel yük yarattığı öngörülen dinleme ve resimler durumundaki denekler kişisel bilişsel yük ölçeğinde diğer gruptakilere oranla istatistiksel olarak daha yüksek bilişsel yük rapor etmişlerdir.

Sonuçlar Mayer'in (2001) multimedya öğrenme teorisi ve Sweller'in (1988) bilişsel yük teorisi altında incelenmiştir. Bu çalışmanın sonuçları Mayer'in (2001) multimedya öğrenme teorisi'nin bilginin işitsel ve görsel (dinleme + resimler) olarak sunulmasının sadece görsel (okuma + resimler) olarak sunulmasından daha iyi öğrenmeyi sağlayacağını ileri süren modalite prensibine tam olarak uygun olmasa da bu prensiple uyuşmamak yerine onu tamamlar niteliktedir. Sonuç olarak modalite prensibinin öngörülerinin ikinci dil öğrenicilerinde testin zamanına, işlek bellek kapasitesine ve değişik anlama başarısı türlerine göre farklılaşabileceği ileri sürülmüştür.

ACKNOWLEDGEMENTS

It is a great pleasure for me to acknowledge the intricate contributions of those without whose help this thesis would not be completed. My wholehearted thanks go to my thesis committee, Assist. Prof. Gülcan Erçetin, Assist. Prof. Senem Yıldız, and Assist. Prof. Günizi Kartal. I am specifically grateful to my thesis advisor Assist. Prof. Gülcan Erçetin for her conscientious and meticulous guidance.

While some thanks go to the people who have made my light go out thus triggering my life-long persistence, I owe deepest thanks to those who have rekindled this light. Among these are Nevin Kahriman, Çiğdem Çilek Anıl, Ayşe Banu Demir, Mustafa Tezcan, Volkan Özekinci, Dilek Çankaya, Ayşe Eroğlu, Pınar Kurt, Seda Kan, and Kadir Gökgöz. Nevin deserves heartfelt thanks for her contributions in the piloting of the study and for her spiritual back up. My home-mate Volkan also deserves extra thanks for his help in the piloting and marking the tests of the study. Dilek's unforgettable efforts in helping me find more participants make her contributions invaluable.

As for my beloved mother, Suzan Kozan, and my beloved aunt, Şerife Gökçen, I cannot deny their contributions to the work I have done to complete this study. I am totally indebted to them to the full extent for everything I have succeeded in the name of both this study and life itself in general. I thank my father, Salim Kozan, for his financial support as well.

My special thanks also go to the participants of this study who made the research part possible and without whom it was impossible to complete the overall thesis. Finally, it is a real proud of me to state the thanks that go to my people, a lot of heroes and heroines, who presented the modern Turkey to us by sacrificing their lives before, in, and after the Turkish War of Independence.

DEDICATION

The present work is wholeheartedly dedicated to my mother, Suzan Kozan, and my aunt Şerife Gökçen, who have continuously been encouraging me to trust life itself and myself.

CONTENTS

FIGURES

TABLES

PREFACE

Sometimes, people cannot escape learning: when we watch a movie, when we read a book, or even when we surf the internet just for fun. However, especially for educational purposes, we aim at making the process of learning an effective and efficient experience, which is also the case in L2 education. This is a challenging task for the field, since it is not only related to instructional aspects but also to learner and task characteristics. Even though efficiency and effectiveness of learning are not only bound by these, they are at the heart of meaningful learning requiring a learner, to-be-learned information, and the presentation of the information interact with one another. Therefore, one of the fundamental questions is how to present information in instructional materials in order to optimize learning. One way of enhancing learning is using technology: With the rapid advent of technology, we have a lot of different sorts of presentation types ranging from texts to animations, which might help us promote learning. However, the wide range of technological advancements might also bombard the learner with information presented in different sorts of multimedia presentations.

Technological developments and their byproducts, multimedia presentations, have been playing a great role in language learning as well, which provides the learners not only with the chance of gaining language information but also with the chance of interacting with that information. Moreover, in a world where communication among people has become essential, learning an L2 carries great importance. Needless to say, in a millennium of knowledge, knowledge itself -be it language knowledge or any other kind of it-, and how to learn it in an effective way have earned importance, which triggered a scientific exploration of how to learn and teach a mass of information by means of a mass of different multimedia.

Educational science, unfortunately, has not yet managed to provide us with a clear-cut and coherent body of guidelines for designing multimedia instructions that would promote learning. However, two recent lines of research informed by cognitive load theory (Sweller, 1988) and cognitive/generative theory of multimedia learning (Mayer, 2001) have provided us with several instructional guidelines based on the limits of human cognitive architecture. Both theories pinpoint that our working memory capacity, which governs selection, organization, and integration of to-be-learned information, is severely limited and that instructional design should bypass the limits of working memory. This partly stems from one of the basic concerns of both theories: working memory is implemented both in learning new information and integrating this new information into already-known prior knowledge in a given domain. If the capacity-limited working memory resources cannot be used in an efficient way, cognitive load or working memory load that occupies existing limited cognitive resources show up, which deteriorates learning.

The guidelines yielded by both cognitive load theory (Sweller, 1988) and cognitive multimedia learning theory (Mayer, 2001) aim at a more efficient allocation of our cognitive resources for effective learning. One of the guidelines is that text corresponding to a picture or animation should be presented aurally not visually, since presenting corresponding pieces of information in two sensory modalities reduces any possible overload in visual cognitive resources, thus leading to a more optimal use of them. This is called the modality effect in cognitive load theory and modality principle in cognitive theory of multimedia learning. Application of the modality effect and the other guidelines has been tested by a wide range of experiments. The results indicated that application of guidelines created less cognitive load and better learning outcomes.

However, to our best of knowledge, most of the earlier research on modality principle has been conducted in first language (L1 henceforth) learning contexts and instructions used in previous studies included subject matters from such exact sciences as engineering or geometry. Instructions were also either system-paced or paper-based instructions both of which were limited to the total duration of the narration. Moreover, most of the research took place in strictly controlled laboratory settings and used short-term learning tests. Likewise, any possible effect of working memory capacity or its interaction with presentation mode was not addressed by most of the previous research agenda. All these raise the question whether the modality effect also applies with instructional materials from the field of L2 education, if the learners are given control over the pacing of instructions, and whether the effect can be generalized to more authentic educational settings and long-term learning outcomes. Specifically speaking, what would happen, for example, in a learner-paced multimedia L2 learning environment over time? Or, what is the role of verbal working memory capacity and its interaction with presentation mode in L2 comprehension? Therefore, the main purpose of the present thesis is to test whether the modality effect expands to learner-paced L2 learning contexts including L2 education in a more ecologically valid educational setting. The experiment took place in a computerized L2 learning laboratory, where university students learn Italian as an L2.

The results of the current research should be read cautiously, since it has some shortcomings. First, time-on-task was not limited in order not to destroy the authentic nature of the learning environment, which could have affected participants' comprehension performance. Second, sample size was small (N= 29). However, significant combined effects and main effects have been found, which suggests that these effects are very large effects. Cohen's effect size (*d)* and eta squared statistics confirmed this and indicated that the statistically significant effects found in the current research were practically significant as well.

After all, the current study seeks to test whether the modality effect can be replicated in the comprehension of an authentic, expository L2 text in an authentic educational setting by taking into account both English verbal working memory capacity and time of testing. Finally, it tests the effects of modality of presentation on both retention of information and transfer of information, thus gauging the applicability of the effect on different types of comprehension.

CHAPTER 1

INTRODUCTION

The effectiveness of multimedia on language learning and teaching has been of interest for researchers accommodating in different fields of inquiry ranging from L2 acquisition to cognitive psychology. There has been a tendency to apply findings from these fields to learning and especially to how to lead to better learning under what instructional circumstances.

Numerous numbers of multimedia software have been used to teach and learn an L2, because multimedia has been considered an effective way in the field of L2 education. It is not only effective in terms of learning outcomes because it facilitates language teaching and learning (Kramsch & Andersen, 1999), but also of motivating learners and expanding language teaching and learning beyond the walls of the language classroom. Besides, multimedia technology has the capacity to promote overall aim of language learning: communication. That is because of the fact that multimedia instruments provide learners with opportunities to develop their communicative competence in an L2 by means of such productivity tools as sound-recorders, writing equipments, and feedback providers.

As use of multimedia has become popular in the field of language teaching and learning, Sweller's (1988) cognitive load theory has also become popular for learning and teaching studies. Mayer (1997, 2001) applied the principles of the cognitive load theory to multimedia learning in different areas. Then, it was realized that not all multimedia instructional designs cause the same amount of cognitive load on the part of the learners. Thus, the question of how to design instructions to lead to better learning has taken a lot interest. For instance, Chen and Macredie (2002) question how to organize and present information through media to enhance effective learning, particularly in computer-assisted learning and hypermedia delivery modes. Grimly (2007) argues that "few theories consider individual differences effectively" (p. 465), and suggests that how individual differences affect learning in instructional environments designed according to Mayer's (2001) cognitive theory of multimedia learning and Sweller's (1988) cognitive load theory be investigated. He further pinpoints that individual learner characteristics are of great importance for cognitive load and multimedia effectiveness in a learning environment.

As generally defined, cognitive load is the amount of cognitive resources allocated for information processing. High cognitive load causes the expansion of more extra mental resources for dealing with upcoming information on the part of the learners. However, low cognitive load frees more cognitive resources that can be assigned to schema construction and activation. It is also known that working memory gets help from long term memory in the form of prior knowledge- a factor that decreases the overall cognitive load imposed upon learners. This means that prior knowledge lessens cognitive load, thus promoting better performance.

Empirical studies focusing on the educational implications of multimedia learning and cognitive load theory have been using divergent instructional materials from different subject areas ranging from history to mathematics (e.g., Mousavi, Low, & Sweller, 1995; Brünken, Steinbacher, Plass, & Leutner, 2002). Even though multimedia use is very popular in L2 learning, the issue of the amount of cognitive load multimedia learning creates for L2 learners have taken less interest.

Most of the studies on cognitive load theory and multimedia learning have been all conducted in L1 learning environments. In other words, studies up to now have been mostly concerned with cognitive load scenarios imposed by instructional learning materials in L1 English. Not surprisingly, learning environments in different areas have been subject to cognitive load research. So far, less and less attention has been paid to L2 learning and teaching materials, and L2 learning materials for learners with different L1 backgrounds.

As such, studying cognitive load and multimedia use in L2 education can provide us with insights into how language learning and teaching materials should be prepared and presented to the learners so that their naturally limited cognitive processes can deal with the complex job of learning an L2. It should be noticed that studying cognitive load through multimedia in L2 education is also directly related to the issues of whether learning or understanding in L1 and L2 carry different amounts of cognitive load for learners.

Furthermore, cognitive approach to learning challenged the behaviorist views of the late 1950s and early 1960s. The cognitive approach to learning emphasized the value of mental processes involved in learning. On the threshold of this approach, L2 learning and teaching focused on learners' use of cognitive strategies and emphasized learning materials that are in line with how learners cognitively process language information (Plass & Jones, 2005). Plass and Jones (2005) also emphasize that due to cognitive approach, the practice of language learning and teaching went beyond passive repetitive grammar-based activities, and began to give importance to the development of "linguistic competencies based on prior knowledge, linguistics knowledge, interaction with and understanding of text" (p. 468).

Likewise, in their commentary, Rikers, Van Gerven, and Schmidt (2004) argue that cognitive load theory (Sweller, 1988) can help us to select learning approaches and materials to increase meaningful learning in a specific domain.

Rikers et al. (2004) further claim that cognitive load theory can help to increase expertise by helping in the choice of adequate tasks, problems or materials that are at the level of the learners' expertise. This, according to Rikers et al. (2004), not only leads to better learning but also increases learners' motivation.

As for the relationship between working memory capacity and L2 comprehension, the results are contradictory: Leeser (2007) pinpoints that working memory capacity of L2 learners does not have a unique effect on their reading comprehension. Leeser (2007) argues that topic familiarity or readers' prior knowledge about a given passage topic directs the effects of working memory capacity on reading comprehension. Specifically, he claims that learners with high working memory can outperform those with lower levels of working memory capacity only when they have prior knowledge. Likewise, Chun and Payne (2004) make the point that working memory capacity did not lead to any performance differences in reading comprehension measures of their study. However, Harrington and Sawyer (1992) report that participants with higher L2 working memory capacity scored higher on reading skill measures. In addition, the researchers found a large significant correlation between L2 reading span and TOEFL reading. Similarly, investigations seeking the effects of working memory on higher-level cognitive skills like memory performance suggest that working memory capacity has effects on higher-level cognitive performance (e.g., Hambrick & Engle, 2002; Hambrick & Oswald, 2005). Therefore, due to the fact that language comprehension is one of the higher-level cognitive skills, it is not unreasonable to expect that working memory capacity affects L2 comprehension.

Theoretical Framework of the Study

Cognitive Load Theory

Cognitive load theory (Sweller, 1988) is one of the limited-capacity-based theories that rooted in cognitive science research. At the heart of the theory there is working memory that has a limited capacity. Cognitive load theory bases learning on an information processing system that consists of limited working memory that performs tasks related to learning and an unlimited long term memory that stores learned knowledge and skills. Information stored in long-term memory firstly needs to be dealt with and processed by working memory (Baddeley, 1992). The main premise of cognitive load theory is that in order to increase the quality of instructional design, limitations of working memory should be taken into account.

The concept of cognitive load is divided into three (Sweller, 1994):

1. Intrinsic load: Intrinsic cognitive load stems from the internal complexity of a learning material that is directly related to the number of interacting elements.To put it another way, it is the load imposed by the nature of the material to be learned. To illustrate, calculating "10 – 2" would lead to less intrinsic cognitive load than solving an equation like "10 – 2 + x = 23", since the number of interacting elements to solve the equation is higher.
2. Extraneous load: Extraneous load is the cognitive load that is created by the instructional design itself. For instance, an audiovisual presentation (narration + pictures) of a learning material will lead to less extraneous cognitive load than a visual-only (text + pictures) presentation of the same material, because

the audio channel is also used to deliver information thus reducing the possible overload on the visual channel in visual-only format.

3. Germane load: Germane load is imposed by the free cognitive resources allocated for schema construction and automation. This means that germane load is related to automating new information based upon prior information schemas. For example, while reading a text in an L2, advanced learners having knowledge of vocabulary and phonological knowledge of an L2 in question can easily make use of their prior knowledge existing in the form of schemas.

Moreover, even though intrinsic cognitive load is hard to manipulate, extraneous cognitive load can be changed easily through well designed instructions, which promotes both efficiency and effectiveness of learning (Clark, Nguyen, & Sweller, 2005). For that reason, one of the implications of cognitive load theory for multimedia learning is the modality effect that helps to decrease extraneous cognitive load in multimedia presentations. Basing his arguments on cognitive load theory, Mayer (2001) suggests a cognitive theory of multimedia learning.

Modality Effect

Cognitive load theory pinpoints that addition of an audio component to a visual presentation enhances learning. To illustrate, it is more effective to present content of an explanatory reading text in an audio format together with visual input like pictures than presenting the body of text in a written format together with corresponding pictures. Presenting text in an audio format decreases the overload on limited visual channel loaded by both pictures and written text. Several studies showed that audiovisual presentation of learning material tailoring to both audio and visual channel increased performance (e.g., Mayer, 2001; Leahy, Chandler, & Sweller, 2003).

Purpose of the Study

The current study seeks to investigate the effects of different levels of extraneous cognitive load imposed by multimedia design under the condition of high intrinsic cognitive load on L2 reading and listening comprehension of advanced learners of English both in the short term and in the long term. While trying to explore the effects of extraneous cognitive load on L2 reading and listening comprehension, this study also questions the interaction of cognitive load and L2 verbal working memory capacity. In other words, the present study looks for any individual differences stemming from L2 verbal working memory capacity.

The instructional design in the current study is as follows: high intrinsic cognitive load, relatively low extraneous cognitive load, and low germane load. In addition, it should be remembered that the conditions in the present study subsume relatively little extraneous cognitive load. Sweller, van Merriënboer, and Paas (1998) state that when dealing with high element interactivity materials, it may be crucial to decrease extraneous cognitive load in order to reduce total cognitive load to the manageable amounts that can be handled by our limited working memory. Therefore, in the present study, design conditions that have relatively small amount of extraneous cognitive load were chosen so as not to exceed the capacity limits of working memory. After all, one of the main aims of the study is to estimate the role of L2 verbal working memory while understanding an L2 text. For this reason, a design that already exceeds the limits of working memory would be inappropriate for the present study to gauge the role of L2 verbal working memory capacity in a situation where the total cognitive load is already beyond the top limits. Moreover, another related important aim of the study is to estimate the role of L2 verbal working memory in cognitive theory of multimedia learning and in the instructional designs suggested by the theory.

Not only the short-terms effects of cognitive load, but also the long-term effects of cognitive load are of concern for this study. More specifically, the research questions focus on the effects of multimedia conditions with different levels of extraneous cognitive load and L2 verbal working memory capacity on L2 reading and listening comprehension through both immediate and delayed comprehension tests.

Definitions of Key Terms Used in the Present Study

A brief clarification of some of the key terms used in the current study might be necessary for the reader based on Tuovinen and Paas (2004), Mayer (2001), Sweller (1994), and Baddeley (1992):

Multimedia: Multimedia means, as it is used in the present study, presentation of an L2 text by using both words and pictures. By words, it is meant that the L2 text is presented in verbal form: written text on a web-page format or recorded text (listening) on the computer. By pictures, it is meant that the text also consists of corresponding pictorial information: static pictures.

Modality: Sensory information processing channels used by learners to process presented information (visual versus aural modality).

Cognitive Load: The amount of mental resources allocated for information processing.

Extraneous Cognitive Load: Cognitive load imposed by instructional designs and presentation techniques.

Intrinsic Cognitive Load: Cognitive load imposed by the internal nature of the learning material that is linked to the number of interacting elements existing.

Germane Load: Cognitive load imposed by allocating free cognitive resources for schema construction and activation, thus promoting meaningful learning.

Working Memory: Component of the human cognitive architecture that deals with temporary storage and processing of upcoming information.

Prior Knowledge: The amount of existing knowledge people have in a specific domain.

Retention: Remembering, and recognizing studied and learned knowledge, thus being able to answer content related questions.

Transfer: Applying learned knowledge to new situations, thus being able to use prior learned material to solve new problems or to use in new situations.

Instructional Efficiency: High performance associated with low cognitive load or mental effort.

Conceptual Question: Comprehension question that asks for uncovering any underlying phenomenon, factor or principle.

Trouble Shooting Question: Comprehension question that asks for solving out why certain expected phenomena do not occur or how unexpected results happen.

Redesign Question: Comprehension question that asks for finding a solution to a problem by redesigning what is covered in the text.

Prediction Question: Comprehension question that asks for describing the role of a particular phenomenon or element.

Instructional Time: A certain total amount of time given to the participants to read or listen to the treatment text.

Reading/Listening Time: The amount of time a participant can read or listen to the treatment text (as many as possible) in the total amount of the allotted instructional time.

CHAPTER 2

LITERATURE REVIEW

This chapter discusses the theoretical and methodological issues that are related to the present study. The main focus of the current study is on extraneous cognitive load (referred to as ECL henceforth) and multimedia learning. First of all, the chapter presents Mayer's (2001) cognitive multimedia learning theory. Second, a review of previous research and theory on cognitive load (hereafter CL), and an overview of the research on the measurement of CL are provided. Third, the chapter reviews the concept and measurement of working memory (WM henceforth) and its relationship to L2 comprehension. Next, discussing the relationship between L2 instruction through multimedia learning and CL theory, it summarizes the findings of previous research agenda. Finally, it sums up the theoretical issues and insights that bring CL theory, L2 education, WM capacity and multimedia learning together.

Cognitive Multimedia Learning Theory

CL theory was used to understand the knowledge and human cognitive architecture interactions and the effects of these interactions on instructional format (Sweller, 1988; Chandler & Sweller, 1991; Sweller et al., 1998). In his 1997 article, "Multimedia Learning: Are we asking the right questions?" Mayer tried to apply the ideas blossomed by CL theory to the field of multimedia learning. Mayer (1997) suggests a generative theory of multimedia learning (see Figure 1) that sees learners as active knowledge constructors. According to Mayer (1997), information is processed through two channels: visual and verbal. This means that verbal information is processed through verbal system and visual information through visual system. After relevant verbal and visual information are selected (added to WM) a propositional or pictorial representation is constructed. In other words, the selected relevant information from the presented input is stored either in the text base (as a propositional representation) or in the image base (as a pictorial representation). Afterwards, the selected information is organized into coherent verbal or visual models depending on the type of input. The last step is to connect the verbal and visual models, a process called "integration". During integration, learners do not only make connections between verbal and visual models but also connect the organized information to prior information already held in long-term memory. In other words, verbal and visual models are integrated into already existing knowledge representations or schemata.

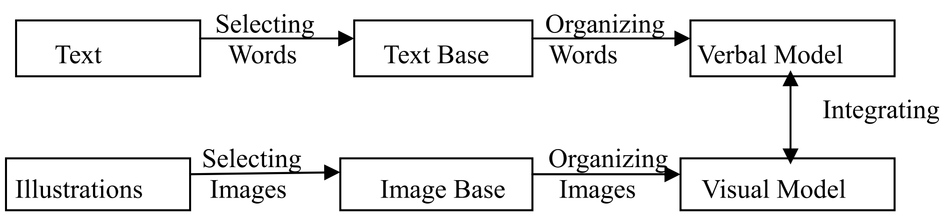


Fig. 1. A generative model of multimedia learning (taken from Mayer, 1997)

Mayer's (1997) generative theory of multimedia learning draws on Wittrock's (1974, 1989) generative theory and Paivio's (1986) dual coding theory. He states:

> From generative theory I take the idea that meaningful learning occurs when learners select relevant information from what is presented, organize the pieces of information into a coherent mental representation, and integrate the newly constructed representation with others. From dual coding theory, I take the idea that these cognitive processes occur within two separate information processing systems: a visual system for processing visual knowledge and a verbal system for processing verbal knowledge. (p. 4)

Mayer (2001) extends the borders of his previous (1997) generative theory. Mayer (2001) describes multimedia learning as learning from multimedia materials that bring together more than one mode of delivery. For instance, learning from text + picture format constitutes a multimedia learning material or environment according to Mayer (2001). By applying basic principles of CL theory to multimedia learning, Mayer (2001) presents a cognitive theory of multimedia learning that is an information-processing model (see Figure 2):

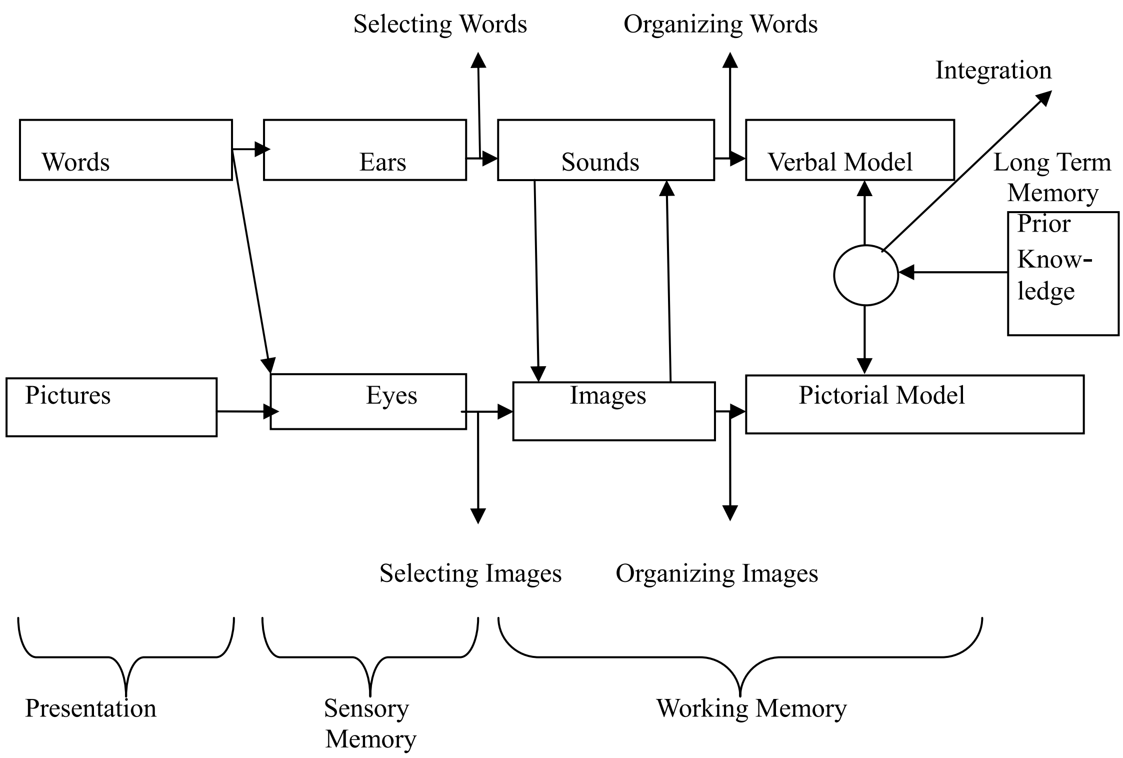


Fig. 2. Mayer's (2001) cognitive multimedia learning model (adapted from Grimley, 2007)

Words and pictures enter sensory memory as spoken (sound) or written (visual) stimuli through eyes or ears. Selected information in the sensory memory is then transferred into WM. Written and visual information is organized in the WM to form a model that makes sense for the learner. When information enters into WM by depending upon modality (visual and auditory information), the models of visual and auditory information should be integrated with both each other and the prior knowledge for meaningful learning to occur. Obviously, while dealing with novel information there is no prior knowledge to integrate into the final information model that encourages meaningful learning. As one can see in the model above (figure 2: the arrows going between sounds and images), auditory information and visual information can be transferred into different forms (auditory or visual, spoken or written) after being selected into WM. It is important to notice that meaningful learning occurs when pictorial and verbal model of the same information is integrated into one another, a case which explains the multimedia principle of Mayer's (2001) cognitive theory of multimedia learning. In other words, in order to learn meaningfully, learners need to develop both a verbal and a pictorial model of the information to be learned and make connections between the two models.

Mayer (2001) outlines seven principles of multimedia learning:

Multimedia Principle

Better learning occurs when learners attend to both words and pictures simultaneously than words alone. The theoretical rationale behind the multimedia principle is that when presented with both words and pictures, learners are more likely to build up verbal and pictorial models and make connections between them. On the other hand, when presented with words alone be it auditory or visual, learners have the opportunity to construct a verbal mental model but not a pictorial mental model. In such cases, learners cannot build connections between the two models in the absence of a pictorial model.

Spatial Contiguity Principle

There is better learning when learners receive words and corresponding pictures that are presented near each other rather than far from each other. The rationale is that when pictures and words are near to one another, they can be more easily held and processed in WM at the same time; and that learners do not have to use more cognitive resources to search for relevant material.

Temporal Contiguity Principle

When words and corresponding pictures are presented simultaneously rather than successively, people learn better. If the words and pictures are presented at the same time, learners are more likely to be able to process mental representations of both in WM simultaneously. Moreover, when words and corresponding pictures are presented simultaneously, they are more likely to be integrated into pictorial and verbal mental models and learners are more likely to be able to make connections between them. Mayer, Moreno, Boire, and Vagge (1999) argue that learners are more likely to make connections between corresponding pictorial and verbal representations when both of them are processed and held in WM at the same time. This effect reaches a maximum point under contiguous presentation and is minimized by successive presentation, which is a conclusion supporting temporal contiguity principle of cognitive theory of multimedia learning. Moreover, what is new in Mayer et al. (1999) is that one of the experimental conditions was successive small bites condition in which small narration portions were presented before or after the corresponding small portions of animation in each successive portion of the whole presentation. Interestingly, learners in the “successive small bites” condition scored as good as the concurrent group who were presented with corresponding animation parts and narration parts simultaneously. Mayer et al. (1999) also emphasize that further research should try to collect multiple measures of understanding and learning. Therefore, this study includes a retention test consisting of matching items, recognition items, open-ended questions, and a transfer test, which are measures of different cognitive learning styles and outcomes.

Coherence Principle

Multimedia instructional materials lead to better learning when extraneous information sources (irrelevant, unneeded words or pictures, though they may be interesting) are excluded than included. This depends on the assumption that extraneous words or pictures compete for limited cognitive resources. Mayer and Moreno (2000) reported that students achieved better scores on both retention and transfer tests when extraneous, irrelevant sounds were excluded from the multimedia environment. This result is totally in line with the assumption that extraneous words and pictures are detrimental to students’ understanding stated in the form of “coherence principle” in Mayer (2001). Mayer and Moreno (2000) differentiated between the diminishing effects of seductive details in the verbal learning material and extraneous auditory adjuncts. They pinpointed that while auditory adjuncts led to cognitive overload in the auditory channel, seductive details appeared to promote “wrong” schemas into which the information to be learned is assimilated. Mayer and Moreno (2000) concluded that irrelevant auditory adjuncts lead to worse learning outcomes by decreasing the amount of to-be-learned learning material and by reducing the cognitive resources used for making connections between verbal and visual models. In line with the conclusions of Mayer and Moreno (2000), the current study guaranteed that experiments were conducted in a learning environment without any sort of detrimental sounds that could function as a contaminating variable for the results of the study.

Modality Principle

Better learning occurs, when words are presented in an auditory form rather than visual form. In other words, people learn better from narrated text with pictures than from text with pictures, because both auditory and visual channels are used (auditory one for the words and visual one for the pictures) rather than only the visual channel which would be overloaded by both words (textual presentation) and pictures. Moreover, when both of the two channels are used, it becomes easier for the learner to set up verbal and pictorial models and integrate them.

Redundancy Principle

There is better learning when redundant information is excluded. For instance, students learn better from narration and pictures than narration, pictures and text. This is because the redundant textual information overloads the visual channel. As a result, cognitive load increases, which in turn impedes successful learning.

Individual Differences

There are stronger design effects for learners with low prior knowledge rather than learners with high prior knowledge and for high spatial ability learners rather than for low spatial ability learners. Mayer (2001) explained individual differences through two factors: prior knowledge and spatial ability. He suggested that multimedia design effects are on the stage for low prior knowledge students and students with high spatial ability. Expending this conclusion, Grimley (2007) asserted that multimedia instructional design effects were directly related to learner characteristics including cognitive style, gender and prior knowledge. Mayer (2001) claimed that empirical findings on individual differences have been far from being conclusive partly because it is hard to define individual differences and to gauge them properly. As for suggestions for further research, he stated:

> (…) Some worthwhile venues for future research include the role of visual and verbal working memory capacity and the role of visual and verbal learning style. (p. 181)

In line with Mayer's (2001) suggestion given above, the current study looks at the effects of WM capacity differences as a whole. Specifically speaking, this study examines the effects of L2 verbal WM capacity on L2 reading and listening comprehension.

Mayer (2001) accepts that the empirical studies covered in his book seek for ideal multimedia designs to reduce ECL. In other words, the multimedia principles discussed above deal with how to reduce ECL of multimedia instruction. Similarly and naturally, much of the work in multimedia learning that try to combine CL theory and multimedia learning deal with ECL, simply because multimedia instructional designs are directly related to ECL. As such, most studies tackled with looking for better multimedia instructional designs in which ECL can be reduced as much as possible.

Plass, Chun, Mayer, and Leutner (2003) emphasize that multimedia learning which is thought to be an effective way of learning can be detrimental to effective learning depending upon learning conditions and individual differences among the learners. The investigators highlight that processing of visual annotations (annotations given for vocabulary items) given in an L2 reading text led to the worst text comprehension than no-annotation, visual and verbal annotation, and verbal annotation conditions. In other words, according to Plass et al. (2003), high cognitive load imposed by processing of visual annotations hinders text comprehension, which in turn means that types of annotations provided to make vocabulary meaning clear led to different amounts of CL, which in turn decreased reading comprehension performance.

Grace-Martin (2001) states that while preparing multimedia presentations the complication is that not all WM load is bad for learning. He warns that because not all amounts and sorts of CL is bad for learning, attempts to minimize amount of CL in educational multimedia may turn out to be detrimental. Grace-Martin (2001) argues that any learning environment including educational multimedia should be appropriately challenging for learners, which can be promoted by the amount of CL a learning material carries. According to Grace-Martin (2001), optimal CL may be between so-called minimal load and a maximum load that overwhelms learners. He concludes that good educational multimedia challenge learners without exceeding their limited capacity (thus not overwhelming them) regardless of learners' prior knowledge.

Lee, Plass, and Homer (2006) point out that intrinsic and extraneous CL on computer-based simulations can be manipulated. The investigators emphasize that the effectiveness of intrinsic and extraneous CL manipulations in visual displays depends upon learner's prior knowledge. Lee et al. (2006) state that when intrinsic CL is low, manipulation of extraneous CL facilitates learning for low prior knowledge learners, but not for high prior knowledge students. They expand this conclusion by further stating that when intrinsic CL is high, ECL manipulations are beneficial for all learners independent of their level of prior knowledge.

Cognitive Load and Cognitive Load Theory

CL refers to the load that performing a specific task imposes on the human cognitive system (Sweller et al., 1998, p. 266). It can be divided into two concepts: mental load and mental effort (Paas, 1992, p. 429). According to Sweller et al. (1998), mental load refers to the load that is created by the characteristics (e.g. element interactivity) of a particular task in question while mental effort is the amount of cognitive capacity or resources allocated by the learner to do a given task. CL theory distinguishes between three types of CL: intrinsic, extraneous, and germane (Sweller, 1994). Intrinsic CL is imposed by the number of information elements and their interactivity; ECL by the presentation design; and germane load by the free cognitive resources allocated for learning.

Paas and van Merriënboer (1994) argue that CL has a causal dimension referring to the interaction between task and learner characteristics; an assessment dimension that refers to the concepts of mental load, mental effort and performance. The authors state that such task characteristics as "task novelty, time pressure and reward systems" and their interactions with learner characteristics consisting of "cognitive capabilities, cognitive style, preferences, and prior knowledge" are sample causal factors leading to CL (Paas & van Merriënboer, 1994, p. 354). Likewise, Paas Tuovinen, Tabbers, and Van Gerven (2003) point out that mental load stems from the interaction between task and learner characteristics while mental effort is the actual cognitive capacity allocated for performing a specific task and for meeting the demands imposed by that task, which is also accepted by Paas and Merriënboer (1994).

In addition, Paas and van Merriënboer (1994) warn that environmental factors like high noise or heat may also contribute to CL, thus increasing it. Brünken, Plass, and Leutner (2004) report that seductive background music did not interfere with learning in their study. On the other hand, Mayer and Moreno (2000) argue that seductive background music leads to a decrease in the performance of retention and transfer tests, thus overloading the auditory channel. Mayer and Moreno (2000) point out that participant performance decreases when background music is added to narration than when narration is presented alone. By pointing out the discrepancy between their study and that of Mayer and Moreno (2000), Brünken et al. (2004) argue that the background music used in their study was not relevant to the narration presented to the participants. They further state that the load of background music (a movie soundtrack without vocals) did not impose load in the auditory WM, since it was not related to the process of knowledge construction.

Controlling CL is crucial so that WM is not overly loaded in completing a task. WM processes information in two streams that are separated from each other by modality, one making use of a visual-spatial sketchpad and the other a phonological loop. This basic assumption implies that more cognitive capacity should be available or cognitive capacity is being used more efficiently when information is split between the auditory system and the visual system. As such, any amount of ECL imposed on the limited WM will hinder and prevent learning and understanding of learning materials. Heo and Chow (2005) argue that when instruction is delivered to the learners in a way that effectively decreases extraneous cognitive load, free WM resources or capacity can be allocated for processing of new information. They conclude that teaching a programming course on an on-line tool that effectively reduces the cognitive load has the promise to ease the burden on both instructors and students.

CL theory (Sweller, 1988, 1994) assumes that humans have limited WM capacity but unlimited long-term memory that can handle an enormous amount of information stored as schemas, which was also predicted by previous studies: Miller (1956) claimed that we cannot deal with more than seven information items at a time and Simon (1974) even stated that this number can be reduced to five (as cited in Sweller & Chandler, 1994, p. 186). Therefore, the basic assumption of CL theory is that learning occurs through limited WM and an unlimited long term memory. In other words, CL theory suggests that WM load or the CL imposed upon WM is the most critical factor for whether effective learning will happen or not. So, if the load on WM exceeds the limits of WM, then effective and meaningful learning will not occur.

Both CL theory (Sweller, 1988) and cognitive theory of multimedia learning (Mayer, 2001) assume that human WM capacity in learning situations is limited and has to be distributed through several cognitive processes at the same time. This assumption is based upon the basic premise that human WM has a limited capacity. For instance, Brünken et al. (2004) showed that cognitive resources used to process verbal information presented in an auditory format are not available for processing a secondary simultaneous auditory task. In other words, the simultaneously presented second auditory task overloads the auditory channel that is already occupied by the primary auditory task. This finding is in line with the finding of Brünken et al. (2002), which demonstrated that a primary visual task decreased the performance of a secondary visual task that overloads the visual channel. Similarly, Carlson, Chandler, and Sweller (2003) suggest that instruction is less likely to lead to learning, if limited WM has to process multiple information elements at the same time. On the other hand, the researchers state that if multiple elements of information are processed successively or serially, corresponding WM load is light.

CL theory presumes a cognitive architecture including a WM component that is very limited especially when dealing with novel information. On the other hand, this limited WM can become reasonably and effectively unlimited while dealing with familiar or already stored information that is kept in the unlimited long term memory. Sweller (2004) argues that information can enter WM in two ways: in the first way, information already held in long-term memory enters WM to be processed. In this case, irrespective of the amount of information entering WM, no WM load is observed. In the second way, information that is new for a learner enters WM from sensory memory. In this case, depending upon the instructional design and inner characteristics of the material to be learned, CL on WM is likely. By highlighting the resemblance between the capacity of long-term memory to store information and that of genomes to hold genetic information, Sweller (2004) points out that the main function of instruction is to store information in the long-term memory. According to Sweller (2004), both genetic information held in genomes and information stored in long-term memory function as a central executive that directs human behavior. He further states that such a central executive is missing when limited WM has to process novel information. As such, the aim of instructional design should be to replace the missing central executive in non-existence of prior knowledge, and lessen this support gradually as more and more information is stored in the long-term memory. Sweller (2004) asserts that WM is limited when dealing with novel information, since there is no already-organized information held in long-term memory that indicates how new information should be organized.

Likewise, Sweller et al. (1998) claim that the main goal of instruction is the construction and automation of schemas. They further argue that schemas are stored in long-term memory; however, to construct schemas, information is processed in WM. Long term memory stores information in the form of schemas that categorize information elements on the basis of the manner in which they will be made use of in the future. Meaningful learning and skilled performance occurs by constructing complex schemas through combining elements of information consisting of lower level schemas. Schema automation leads to the spontaneous and unconscious use of schemas in learning and practice situations, which can free limited WM capacity. As such, both schema construction and schema automation can bypass WM limitations and lead to more meaningful learning that will later be combined with already existing schemas held in long term memory. What this suggests for instruction is that instruction should be designed in such a way that encourages schema construction and automation. In addition, Sweller et al. (1998) pinpoint that WM and consciousness can be equated, since only the contents of WM can be monitored consciously. So, it is reasonable to say that extraction and manipulation of information in WM play a crucial role in the construction of schemas that lead to understanding.

CL imposed on limited WM heavily depends upon the level of prior knowledge in a specific domain. Kalyuga and Sweller (2004) argue that the reason why instructional design that help novices to learn better fails to do so with expert learners or with learners having prior content-related information is that more knowledgeable learners do not need the well guidance novices need in terms of both the content and presentation of the material to-be learned. Likewise, Sweller (2003) states that the manner in which information is processed changes radically as the familiarity with that information increases (as cited in Kalyuga & Sweller, 2004, p. 559). Furthermore, Paas et al. (2003) stress that intrinsic CL through element interactivity is determined by both the nature of the learning material and the expertise of learners. It is clear from Kalyuga and Sweller's (2004) argument that optimal learning depends upon instructional designs that are constructed on the basis of learners' level of content knowledge in a subject area, which ultimately alters WM load.

Yeung, Jin, and Sweller (1997) claim that instruction efficiency depends upon the amount of ECL it imposes on the learner through split attention or redundancy effects. However, the researchers emphasize that the effects of extraneous ECL on instructional efficiency are directly linked to the learners' expertise. In other words, badly designed instructions lead to less learning depending upon learners' expertise or prior knowledge in a given area. In Yeung et al. (1997) study, as for performance on reading comprehension, learners with more knowledge preferred the less complete text and less knowledgeable learners chose the reading text with the additional information. Not surprisingly, in this study, performance of more expert learners improved in the condition of the incomplete text, while the performance of less expert learners improved with the additional material. Yeung et al. (1997) conclude that an instructional design in a specific format of presentation may facilitate learning for some learners, but also may retard learning for some other learners, referring to learner expertise or level of knowledge in a specific area. Likewise, in Schnotz's (1993) study that investigates how readers with different levels of prior knowledge adapt their reading processing to continuous and discontinuous text organization, readers with high prior knowledge could easily deal with a discontinuous text format that is more difficult and that carries more CL. Moreover, their comprehension scores did not change when they read the continuous text. Even more, readers with high prior knowledge did gain even better recall scores after reading a discontinuous text. Schnotz (1993) concludes that coherence formation in a reading text highly depends upon readers' prior knowledge.

McCrudden, Schraw, Hartley, and Kiewra (2004) claim that familiar information to be learned imposes less intrinsic CL than unfamiliar information because of learners' existing prior knowledge. The researchers exemplify this point by stating that a text on lightning formation leads to low intrinsic CL for a "meteorologist" but to high intrinsic CL for a learner who does not have prior knowledge about lightning formation. The participants enrolled in the present study are first year and third year students in the Faculty of Education at Boğaziçi University. So, from the perspective of the field of study, the participants are assumed to have no prior knowledge about how a tornado forms simply because the departmental programs of the faculty does not include any course on tornadoes or any other environmental events. As such, the reading text used in the current study can be considered to carry high intrinsic CL for the participants of the study, which will be confirmed by the results of the prior knowledge test.

McCrudden et al. (2004) also state that text presentation can be considered to influence ECL. In their study, the investigators compared a sentence by sentence presentation of a reading text with a whole text presentation. According to the researchers, sentence by sentence presentation led to more extraneous demands by increasing the amount of previous information to be held in WM. In his paper discussing reading from paper versus screens, Dillon (1992) also claims that splitting sentences across screens might destroy comprehension by putting extra load on the limited WM that tries to hold the previous information on one hand, and the current information on the other. Even though the text was not presented across screens in the present study, pictures were placed between the sentences, which might have increased ECL. However, this ECL must have been bypassed easily by the participants since the presentation was learner-paced. In other words, it is unlikely that the presentation of the text body contributes extremely to ECL demands in the present study.

As a result, CL theory is concerned with instructional techniques for managing WM load and its limited capacity in order to facilitate the learning changes in the long term memory. CL theory primarily focuses on the learning of complex cognitive tasks consisting of quite large number of information elements and their interactions that are needed to be processed simultaneously by learners in order to achieve meaningful learning. In result, the number of interacting information elements lead to either high or low CL on the part of the learners. Paas, Renkl, and Sweller (2004) emphasize that element interactivity in a learning material is the driving force of intrinsic CL that is a natural aspect of instruction. Paas et al. (2004) claim that under the conditions of both underload and overload, meaningful learning may not occur. As a result, the investigators suggest that underload instructional conditions should be accompanied with practice that increases the load level to a reasonable level thus challenging the learners. On the other hand, overload learning situations should be followed by practice sessions that aim at reducing the CL to a manageable point. Because both underload and overload prevents meaningful learning from occurring, there should be a balance in any instructional design on the basis of the expertise or prior knowledge of the learners as highlighted by Paas et al. (2004):

> Instructional designers need to realize that reducing cognitive load is not necessarily beneficial, particularly in cases where working memory capacity limits are not exceeded and the load is already manageable. As long as the load is manageable, it is not the level of load that matters, but its source. If the load is imposed by mental activities that interfere with the construction or automation of schemas, that is, ineffective or extraneous load, then it will have negative effects on learning. If the load is imposed by relevant mental activities, i.e. effective or germane load, then it will have positive effects on learning. (p. 3)

Brünken et al. (2004) argue that with respect to the effects of ECL a less is more rule could oversimplify the instructional design in which any learning material is presented. Instead, the researchers advise that we should try to optimize ECL by focusing on the interaction between the demands of the learning material, learning and the mode of presentation.

Both Moreno and Mayer (1999) and Mayer (1997) point out that low-experience learners are more likely to exhibit modality effect. Low-experience means that learners have low prior knowledge, and that information presented is not familiar. As stated earlier, McCrudden et al. (2004) state that such information decreases intrinsic CL. When these assumptions are combined with Sweller's (1994) point that effect of ECL may only be seen under high intrinsic CL, a research design testing modality should have high element interactivity learning material and learners should not be familiar with the content, which is the case in the current study.

Paas, Renkl, and Sweller (2003) hypothesized that in the case of learning materials that have high element interactivity, thus increasing the number of related chunks of information to be held and processed simultaneously in WM, it becomes important to decrease the amount of CL imposed upon visual channel. Therefore, a modality effect may not come up under the condition of low element interactivity. Within the same line of logic, Ginns (2005) meta-analysis of modality effect shows the effect of the difficulty of the learning materials on modality effect. Specifically speaking, in Ginns' (2005) study, average modality effect sizes are smaller for low element interactivity materials than for high element interactivity ones and lower for self-paced presentations than for system-paced presentations. Likewise, Tabbers, Martens, and van Merriënboer (2001) and Tabbers (2002) claim that modality effect interacts with pacing of the presentation in that modality effect shows up with system-paced instructions not with learner-paced instructions. Even though in most of the previous research the L1 was English, Tabbers et al. (2001) and Tabbers (2002) employed instruction in L1 Dutch. ECL is created by the manner in which information is presented to the learners and by the learning activities learners are required to fulfill. In contrast to intrinsic CL, ECL can be manipulated by instructional changes. The only way to reduce intrinsic CL is to reduce the number of interacting elements. If learning activities are unrelated to schema construction and automation or, if you wish, to meaningful learning, they impose extraneous or ineffective load on the part of the learners. However, if the learning activities required of learners are directly related to meaningful learning, they lead to germane load or effective load. All these three types of CL are considered additive in that the total load should not exceed WM capacity in order for learning to occur (see Figure 3).

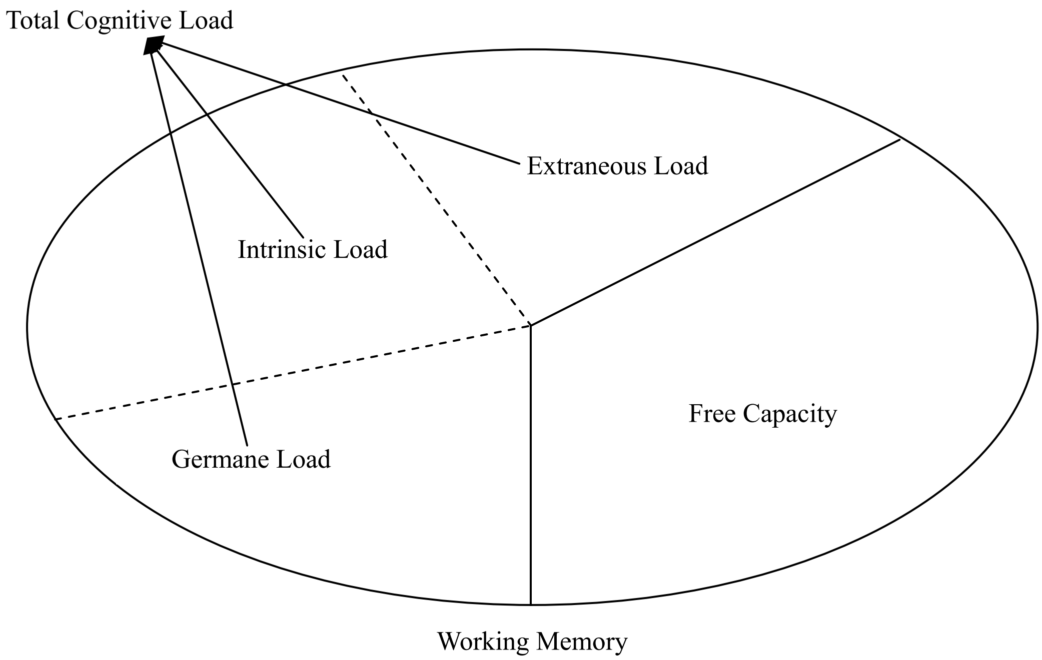


Fig. 3. Additive nature of types of cognitive load

Sweller and Chandler (1994) suggest that should the intrinsic CL be low, ECL might not be very important. They further state that ECL can become critical for a learner when tackling with intrinsically loaded learning materials. Sweller and Chandler (1994) describe an element of information as low-order-schemas that may interact with each other, and constitute the whole information to be assimilated into the long-term memory. The researchers argue that intrinsic CL is imposed solely by the element interactivity, not by the total number of elements to be learned, the difficulty of learning materials in an area is determined by both the number of elements and their interactivity though. As Paas et al. (2004) point out; CL theory primarily focuses on element interactivity in which relevant elements interact with one another because high-element interactivity makes instructional materials difficult to understand. As such, information may be difficult to learn due to the great number of elements to learn, but may cause low CL because of the low element interactivity existing. In other words, even if the total number of elements is small, high element interactivity leads to high intrinsic CL for learners. Sweller and Chandler (1994) emphasize that the difficulty of a learning material due to element interactivity totally depends on the prior knowledge of the learners or already existing schemas one has in her or his long-term memory. Therefore, the difficulty of information changes from person to person depending upon their level of already-assimilated schemas that are held in one's long-term memory.

Sweller et al. (1998) highlight that understanding occurs only when elements of a high element interactivity material are held and processed simultaneously in WM. The researchers state that learning materials consisting of a procedure in which all steps are dependent upon each other constitutes an example. In the present study, tornado formation is explained in seven critical steps that lead to one another in turn. In other words, for the 4. phenomenon in tornado formation to occur, the previous 3. step and its obligatory interactivity with the 1. and 2. phenomena are essential. Actually, this is the case in most natural phenomena including earthquakes, volcanic eruptions, because none of these events happen suddenly in one step. Thus, as far as the current study is concerned, learners should be able to process all the desirable seven steps of tornado formation in order to understand how a tornado forms. Even though the text is used within the lieu of L2 reading and listening, in order to answer the comprehension questions (especially transfer questions) learners must understand the basic seven steps.

Sweller et al. (1998) also claim that under the condition of high element interactivity material, it may be crucial to decrease ECL to manageable levels that can be handled with by our limited WM in order to reduce total CL. Within the same line of logic, Carlson et al. (2003) point out that instructional techniques reducing ECL can contribute to superior learning by reducing the total CL on WM under the condition of high element interactivity learning materials that carry high intrinsic CL. However, as for learning elements that can be processed serially, Carlson et al. (2003) highlight that they impose little intrinsic CL on WM. Under such learning situations, the researchers suggest that ECL imposed upon WM by instructional designs may not be of central importance to effective learning. The investigators conclude that such instructional helpers like diagrams that help to lessen the amount of extraneous WM load may facilitate learning under certain conditions where learning material has high element interactivity. Otherwise, there could be no performance differences among learners under low ECL and high ECL circumstances.

On the other hand, Bannert (2002) emphasizes that although reduction of ECL lessens the total load on WM, it does not guarantee that available cognitive resources will be allocated for meaningful learning. So, Bannert (2002) points at the lack of concentration on individual management strategies in CL research and argues that it is also important to know how learners deal with high CL learning environments. According to Bannert (2002), computer-paced multimedia environments are full of information and should have a good design to keep ECL under control. Moreover, Bannert (2002) states that there should be a distinction between short term and long term learning benefits of certain instructional manipulations in order to reach a better understanding of the effects of extraneous and germane CL on learning outcomes. In line with this suggestion, the current investigation tests both immediate and delayed effects of ECL and WM capacity on retention and transfer.

Sweller (1994) points out that CL theory concerns with learning and understanding difficulty. The learning difficulty if it is due to ECL is described to be artificial in that instructional design can change it. On the other hand, intrinsic CL is fixed in a given area simply because it is related to the nature of the learning material, which may hinder understanding. Sweller (1994) also suggests that ECL only matters when intrinsic load of an instructional material is high. In other words, according to Sweller (1994), under low intrinsic CL, instructions designed to reduce ECL may not lead to significant learning outcomes. Likewise, Leahy et al. (2003) point out that test score differences between audio-visual and visual only groups were greater under high element interactivity. It is claimed that level of element interactivity may explain not only why some material is difficult to learn but also why it can be difficult to understand the content of the material. Therefore, understanding becomes an issue when a high element-interactivity instructional material is to be learned. Ginns (2005) suggests that the choice of learning materials should be based on an intricate match between learners and learning materials, which takes into account high (six to eight interacting elements) or low (one to two interacting elements) element interactivity and prior knowledge of learners. So, it is reasonable to argue that the reading and listening text used in the current study, with seven interacting elements carry high intrinsic CL.

Marcus, Cooper, and Sweller (1996) pinpoint that whether we understand instructions or procedures easily depends upon two general factors: “the intrinsic complexity of the information and the manner the information is presented” (p. 49). The researchers add that these factors interact both with each other and human cognitive system. The presentation conditions in the present study, “a text with pictures” (T+P henceforth) and “a narration with pictures” (N+P hereafter), do not in fact constitute a strict hierarchy in terms of the ECL they impose on participants. In other words, the conditions mentioned above actually consist of relatively reasonable levels of ECL depending upon the presentation manner they include; however, there is a difference between the conditions in terms of the amount of ECL they cause. Marcus et al. (1996) argue that when learning is aimed, instructional materials should be designed by keeping in mind the fact that WM capacity is limited. In the current study, learning is aimed in the sense that in order for participants to answer retention and transfer questions, they should learn about how a tornado forms. Marcus et al. (1996) emphasize that understanding instructions depends upon the degree of element interactivity if the elements of information cannot be linked to pre-existing knowledge schemas that reduce any possible amount of CL. The investigators further suggest that understanding depends upon not on the amount of information to be learned but on the amount of information that must be held simultaneously in WM. The current study involves a 254-word reading text on how tornadoes form. Even though the text is short or the total amount of information to read/listen and understand is small, the number of interacting elements to be processed simultaneously in WM makes it a hard passage to understand. The text used in the present study defines a procedure "tornado formation" that consists of seven successive interdependent steps, which makes it an expository reading passage with high element interactivity.

Unlike investigations of CL and modality effect on comprehension and learning in a wide range of subject areas in L1 English, studies of CL and modality effects on comprehension and learning in L2 English contexts are scarce including English language education. In addition, most of the studies in L2 contexts are limited to research conducted on presentation modes of annotations (i.e., glosses) not that of the whole learning material. For instance, Plass et al. (2003) argue that while reading an L2 text, learners should be given the right to choose in which mode learners prefer to read and process the given information (visual or verbal etc.). This argument stems from the performance differences among learners in terms of their spatial and verbal ability. Depending upon the type of annotations that provide information about the meaning of vocabulary items in an L2 reading text, text comprehension performance of learners differ. According to Plass et al. (2003), it is the visual annotations that impose the highest level of CL on learners. So, text comprehension performance of those who were exposed to visual annotations in Plass et al. (2003) was the lowest. Plass et al. (2003) conclude that multiple representations of information in the form of both verbal (text translation of a word) and visual annotations (a static photograph or a short video) -like the case in reading an L2 text with visual aids- may not always help learning. In fact, this could decrease learning in low-ability learners to a great extent.

In their study aiming at exploring the effects of explanatory notes on reading comprehension, Yeung et al. (1997) argue that the effects of instructional formats on reading comprehension change from learner to learner depending upon the level of expertise of the learners. The study of Yeung et al. (1997) yielded that vocabulary items are learned best when their meanings are separated from the reading text. Yeung et al. (1997) pinpoint that comprehension or learning outcomes cannot be explained only through instructional design, and that nature of the learning material and learner characteristics should also be taken into account. They emphasize the importance of the interaction between reading materials and learner expertise in explaining any increase or decrease in reading comprehension.

Plass and Jones (2005) suggest a model of cognitive processing in L2 acquisition that depends upon interactionist models and a cognitive theory of multimedia learning (see Figure 4). Plass and Jones (2005) give the following definition of L2 acquisition with multimedia:

> (…) For the purpose of this chapter, we will focus on second language acquisition with multimedia, that is, the use of words and pictures to provide meaningful input, facilitate meaningful interaction with the target language and elicit meaningful output. (p. 469)

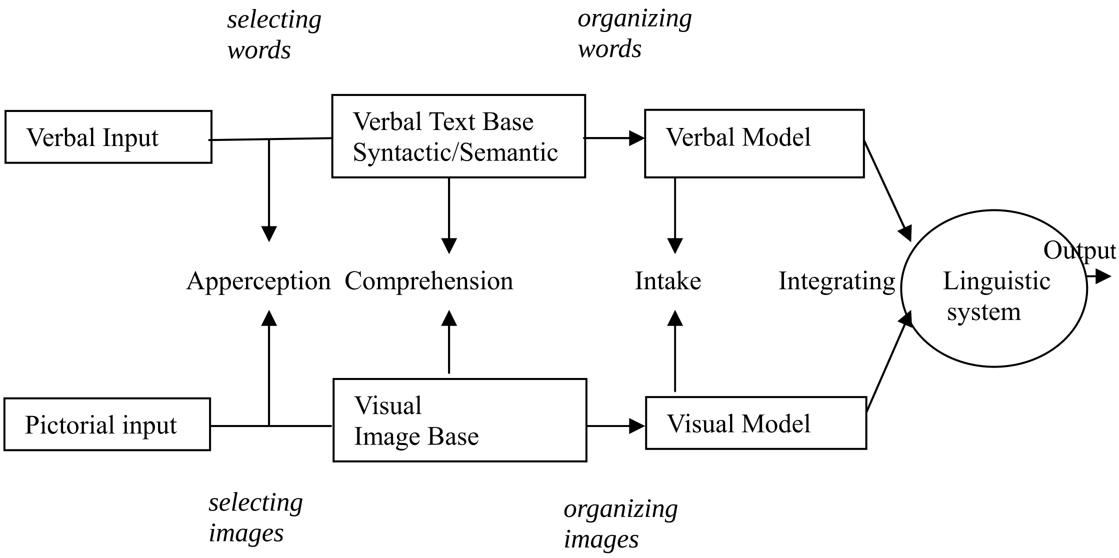


Fig. 4. Plass and Jone's integrated model of SLA with multimedia (adapted from Plass & Jones, 2005)

Plass and Jones (2005) argue that L2 learners should have the chance to interact freely with multimedia learning materials. That is to say, according to Plass and Jones (2005), L2 learners should have control over multimedia materials. Basing their argument on an interactionist approach to L2 acquisition/learning, Plass and Jones (2005) claim that multimedia materials are interactive learning environments that provide learners with the chance to interact meaningfully with the learning material.

The Measurement of Cognitive Load

According to Sweller et al. (1998) CL consists of complex relationships among mental load, mental effort and performance in a task. In other words, CL is a multidimensional construct, which makes it challenging to measure it.

Whelan (2007) presents three forms of measuring CL: a) dual task paradigm; b) physiological measures of CL; and c) subjective measures. After reviewing literature on these three paradigms of measuring CL, Whelan (2007) provides fMRI neuroimaging technique as an alternative and reliable way of assessing CL. He argues that especially the intrinsic and extraneous components of CL may be directly observable through neuroimaging studies, since other paradigms are sensitive to different types of CL under different types of situations or to some combination of them but not to the others.

In the dual task method, a secondary task performance is assumed to be reflective of the level of CL created by the primary task. According to Brünken et al. (2002), dual method is based on the assumed premise that even though WM capacity is limited, it can be distributed with flexibility. The methodology assumes that if two simultaneously presented tasks need the same cognitive resources, then while performing these two tasks cognitive resources are split between them. Accordingly, what is left for performing the second task is the remaining cognitive processing capacity after performing the first task. Thus, performing the simultaneously presented second task depends upon the CL imposed by the first task. The secondary task should be of the same type of which the primary task is. To illustrate, if the first task is a visual task then the secondary task should be a visual one as well. What is more, it is important that the second task is always easier than the first one. In dual-task paradigm, CL is gauged through participants' reaction time performance in the secondary task. In their (2002) study, Brünken et al. (2002) found that reaction times in the visual-only multimedia group (on-screen text & pictorial information) were significantly higher than those in the audiovisual multimedia group (narrated text &

pictorial information). This simply shows that in the first group visual channel was overloaded by the primary task first and then overloaded additively by the secondary task. However, since in the audiovisual group the visual channel was not as overloaded as it was in the first group, reaction times decreased. These findings are all in line with the assumption of Mayer (2001) with regard to modality effect principle of the cognitive theory of multimedia learning. In short, Brünken et al. (2002) provided evidence that secondary task processing in the audiovisual multimedia design was significantly quicker than in the visual-only presentation (as measured by reaction times), which shows that simultaneous text and picture presentation leads to higher CL.

Brünken, Plass, and Leutner (2003) state that the benefits of dual-task methodology is superior to others because both the primary task and the secondary task are processed simultaneously, which makes it possible to detect the CL exactly when the load is imposed on the learner. On the other hand, for instance, subjective load scales can only be implemented after the task is done. Whelan (2007) points out that dual-task methodology has certain limitations. He argues that one of the limitations is the risk that the secondary task can interfere with the primary task, thus destroying the performance on the primary task. He further asserts that the intrusion of the secondary task into the first one reaches maximum when the response modality is the same. According to Whelan (2007), the nature and the structure of the primary task other than response modality can affect the usability and efficiency of the second one.

Another problem may be related to the easiness or difficulty of the second task. Kirschner (2002) argues that a learning task, which is comparatively more difficult in general, entails more cognitive resources to be processed, suggesting that difficult tasks are more likely to impose greater CL independent of the modality used. Meshkati and Loewanthal (1988) claim that secondary tasks should not interfere with the primary task, so they should be easy to learn, “self-pacing”, “constant”, and “compatible” with the primary task (as cited in Whelan, 2007, p. 3). Whelan (2007) pinpoints that secondary task performance may be impacted by individual differences. For this reason, he suggests within-subject design for the measurement of CL, because it does not involve error due to differences among individuals from different groups. Finally, secondary task methodology can lead to increasing task difficulty because of the two tasks: In such a case, people might compensate for the increasing difficulty of the tasks by increasing their mental effort, which eventually will affect their performance. In this scenario, the performance of participants might not change and stay stable because of the increased mental effort despite the increased mental load of the task.

A second paradigm, physiological measures of CL, uses such metabolic mechanisms as “breathing rate, pupil diameter, pulse rate, brain waves, respiration content, auditory canal temperature, voice pattern, endocrine, galvanic skin response”. Paas (1993) used heart-rate as a measurement tool of CL and found that correlations between heart rate and participants’ performance level on divergent conditions of CL were significant but too low (as cited in Whelan, 2007, p.3). As such, it was impossible for Paas (1993) to conclude that heart rate was a reliable measure of CL. With respect to physiological measures, Brünken et al. (2003) warn that physiological measures are vulnerable to both environmental and personal factors like stress or emotional mood independent of the primary task.

The third paradigm used to gauge CL is subjective CL scales completed by learners after the task execution. Zhang and Luximon (2005) argue that subjective

CL or mental workload measures appear to be best compared to other paradigms. This claim depends upon a comparison of CL measurement paradigms on several criteria: validity, sensitivity, diagnosticity, intrusiveness, repeatability, selectivity and convenience. A short description of these criteria might be necessary here: Validity refers to the extent to which a measure can gauge the attribute in question. Sensitivity is the power of a measurement tool to detect changes in the difficulty of a given task. Diagnosticity is the capability of discriminating the levels/amount of CL imposed on different cognitive resources. Intrusiveness is the extent to which a load index interferes with the primary task. Repeatability or reliability is the ability of a measure to give the same workload result on different executions of the same test. Selectivity is the ability of a measure to be sensitive solely to changes in cognitive demands of the task. Finally, convenience is the practicality of a measure such as cost and equipments used.

Luximon and Goonetilleke (2001) state that physiological measures are high in cost, highly intrusive, poor in diagnosticity, and lack repeatability even though they use observable ways of detecting CL. Zhang and Luximon (2005) argue that physiological measures are prone to environmental effects like temperature, noise, and lighting and to personal internal states like emotions and illness. According to Zhang and Luximon (2005), secondary task measures performance rather than CL or mental load, thus lacking validity. In addition, the researchers highlight that secondary task measures are highly intrusive, since learners have to do a second task that can easily interfere with the performance of the primary task. Zhang and Luximon (2005) claim that subjective rating scales are important because of: 1) their ease of use; 2) non-intrusiveness; 3) low cost; 4) high face validity; and 5) high sensitivity to changes in workload (p. 201). Zhang and Luximon (2005) discuss

National Aeronautics and Space Administration - Task Load Index (henceforth NASA-TLX) and SWAT (Subjective Workload Assessment Technique) in regards to the seven criteria stated on the previous page. The researchers conclude that subjective workload measurements are not suitable for simultaneous load measurement, since they may become intrusive.

Burkes (2007) state that dual-task methodology may not yield good results of CL estimation since learners may bypass increasing task difficulty (mental load) by increasing their mental effort. Therefore, she suggests that it is necessary to measure learners' mental effort to index CL. She also proposes Paas' (1992) CL rating scale as a good measure of CL because of its reliability and sensitivity. Although people's self-rating of CL may be questionable, Gopher and Braune (1984) pinpoint that people are quite able to indicate mental load they experience numerically (as cited in Burkes, 2007, p. 18). Paas (1992) state that Bratfisch, Borg, and Dornic (1972) found a Spearman rho of 0.9 between subjective and objective measures of task difficulty (p. 429). After all, as O'Donnel and Eggemeier (1986) argue, subjective measures are practical to conduct and obtain data, "nonintrusive", "easy to analyze", and "have very high face validity" (as cited in Paas, 1992, p. 430).

Whelan (2007) argues in his discussion of subjective workload measures that these measures have also their own problems. For instance, they can be affected by the situation or context in which they are implemented as much as by task difficulty. Meshkati and Loewenthal (1988) point out that subjective workload measures may be prone to individual differences and individual states during the time of testing (as cited in Whelan, 2007, p. 4).

Upon arguing that all knowledge depends upon subjective experience, Annett (2002) states that differentiation between subjective and objective knowledge is not that much clear-cut. He further asserts that objective measures used in the scientific arena are not totally independent of subjective experience or subjectivity coming from either the participant or the experimenter. He concludes that "the choice of measures, whether subjective or objective, always has to be justified in terms of the specific aims of the investigation" (p. 984). In light of Annett (2002), the present study also discusses test-retest reliability, internal consistency and convergent validity of the subjective CL scale used.

Working Memory

Information first should be processed through WM. WM is the cognitive network where present cognitive activity takes place and is severely limited (Miller, 1956; Simon, 1974; as cited in Carlson et al., 2003, p. 629). Information successfully processed in WM is transferred to and stored in the long term memory that has an enormous capacity with no known limits.

Baddeley (1992) defines WM as a brain system that does the job of temporary storage and manipulation of upcoming information needed for such complex cognitive phenomena as learning and language comprehension. In his theory of WM, Baddeley (1992) divides WM into three subparts: The central executive directs attention and the other two slave systems. The phonological-loop processes speech-based information like spoken language. The visuo-spatial sketchpad is responsible for processing visual information such as pictures and diagrams. This theory assumes that the two slave systems, the phonological loop and the visuo-spatial sketch pad, work independently of each other and that the visuo-spatial sketch pad and the phonological loop have a limited capacity of processing information at one time.

Thus, a further assumption of the theory is that presentation of the same information to both of the slave components of WM may facilitate processing of information.

Baddeley and Gathercole (1993) state that the work of Daneman and Carpenter (1980) have contributed widely to the idea that "general-purpose cognitive resources" that are backed up by WM are the main cognitive structures that are responsible for language comprehension, be it oral or written (p. 222). Baddeley and Gathercole (1993) pinpoint that the paradigm of Daneman and Carpenter (1980) has three main principles: 1) The first principle assumes that language comprehension consists of both processing and storage functions. Processing includes recognizing words, accessing their semantic and syntactic properties, and interpreting propositions. The representations created by the processes need to be stored, as they are needed for further understanding of language; 2) The same limited cognitive resources implement both processing and storage, so a tradeoff is inevitable whenever a language task exceeds the limited cognitive capacity; 3) There are individual differences among people and these differences stem either from variation in cognitive capacities of human beings or from the efficiency with which cognitive functions are executed.

Daneman and Carpenter (1980) developed a reading span test, which they interpret as measuring WM capacity. According to Baddeley and Gathercole (1993), the reading span test of Daneman and Carpenter (1980) is different from other span tests in that it measures not only storage capacity but also processing capacity of WM system. In their study, Daneman and Carpenter (1980) provided participants with sentences and required to read them aloud. The sentences were presented on cards in sets of two to six sentences. At the end of each sentence set, the participants were asked to recall the last words of the sentences presented. The presentation of a blank card was signaling to participants that they were to write down the last words from the sentences they had read. The span was the maximum number of sentences on which the participants were successful in doing the task. In several studies of theirs, Daneman and Carpenter (1980) were able to establish the validity and reliability of their reading span test. According to Baddeley and Gathercole (1993), the reading span task differed from other spans tasks like digit span tasks, simply because the reading span task promoted not only storage but also processing cognitive functions of the human cognitive architecture.

Turner and Engle (1989) showed that reading comprehension correlated with complex span tasks- tasks that include both processing and storage- but not with simple word and digit spans. According to Turner and Engle (1989), the processing part of the complex span tasks may prevent participants from employing memory strategies like rehearsal, thus leading to a more pure measure of WM capacity. By comparing the correlation of four different complex span tasks with reading comprehension and finding that they all have a significant relation with reading comprehension, Turner and Engle (1989) claim that the processing component of a WM battery does not have to be reading and that the primary task (in this case reading comprehension) does not need to be the same as the skill in question (in this case reading). In light of these findings, Turner and Engle (1989) point out that WM capacity is not task dependent. Besides, Turner and Engle's (1989) study also shows that both sentence-word (complex reading span) and operation-word (complex operation span) correlated significantly with reading comprehension in the same manner when participants were exposed to processing reading and operation tasks that differ from each other in terms of difficulty. Interestingly enough, the size of the correlations between complex span tasks and reading comprehension depended upon whether the items that were to be remembered were words or digits. Namely, the correlations of sentence word and operation word span tests with reading comprehension were higher than those of sentence digit and operation digit. Finally, the authors argue that learning from a text may be difficult because of the limited processing capacities of readers.

Waters and Caplan (1996) criticize WM capacity estimation by just taking the maximum number of words correctly recalled, since this does not take into account the possibility that there could be a trade off between processing and recall parts of a WM span test. Furthermore, they highlight that the method of reading like the one used in Daneman and Carpenter (1980) does not necessarily lead to processing, because reading aloud does not guarantee sentence processing without including semantic or syntactic acceptability judgments. Waters and Caplan (1996) made use of a computerized reading span task to which they incorporated sentence acceptability judgments. At the end of the study, the correlations between reaction times for correct sentence judgments and wrong recalls and those between reaction times and judgment errors were found to be negatively significant. This suggests that most of the participants enrolled in the study were really trading off between processing (accuracy judgment) and storage (recall). The investigators then turned each measure scores into *z* scores and took the average of it as a composite WM capacity index. Test-retest reliability scores (0.75 for cleft subject sentences, 0.83 for subject object sentences) were higher than that (0.41) of Daneman and Carpenter (1980). Waters and Caplan (1996) conclude that sentence processing (reaction time and sentence judgments in their studies) prevents trade offs between processing and storage functions of WM.

Conway et al. (2005) state that reading span tasks are designed to tap both storage and processing functions of WM. Conway et al. (2005) criticize all WM span tasks including the reading span on some general points. For instance, they argue that reading is not necessary for measuring WM capacity, which is in line with the findings and arguments of Turner and Engle (1989). Instead, an operation task like one that entails solving mathematical problems may be a reliable measure of WM capacity. Despite all their criticisms, Conway et al. (2005) conclude that all types of span tasks correlate significantly with each other and provide a plausible measure of WM capacity. The investigators state that when carefully designed and implemented the span tasks are highly reliable and valid.

Kane, Conway, and Engle (1999) argue that such secondary tasks as sentence processing in reading span tasks are good estimates of WM capacity because of the divided-attention to which they lead. The investigators further claim that divided-attention is a desirable part of the WM reading span tasks because it strengthens the predictive validity of a reading span task. Moreover, according to Kane et al. (1999), people's cognitive capability to keep mental representations in the cases of attention shifts or distractions is the factor that constructs a relationship between WM capacity and other cognitive activities. Engle (2002) claims that reading span tasks and other WM tasks should embrace some aspects of cognition or attention-demanding components, because performance on those tasks are closely related to other higher-order tasks and performance on cognitive activities.

The present study makes use of a reading span task to measure WM capacity of the participants. The test consists of twenty sets of English sentences including the number of sentences changing from two to five. The processing part of the test requires participants to indicate whether the sentence they have just read was grammatical or not. The storage aspect requires participants to remember and write down the last words of the sentences they have read in the order they remember. In light of the review stated above, the use of a reading span task in the present study is plausible.

Working Memory and L2 Comprehension

In their a study that investigates the relationship between L2 WM capacity and reading skill, Harrington and Sawyer (1992) claim that reading span tests are a good measure of WM capacity. Harrington and Sawyer (1992) test the assumption that L2 WM capacity and L2 reading comprehension are related. The researchers also made use of both L1 and L2 digit and word span tests to see whether they were related to one another or not. The results of the study indicated that L2 digit and word span test scores did not correlate significantly either with reading ability or with L2 reading span. On the other hand, L2 reading span scores reached a significant correlation level with L2 reading measures used in the study. On the basis of these findings, Harrington and Sawyer (1992) conclude that L2 digit and word span test have little relationship to L2 reading skill as measured in their study. As for L1 reading span test, there was no significant difference between it and L2 reading span test scores. However, both L1 digit and word span test scores were significantly higher than their L2 counterparts.

According to Harrington and Sawyer (1992), reading span scores are reliable indexes of individuals' WM capacity as shown by the significant correlation coefficients found between the measures of the two. The investigators make the point that L1 reading span and L2 reading span can be related significantly as shown by the significant correlation coefficient in their study, but warn that the correlation was a moderate one not large enough to interpret. So, L1 advantage present in L1 digit and word span tests compared to L2 digit and word span tests did not exist in L1 reading span test compared to L2 reading span test. This led Harrington and Sawyer (1992) to suggest that what is called L2 WM capacity may not be different from L2 proficiency. In such a case, the researchers argue that L2 reading span test may prove to be a reliable measure of L2 proficiency. The point to keep in mind about Harrington and Sawyer (1992) is that participants were Japanese-English bilingual students.

Osaka and Osaka (1992) examined whether WM capacity was language dependent. L1 Japanese speakers of English were given: Daneman and Carpenter's (1980) reading span test, its Japanese version, and English as a second language version. Like Harrington and Sawyer's (1992) study, Osaka and Osaka (1992) required participants to read sentence sets aloud from cards and orally recall sentence-final words of each set. In Osaka and Osaka (1992), the reading span tests did not include a sentence judgment component. The span score of the participants was the highest number of last words they could consistently recall for at least three sets. The investigators found out significant correlations between scores on the Japanese and English reading span tests, as well as between the Japanese version and Daneman and Carpenter's (1980) reading span test. Likewise, Osaka, Osaka, and Groner (1993) reported that there was a significant correlation between L1 German/L2 French participants' scores on German and French reading span tests. Thus, the researchers conclude that WM capacity is independent of language.

Walter (2004) argues that the transfer of L1 reading comprehension skills is related to mental representations of a text in L2 and to L2 WM capacity. The participants in Walter (2004) study were French learners of English. In contrast to Harrington and Sawyer (1992), participants of Walter (2004) were not bilinguals.

They were divided into two proficiency groups: lower-intermediate and upper-intermediate. Correlation analyses conducted to test the relationship between WM and reading comprehension yielded significant English and French correlations for both lower-intermediate and upper-intermediate participants. This means that higher WM capacity (both L1 and L2) lead to better reading comprehension for both lower-intermediate and upper-intermediate groups. This was significantly more the case for the lower-intermediate group, since the correlation coefficient of the group was larger than that of upper-intermediate group. English WM scores correlated more significantly with L2 reading comprehension than those of French. This shows that L2 verbal WM capacity is a better predictor of L2 reading comprehension than L1 verbal WM capacity.

Chun and Payne (2004) investigated the relationship between WM capacity differences and L2 German reading comprehension and vocabulary acquisition. As measures of WM capacity, Chun and Payne (2004) used a non-word repetition task and a version of Daneman and Carpenter's (1980) reading span task. Chun and Payne (2004) argue that the non-word repetition task measures the capacity of learners to temporarily store and maintain sound information while the reading span test gauges the executive function of WM. The researchers categorized participants as "high" and "low" WM groups on the basis of a median split procedure on reading span test scores. Chun and Payne's (2004) study yielded no significant differences between "high" and "low" WM groups- in both WM tests- in terms of any of the reading comprehension and vocabulary measures.

Leeser's (2007) study investigating both individual and common effects of topic familiarity and WM capacity found that WM capacity was indeed a significant variable on reading comprehension but under certain familiarity conditions. In this study, learners with medium and high WM earned higher comprehension scores only when they were familiar with the topic of the reading texts. In other words, in Leeser's (2007) study, high and medium WM learners outperformed low WM learners only if they were familiar with the reading topic. Leeser (2007) made use of a computerized version of Waters and Caplan's (1996) reading span test that measures WM capacity as a composite score of reaction times, sentence judgments and words correctly recalled. Leeser (2007) concludes that WM capacity plays a central role in learners' reading comprehension depending upon their previous knowledge about reading topics. One thing that bears to be kept in mind is that both Leeser (2007) and Chun and Payne (2004) made use of WM capacity measures in L1. So, both of the studies looked at the relationship between L1 WM capacity and L2 reading comprehension.

Conclusion

This chapter presented an overview of the historical basis of the theoretical issues related to L2 reading comprehension in a multimedia environment as well as a review of major findings from studies conducted on CL, WM, and their effects on L2 comprehension. It also discussed the research on the measurement of CL and WM capacity. In addition, previous research agenda suggests that multimedia presentations of L2 texts may be useful in making L2 texts easier-to-understand and quicker-to-deal-with depending upon such variables as CL and WM.

Studies on the effects of CL on L2 reading reveal that both instructional design through which the reading text is exposed to the learner and in-text characteristics like types of annotations play a significant role on L2 reading

comprehension, which is directed by the amount of CL they lead to.
Moreover, most research from the same agenda claim that not only presentation characteristics but also learner characteristics like expertise affect performance on reading comprehension due to their CL-decreasing effects.

Research on the interaction between WM capacity and L2 reading comprehension found significant relationship between the two. However, these interactional effects appeared to depend upon other variables such as prior knowledge. Even though the studies exploring any effects of L2 learners' WM capacity and their L2 reading comprehension performance led to an inconsistent body of findings, they pointed out one common factor directing the effects of WM capacity on L2 reading comprehension: prior knowledge. In addition, it is important to note here that research studying the relationship of WM and L2 reading comprehension, used different types of span tasks ranging from simple spans to complex spans and from span tasks in L1 to those in L2.

Furthermore, multimedia learning research that is based upon Mayer's (2001) cognitive theory of multimedia learning showed that different multimedia instructional designs carry different levels of ECL that eventually affects students' understanding of a text. The theory provides design principles that lead to better understanding of learning materials and suggests an information presentation through both visual and audio channels by eliminating extraneous or un-needed CL. The research body that is based upon Mayer's (2001) cognitive theory of multimedia learning and Mayer's (1997) generative theory of multimedia learning tested the principles of the theories mostly in one's native language.

To conclude, in order to see clearly the effects of multimedia presentation mode of an L2 text, prior knowledge, L2 WM capacity and their interactions with one another on L2 reading and listening comprehension, further studies are warranted. The current study tries to address some of these issues. The methodology of the present study and the research questions are presented in the next chapter.

CHAPTER 3

METHODOLOGY

This section explains the methods and procedures implemented in the present study. In what follows, research questions and related hypotheses are presented followed by independent and dependent variables. Second, descriptions of participant characteristics and materials made use of to collect data are provided. Third, a brief report on the pilot study the researcher conducted is presented. Finally, a "thick" description of data collection and analysis are stated.

Research Questions and Hypotheses

The present study investigates the effects of extraneous cognitive load (ECL) and L2 verbal working memory (WM) capacity on the reading and listening comprehension of an English expository passage in a computerized multimedia environment under the conditions of low prior knowledge and high intrinsic cognitive load (CL). In other words, these two variables were constants in the experimental design of the current research. Moreover, another constant was L2 proficiency because all participants were advanced learners of English. One important aspect of the study that should be kept in mind is that participants who read or listened to the text were given the right to re-read and re-listen to it carefully as many times as possible before they thought that they were ready to take the tests. Reading or listening time was not limited so as to stimulate a real-life like task, since while involved in negotiated interaction, L2 learners can control the rate of processing by stopping the conversation and asking for repetitions, clarifications etc. In addition, participants in the pilot study spent a minimum amount of four minutes on reading the text and a minimum amount of five minutes on listening to the text. As a consequence, if, for instance, reading or listening time had been limited to one time, this would have led to a floor effect on participants' comprehension performance. By re-reading or re-listening, learners were able to control the pace of processing. Moreover, VanPatten (1996) argues that:

> (...) attentional capacity (...) is also affected by task demands of which processing time is a significant variable. It could be that even with simplified input time pressure to comprehend may place sufficient demands on the learner such that attentional resources cannot keep up with the demands of the task. (p. 28) (as cited in Leeser, 2004, p. 605).

The limit of "at most ten minutes" of instructional time was determined on the basis of the maximum time period that took any participant to read or listen to the text in the pilot studies. This is an important aspect of the study, which distinguishes it from most other studies in the literature that asked participants to read or listen only once. This is important simply because how many times participants read or listened to the passage could bypass WM differences among them and the difference of the amount of ECL imposed by presentation modes to a certain extent even though the study expects to find effects of English verbal WM capacity and ECL as stated in the hypotheses of the study below. As such, the current study addresses the following issues:

1. What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on retention of information from an expository multimedia text in the L2?

2. What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on transfer of information from an expository multimedia text in the L2?
3. What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on participants' perceptions of cognitive load involved in the presentation of an expository multimedia text in the L2?

Based on the modality principle of Mayer's (2001) cognitive theory of multimedia learning, Hypothesis 1 assumes that the participants exposed to low ECL will achieve significantly higher scores on comprehension tests (i.e., retention and transfer) than the ones exposed to high ECL over time. The following results are expected out of the first hypothesis:

1. The participants in the N+P (narration with pictures) condition will achieve statistically higher retention scores than the ones in the T+P (text with pictures) condition over time (Hypothesis 1-a).
2. The participants in the N+P condition will achieve statistically higher transfer scores than the ones in the T+P condition over time (Hypothesis 1-b).

Based on the research demonstrating the facilitative effects of WM capacity on language comprehension and other higher-level cognitive skills (e.g., Harrington & Sawyer, 1992; Daneman & Carpenter, 1980; Hambrick & Engle, 2002; Hambrick & Oswald, 2005) and on the studies that investigated modality effect (e.g., Mayer & Moreno, 1998; Moreno & Mayer, 1999) it is hypothesized that there would be comprehension difference between HWM and LWM participants in the low and high ECL conditions. The following are the expected results:

1. Participants with high L2 verbal WM capacity will achieve statistically higher retention scores than the ones with low verbal WM capacity over time (Hypothesis 2-a).
2. Participants with high L2 verbal WM capacity will achieve statistically higher transfer scores than the ones with low verbal WM capacity over time (Hypothesis 2-b).

Based on the modality principle of Mayer's (2001) cognitive theory of multimedia learning and cognitive load theory (Sweller, 1988, 1994), and on the research showing that individuals can accurately rate the amount of CL to which they were exposed (e.g., Paas, 1992; Tabbers, 2002; Burkes, 2007), the hypothesis 3 assumes that presentation mode will affect participants' perception of the amount of CL imposed upon them. In addition, since CL is actually the load on WM capacity it is further hypothesized that WM capacity will affect participants' CL ratings on the subjective CL scale. The following outcomes are expected:

1. Participants in the N+P condition will report statistically lower CL ratings than the ones in the T+P condition over time (Hypothesis 3-a).
2. Participants with high WM capacity will report statistically lower CL ratings than the ones with low WM capacity (Hypothesis 3-b).

Finally, data will be tested for any time effects on retention and transfer comprehension scores, and on participants' perception of CL; however, there are no specific hypotheses on this issue.

Variables and Constants

Independent Variables

Extraneous Cognitive Load

This is a two-level categorical variable consisting of high and low categories realized through combinations of multimedia information. It was assumed that T+P condition would involve high ECL while N+P condition would involve low ECL based on Mayer's (2001) cognitive theory of multimedia learning.

N+P condition reduces ECL by allowing information to be processed through the visual and auditory channels simultaneously. In the T+P condition, however, the visual channel is overloaded because both verbal and visual information is processed in this channel.

L2 English Verbal Working Memory Capacity

This is a continuous variable measured through a reading span test. The variable was turned into a categorical variable with two levels: 1) low WM; 2) high WM.

Time of Measurement

This is a two-level categorical variable with two categories on the basis of the time of comprehension tests (immediate versus delayed).

Dependent Variables

Retention of Information

This is a continuous variable gauged through comprehension questions developed by the researcher.

Transfer of Information

This is a continuous variable measured through comprehension questions developed by the researcher.

Perception of Cognitive Load

This is a continuous variable measured through a subjective CL scale developed by the researcher.

Constants

Prior knowledge (low), intrinsic cognitive load (high), English proficiency (advanced).

Participants

5 freshman (17.2 %), and 24 junior (82.8 %) students who enrolled at the Department of Foreign Language Education, Boğaziçi University, participated in the study. Full participation in the experiment was rewarded with 5% of homework grade in two

courses and 5% of total course grade in one course, which aimed at preventing the loss of participants.

Participant Characteristics

The participants were asked to fulfill the participant profile (see Appendix A) before they were exposed to the treatment. The participant profile required participants to provide information about participants': name, age, gender, L1, department, how many semesters they spent at preparatory class, rating use of world wide web (henceforth WWW), rating use of the computer, proficiency, year of undergraduate study, the number of undergraduate courses taken, the number of academic semesters spent at undergraduate study before the time of testing.

The participants were rather homogeneous in terms of their L1 and educational background. All the participants passed the Boğaziçi University English Proficiency Test (hereafter BUEPT). In order to start with freshman year courses in their department, students were required to pass BUEPT by a minimum score of "C" (equivalent of 60). 7 students (24.1 %) passed BUEPT with a score of C, and 22 students (75.9 %) with B. Moreover, participants had taken twenty- two undergraduate courses and spent four semesters on average at their undergraduate study. As such, the participants could be considered advanced learners of English.

Treatments

The treatments given in the study were multimedia presentations prepared in an html format by using Microsoft FrontPage. They were presented on one single webpage. The treatment was an expository L2 passage on tornadoes presented through two different types of multimedia conditions. The text subsumes 254 words and 26 affirmative sentences. The text was an authentic text taken from the BBC "science and natural disasters" webpage. The speeds of tornadoes that were expressed in "mph" (miles per hour) in the original text were turned into their equivalent "kms" (kilometers) on Microsoft Excel in order to prevent any misunderstanding on the part of the participants. Both the narration and html presentations were learner-paced. Namely, the participants were allowed to go forward and backward by scrolling down and up. Mayer and Chandler (2001) claim that computer-paced presentations lead to higher ECL than self-paced presentations, since computer-controlled presentations do not allow learners to stop or to go back to previous information. A self-paced presentation was also chosen to better stimulate a real-world situation: Dixon (1991) argues that in order to bridge the discrepancy between theory and practice, investigations should be conducted in real-life situations and they should be based upon variables important to real-world learning environments. Moreover, Moreno (2006) states that most of the experiments are one-time laboratory learning experiences, and questions whether the results could be generalized to authentic learning environments, thus implicating the use of authentic learning situations in modality effect studies. Thus, the current experiment was carried out in an L2 learning computer laboratory. A high element interactivity text was chosen because previous research (e.g., Leahy et al. 2003; Sweller, 1994) suggests that ECL may not retard learning under low intrinsic CL, since low element interactivity reduces intrinsic CL which in turn reduces overall CL. Therefore, there were two different conditions in the present study both of which were learner-controlled: high ECL (T+P) versus low ECL (N+P).

Text + Picture Condition

This condition included a reading text on tornado formation with corresponding static pictures. ECL was reduced to a certain extent by placing corresponding sentences and pictures as much close to each other as possible based on spatial contiguity principle of Mayer's (2001) cognitive theory of multimedia learning by not leading to any redundancy effect. It should be noted that if the text or some part of it had been integrated into the passage, this would have made the narrated text redundant in the N+P condition as suggested by Leahy et al. (2003). Leahy et al. (2003) posit that how and when aural information is presented is important to the effectiveness of multimedia instruction, which implies that in experiments comparing audio-visual presentations with visual only presentations, neither auditory material nor visual material should be understandable in isolation. The first experiment of Leahy et al. (2003) in which auditory and visual presentations were not understood separately, yielded that audio-visual presentation lead to better learning when compared to its equivalent visual-only presentation. However, in the second experiment whose design includes aurally presented non-essential text with similar written text included in a diagram, the researchers found that there was a redundancy effect because of the aurally presented non-essential text that hindered learning. In the present study, the corresponding pictures used in both T+P and N+P conditions cannot be understood in isolation. In other words, participants could not learn tornado formation solely by looking at pictures. Therefore, pictures in the current experiment served to promote understanding of both visually and aurally presented text by helping participants to build a pictorial model through their visual channel. In addition, treatment text or some part of it and corresponding pictures were not integrated visually in the N+P condition of the present study. Because of equivalency issues between the treatment groups, text and pictures were not integrated in T+P condition either, which violates the spatial contiguity principle to some extent inevitably. The rationale was that the participants in the N+P condition may not have needed to listen to the redundant narration simply because they could have understood the content of the text by only examining the integrated visual information.

Pictures were placed right above the corresponding sentences (see Figure 5):

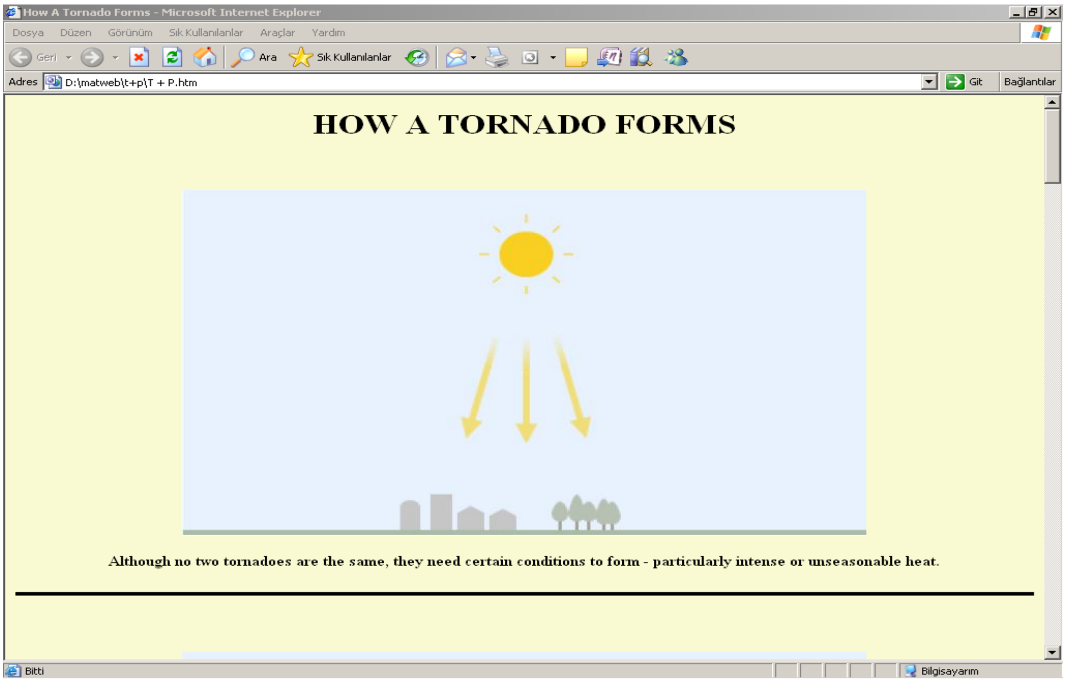


Fig. 5. A screenshot from T+P condition

Participants under the T+P condition were given at most ten minutes to read the text. They were allowed to read as many times as they needed. Furthermore, they could end reading the text any time before the allotted period of ten minutes is over.

Narration + Picture Condition

In the N+P condition, narration and corresponding static pictures were presented on a webpage. The audio material was presented through winamp sound player. The winamp control panel was not embedded into the webpage. Therefore, participants had to control both the website to view static pictures of tornado formation and the winamp panel to listen to the text. This does not mean that they had to control both the webpage and winamp control panel at the same time. However, due to its nature, this design did not allow learners to go back and forth sentence by sentence, thus making them wait until the end of the narration. That is also why a short text was chosen in order not to increase ECL unnecessarily. More specifically, participants in the N+P condition had to synchronize the narration and the corresponding pictures by themselves. Moreover, the presentations of both narration and pictures were learner-paced so that the participants could go back and forth upon their need. Moreno and Mayer (2002) pinpoint that text with audio presentations may lead to greater learning on the part of learners compared to audio alone presentations, which is in contrast to redundancy principle. On the other hand, Kalyuga, Chandler, and Sweller (2004) showed that audio alone presentations resulted in better learning outcomes. Clark et al. (2005) argue that the contradictory research findings in the literature might be attributed to the segmentation of the audio material, and state that a long audio or listening material may be challenging to retain and process in WM and even more difficult to coordinate with relevant visual material (as cited in Burkes, 2007, p. 42). Similarly, Kalyuga et al. (2004) argue that the inconsistency between their result and that of Moreno and Mayer (2002) can be explained by the segmentation of the textual information in their study: Kalyuga et al. (2004) presented the text continuously to participants as a single unit without any breaks. On the other hand, in Moreno and Mayer's (2002) study, the text was presented as consecutive smaller segments with breaks between them. This could have given participants of Moreno and Mayer (2002) enough time to construct partial mental representations from each text segment, thus reducing working memory load (Kalyuga et al., 2004, p. 579). Therefore, a short listening audio record (one minute, fifty-five second long) was used in the present study. This prevented increasing ECL unnecessarily. There was no pause inserted at any sentence, clause, and phrase boundaries. As a whole, the listening passage was delivered at approximately 127 words per minute. At most ten minutes were allocated for the participants in the N+P group and they were allowed to listen to the text as many times as they wished. In addition, they could end listening to the text before ten minutes of instructional time were over. Recording was actualized by an MP3 recorder program called "MP3 recorder XP". Besides, a native speaker of English from the U.S.A was hired to do the recording (see Figure 6):

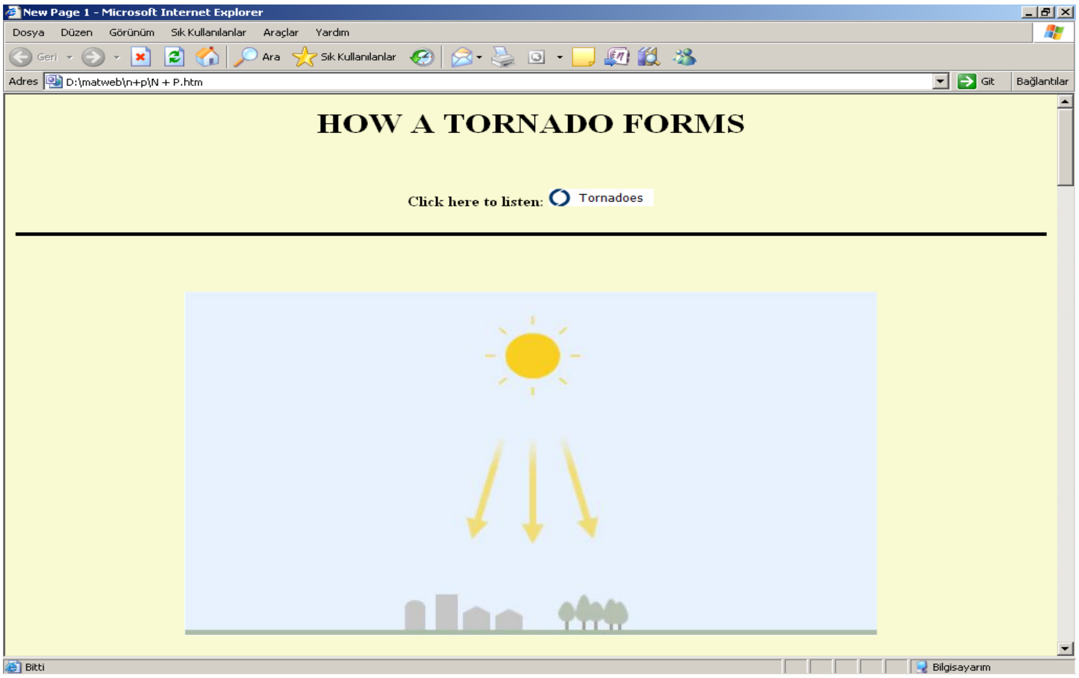


Fig. 6. A screenshot from N+P condition

There were four treatment groups in the present study: high WM participants in the N+P condition (henceforth N+P_HWM), low WM participants in the N+P condition (from now on N+P_LWM), high WM participants in the T+P condition (referred to as T+P_HWM from now on), and low WM participants in the T+P condition (referred to as T+P_LWM henceforth).

The Pilot Study

A pen-and-paper version of the text with pictures was piloted with thirteen students who did not participate in the original experiment in order to determine the feasibility of the reading text, how much time could be allocated for each section of the investigation, and to see whether there could be any "floor effect" because of: difficulty level of the text and the comprehension questions, wording, unknown vocabulary items, and concepts or ideas that are difficult to interpret. Upon suggestion of one of the participants the wording of the second retention question was changed (why/why not? was turned into why/how can this happen?). No unknown vocabulary and unclear concept or idea was reported by the participants. The Shapiro-Wilk normality test, Q-Q plots and histograms showed that both retention and transfer scores of the participants were normally distributed, which also held true for CL ratings. There was a score range of four to eight on the retention test, a range of two to ten on the transfer test, and a range of nineteen to twenty-seven on the CL scale. These suggest that there was no floor effect in retention and transfer performance of the participants. The participants in the pilot study of the pen-and-paper version of the text were asked to read the text in order to comprehend it, since they would be given comprehension questions. They were not given any time limit and it took them eight minutes (maximum) to read the text, three minutes (maximum) to complete the CL scale, ten minutes (maximum) to answer the retention questions, and nine minutes (maximum) to answer the transfer questions, thus making a total of thirty minutes to complete the whole task at most.

As for the narrated text with pictures, the audio record was piloted with nine advanced learners of English who did not participate in the original experiment to detect its understandability (in terms of accent, speed, volume etc.). The participants in the piloting of the listening text were asked to listen with a purpose of understanding and told that they would take comprehension tests afterwards. The participants spent seven minutes (maximum) on listening to the text, four minutes (maximum) on completing the CL scale, thirteen minutes (maximum) on answering the retention test, and eleven minutes (maximum) on answering the transfer test (thirty-five minutes in total). The Shapiro-Wilk normality test, Q-Q plots and histograms showed that the dependent variables (retention, transfer scores, and CL ratings) in the second pilot study were normally distributed with the exception for CL ratings. Besides, there was a score range of three to eleven on the retention test, a range of two to nine on the transfer test, and a range of twenty to thirty-five on the CL scale, thus eliminating any floor effect. The participants in the second piloting did not report any problems either with the audio records or with the comprehension questions and the subjective CL scale. Finally, the other data collection instruments "participant profile, prior knowledge test, and subjective CL scale" were also piloted with the same participants enrolled in the piloting of both T+P and N+P presentations in terms of wording, ambiguity, unclear concepts/ideas, and unknown English words/phrases. None of the participants requested that the researcher make any change in any of the instruments mentioned above.

Van Merriënboer, Kirschner, and Kester (2003) advise that general supportive information should be provided first so that learners can construct a schema to be used during the task. Lund (1991) argues that prior to listening activities listeners should be provided with situational information covered in the context of the listening activity to enable listeners to pay attention to the content. He goes on saying that "the visual aspect of videos provides more, but not necessarily all, of the relevant context" (p.202). For this reason, the present study includes two supportive information sources: 1) In the form of a pre-reading/listening activity, every condition has a reading section consisting of interesting information and myths about tornadoes such as opening the windows will lessen the damage of a tornado (see Appendix B); 2) prior to testing, a video on a real tornado (a one minute and twenty-two second-long video) was presented to all participants (see Appendix C).

Data Collection

For data collection, two sets of instruments were used. The first set consisted of a WM test, and comprehension tests that were used to answer the research questions. Second, just before participants were given the treatment a prior knowledge test was implemented to determine whether participants differ in terms of prior knowledge about tornadoes.

Verbal Working Memory Test

An English reading span test was conducted for all participants as the verbal WM test before they took the treatment. The test consisted of seventy affirmative English sentences run on the Superlab computer program. In the computerized L2 WM test, the participants were shown sentences on the computer screen, one at a time. Each sentence was on the screen for seven seconds. During this time, each participant was required to:

1. Read the sentence and decide whether it was grammatically correct.
2. Try to remember the last word of the sentence.

If the sentence was grammatically correct for the participant, she or he pressed “T” (True) on the keyboard; if not she or he clicked on “F” (False). All the sentences were divided into sets having different sizes ranging from two to five sentences. After a participant saw all the sentences in a particular set, she or he was required to remember the last word of each sentence and instructed to write the last words in a box shown on the screen in the order they remembered. Because the set sizes increased from two to five sentences, the participants were required to remember more words as the test proceeded. There was a practice session before the test started.

Participants’ L2 English verbal WM scores were calculated as follows: 1) each participant’s number of correct judgments (out of seventy sentences) and correct remembering of words (out of seventy last words) was calculated as a total on SPSS; 2) the total scores of the number of sentences correctly judged and the number of words correctly remembered were turned into standardized values (z scores); 3) the mean of the “z” scores were calculated for each participant; 4) participants were divided into high WM and low WM groups by recoding their mean “z” scores through a median split procedure.

The reading span of the twenty-nine participants varied from -.74 to 1.29 (M = .33, SD = .60). The reading span of high WM participants varied from .12 to 1.29

(M = .80, SD = .37) and that of low WM participants from -.74 to .09 (M = -.18, SD = .27). In order to determine whether the groupings of participants into high and low WM groups represented true differences, the mean *z*-scores (reading spans) were submitted to a one-way analysis of variance (ANOVA henceforth) with WM group as the independent variable. The assumptions of normality and homogeneity were true for mean *z*-scores, $p > .05$. In addition, no outliers were detected. There was a significant difference between low WM and high WM groups in terms of composite *z*-scores they got on the WM test ($F_{1,\ 50}$=100.978, $p < .001$, η^2= .657) with a large effect size ($d > .8$). Figure 7 shows the mean reading spans for each WM group:

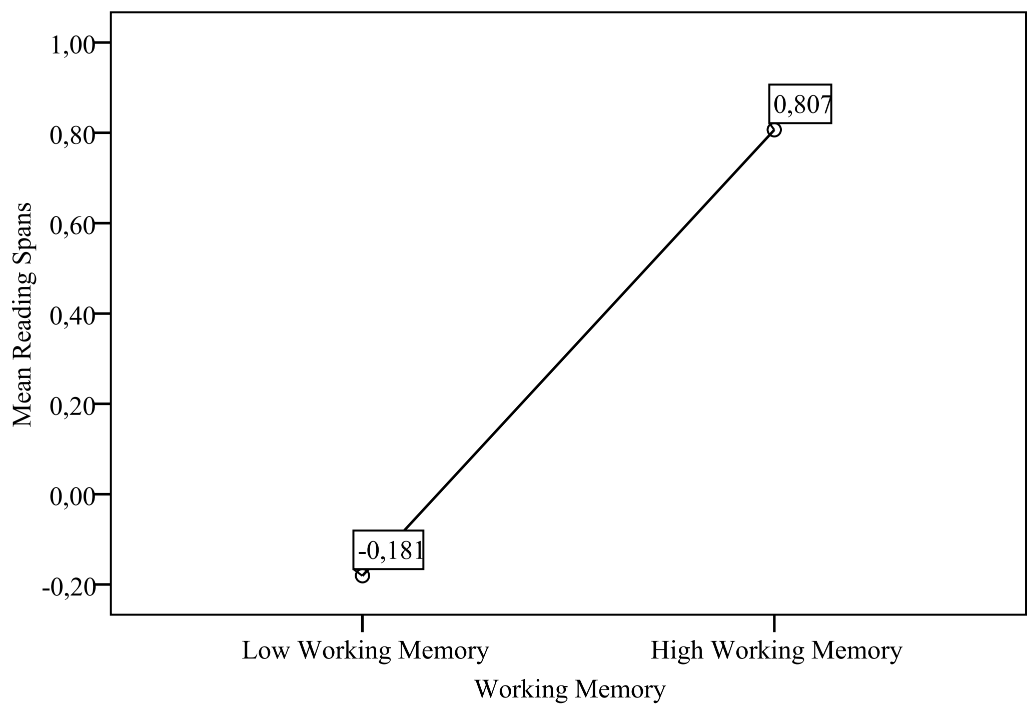


Fig. 7. Mean reading spans for each working memory group

The number of the correct answers of the participants on sentence judgment component of the reading span test ranged from 39 to 67 (M = 56.38, SD = 6.70) and on the last word recall part from 14 to 55 (M = 36.52, SD = 10.40). In order to determine whether participants traded off on one component of the WM test for the other, thereby warranting a composite score, the *z*-scores for each component were correlated. Using Spearman's rank-order correlation (r_s), a positive nonsignificant correlation between sentence processing and word recall was found (r_s = .281, n =29, p > .05).

At the first glance, the positive r_s indicates that the participants did not trade off on sentence processing or word recall for the other one, thus not warranting the use of composite mean *z*-scores as the index of the reading spans of the participants. However, notice that the existing correlation coefficient is neither significant nor large. According to Cohen's (1998) guideline (Pallant, 2001, p.120) a correlation coefficient of .281 refers to a small positive one. Then, it is reasonable to argue that the small positive correlation between sentence judgment and word recall may imply a kind of trade-off between the two components of the WM test than a large positive one. Therefore, composite *z*-scores were used as the English verbal WM index of the participants enrolled in the present study.

Following Daneman and Carpenter (1980), the same WM test was also given to participants in the N+P condition, since Daneman and Carpenter (1980) found that oral and silent reading span tests correlated highly with both listening span test and listening comprehension test. Moreover, Osaka and Osaka (1992) and Osaka et al. (1993) found significant correlations between L1 and L2 WM measures, and concluded that WM capacity is not language dependent. While examining the effects of WM on the comprehension of aurally presented L2 information in the present study, L2 reading span test was used as the index of WM given that previous research found significant correlations between L1 and L2 measures of WM, and that there were significant correlations between L1 oral/silent reading and L1 listening span/comprehension tests. Moreover, Turner and Engle (1989) reported that the secondary task used in complex span measures of WM (reading in the form of grammaticality judgment in the present study) does not have to be the same as the main task (listening comprehension for N+P groups in the current study) on which participants' performance is measured. Specifically speaking, they state:

> (…) These results clearly show that the processing component of the WM span does NOT need to be "reading" related to produce a correlation between the span measures and reading comprehension. (…). (p. 139)

Furthermore, Payne, Kalibatseva, and Jungers (2009) investigate any possible interaction between domain experience and WM in terms of their combined effect on L2 reading comprehension. The investigators used a counting span task in which participants were expected to count the number of dark blue circles surrounded by distracters sharing color and shape aspects on a series of screen displays as the WM capacity index. Furthermore, the counting span task was implemented in participants' L1 English. Payne et al. (2009) state that they preferred the counting span task to reading span task since it depends less on verbal ability, and thus correlations found with L2 comprehension measure would directly be due to updating and maintenance parts of WM, as opposed to shared ability between reading span task and reading comprehension.

Finally, Mayer (2001) argues that both spoken words and written words are turned into a verbal model in WM even though written words start to be processed through the eyes. Besides, according to Mayer (2001), learners can change the processing channel of spoken words and written words. For instance, a learner can create mental images of what s/he is listening to –such as visualizing a cloud and a funnel dropping out when the narration says "A visible cone or funnel drops out of the cloud towards the ground." or s/he can mentally create sounds of what s/he is reading – such as mentally saying “cloud” when s/he views the word “cloud” in the text. Therefore, Mayer (2001) claims that although cognitive processing of written words must initially start in the visual channel, a learner can shift their processing into the audio channel by mentally pronouncing the written information. Then, once written/visual information is shifted into the verbal channel, it is processed in that channel and integrated into a verbal model.

Therefore, it is not unreasonable to use an L2 reading span task as the index of L2 verbal WM capacity and test its effects on L2 listening comprehension, which is the case in the current experiment.

Comprehension Tests

The comprehension tests involved retention and transfer of information from an expository L2 text. The tests were administered immediately after and two weeks after the treatment in order to estimate the immediate and delayed effects of the treatment levels. In order to score the tests, detailed answer keys were prepared by the researcher. Another independent reader volunteered to read and prepare answer keys to the tests. Disagreements between the researcher and the independent reader were resolved by consensus later. In both retention and transfer tests, all the questions were presented on the same sheet of paper, since no questions would lead to any ideas about the answers to the other questions in the test. Since no feedback was given to the participants after both retention and transfer tests, the same questions were again made use of in the second tests implemented two weeks after.

Note-taking was not allowed either during reading and listening to the treatment. The rationale was that note-taking could have compensated for the lack of

L2 verbal WM capacity of the participants with low capacity, thus making it hard or impossible to detect any performance differences between high and low WM participants. In other words, note-taking could have made the main effect of WM capacity undetectable with which the present study is concerned. In the same line of logic, the reading and listening text was not available to the participants during the comprehension tests. This procedure was also based on Pederson's (1986) argument that reading comprehension is more validly measured if participants are not allowed access to the reading text (as cited in Ariew & Erçetin, 2004, p. 247). The reliability of retention and transfer tests was examined by correlating immediate test scores with the delayed test scores. There was a large positive correlation between the immediate and delayed retention test scores (r = .559, n = 29, p < .05), between the immediate and delayed transfer test scores (r = .555, n = 29, p < .01).

<u>Retention Tests</u>

The retention test (see Appendix D) consisted of four questions. The participants were given at most twenty minutes to answer all the questions. Mayer (2001) states that the goal of retention tests is remembering, and defines retention of information as the "ability to reproduce or recognize presented material" (p. 16). As such, the retention test used in the present study consists of questions to promote reproduction and recognition of presented information: The first question asked the participants to explain how a tornado formed. Before the test was given, the researcher determined 7 key steps in tornado formation regarding the first question. While scoring the answers of participants to the first question, the number of the key steps provided in the answer was counted (1 point for each correctly indicated step). This judgment based on not the exact wording but on the meaning of the given answers. Spelling mistakes were ignored during scoring. Unnecessary and wrong information were not scored either. The second question involved matching sample damages with corresponding tornado damage scale. Each correct match was given 1 point, thus the whole question was rewarded with 6 points. The third and fourth questions were multiple choice items (with four choices) each of which was rewarded 1 point. So, the whole retention test was worth of 15 points if all the questions were answered correctly. The third question required the participants to choose the damage scale of a tornado moving with a certain speed. The fourth question required them to choose the damage scale of a tornado that has a specific damage scale level.

The answer key to the retention test (see Appendix D.01) was prepared by the researcher beforehand and checked by another independent reader who was trained on how a tornado forms. Since the second question was a matching item and the remaining two were multiple choice items, only the number of steps involved in tornado formation (question one) were discussed. Disagreements between the researcher and the reader were resolved by consensus. Both immediate and delayed retention tests were scored by another independent rater. Disagreements between scorers were resolved by either consensus or taking average of the marks given by the two separate scorers. Using Spearman’s rank-order correlation, an inter-rater reliability of .934 was found for the immediate retention tests and an inter-rater reliability of .803 was found for the delayed retention tests. Retention performances of the participants were expressed as a raw score out of the total mark “15”.

Transfer Tests

The transfer test (see Appendix D.02) consists of five questions. The participants were given at most thirty minutes to answer the questions. Mayer (2001) defines transfer of information as the "ability to use presented material in novel situations" and states that the goal of transfer tests is to tap understanding (p. 16). The first question was a conceptual question that asked for any cause or causes of a tornado. The first question was worth 3 points because the answer involved three possible causes of a tornado. The second question was a troubleshooting question that presented participants with a possible unexpected scenario and asked for why that would be the case in the light of what participants read or listened. The answer to the second question consisted of two options given one point in the case of each correct answer (2 points as a whole). The third question was a redesign question that required participants to find a solution to a specific problem related to tornadoes. There were two complementary steps to answer this question. Therefore, the question was rewarded with 2 points as a whole. The fourth one was a prediction question that asked participants to find a solution to a process of tornado formation that was not explicitly stated in the treatment passage. There were four possible predictions/answers to this question constituting a whole of 4 points for this question. The fifth question was again a conceptual question asking for the factors affecting the duration and strength of a tornado. There were two possible answers each of which was rewarded with one point. Thus, the whole question was given 2 points if correctly answered. The last question that is a conceptual question asked participants to list environmental clues to watch out for a tornado. Since there were seven clues stated in the passage, the question was rewarded with 7 points in total. Hence, the transfer scores of participants were calculated out of a 20-point total transfer score. The answer key to the transfer test (see Appendix D.03) was prepared by the researcher; another independent reader who was trained on the content of the text checked it. The raters reached agreement about the disputed answers. As with the retention test, answers to the transfer tests were marked by two independent raters. Disagreements were resolved by either consensus or taking the average of the two scores given by the raters. Using Spearman's rank order correlation, an inter-rater reliability of .845 for the immediate transfer test and .912 for the delayed transfer test were obtained. The answers of the participants were scored regardless of their wording or spelling errors, which was also the case in the retention tests. Unnecessary and wrong answers were not scored either.

Prior Knowledge Test

The prior knowledge test conducted in the study included two parts (see Appendix E). In the first part, participants were asked to indicate whether they had experienced a real tornado (together with a when and where question if the answer was "yes") and to write down anything they knew about tornadoes. Each acceptable answer was given 1 point. In the second part, participants were provided with four open-ended conceptual questions on tornado formation. A detailed answer key for the second question in "Part I" of the test and the second part of the prior knowledge test (see Appendix E.01) was prepared by the researcher and checked by another independent reader who was trained on the treatment text. The second question in "Part I" was scored on the basis of the 7 steps involved in tornado formation as indicated in the L2 text. Disagreements between the researcher and the independent reader were resolved by consensus. Spearman's rho was used to measure the inter-rater reliability. There was an inter-rater reliability of .613. The participants got very low scores on the prior knowledge test. The prior knowledge test scores marked by the researcher had a range of 4 with maximum and minimum scores of 0 and 4 respectively (M = .72, SD = 1.16); the range of the prior knowledge test scored by the independent rater had a range of 3 with maximum and minimum scores of 0 and 3 respectively (M = .62, SD = .89). All the questions were presented on the same sheet of paper, since none of them provided answer clues for the other questions. Moreover, because the prior knowledge test was given right before the treatment, it asked for general information about tornado formation process and included different questions from the ones on the comprehension tests in order not to sensitize participants to the tests. In the same line of logic, no feedback was provided to participants after the implementation of the prior knowledge test. The additive raw scores of the two parts were used as the total prior knowledge score of each participant. A two way independent ANOVA with "WM and ECL" as the independent variables was performed on the prior knowledge test scores to determine whether experimental groups differ in terms of prior knowledge. The results revealed no significant main effect either for WM ($F_{1, 25} = 3.372$, $p > .05$) or for ECL ($F_{1, 25} = 1.071$, $p > .05$). Pair-wise comparisons with Bonferroni adjustments confirmed these results. Moreover, the interaction effect between WM and ECL on prior knowledge was statistically non-significant ($F_{1, 25} = .697$, $p > .05$). Moreover, 62 % (18) of the participants received 0, and only 3.4 % (1) received 4. These results indicated that experimental groups did not differ significantly from each other in terms of prior knowledge, which in turn means that prior knowledge is a constant in the study. Moreover, since the highest score on the prior knowledge test is "4", it is reasonable to claim that all participants had “very very low” or “no” prior knowledge before the treatment.

Paas’ (1992) Subjective Cognitive Load Scale

In order to measure the amount of CL involved in the different practice-statistical problem types, Paas (1992) developed a subjective CL scale (see Appendix F). Paas (1992) states that CL is a multidimensional construct, and that it consists of mental load and mental effort. According to Paas (1992), mental load is generated by instructional characteristics including task structure while mental effort is the amount of cognitive resources allocated to meet the instructional demands. This implies that learners try to bypass mental load imposed by instructional demands by investing mental effort. Paas (1992) claims that increased mental load will only be effective if learners are motivated and really invest mental effort. Therefore, Paas (1992) argues that the amount or intensity of mental effort is an index of CL. Paas’ (1992) scale is a unidimensional scale in the sense that it has one item asking people to indicate the amount of mental effort they have invested. The scale is a 9-point scale on which the numerical values and their labels ranged from *very, very low mental effort* (1) to *very, very high mental effort* (9). Paas’ (1992) scale is subjective because participants reported the mental effort they spent by translating the perceived amount/intensity of mental effort into numerical values.

Paas, van Merriënboer, and Adam (1994) showed the reliability and sensitivity of Paas (1992) CL rating scale, and that one-dimensional scales have sensitivity with respect to CL differences and such scales are valid, reliable, and unintrusive. Burkes (2007) claims that the reliability and sensitivity of the scale and the ease of its use “have made this scale, and variants of it, the most widespread measure of working memory load within CLT research" (p. 17). She further states that Paas' (1992) scale was the first one used within the framework of CL theory. Gimino's (2000) study has also demonstrated the reliability, convergent, construct, and discriminate validity of the scale (as cited in Burkes, 2007, p. 18).

Paas' (1992) CL scale was used as the criterion in the convergent validity analysis of the subjective CL scale used in the current study. In order to estimate the convergent validity of the subjective CL scale used in the current study, participants' scores on it were correlated with their scores on Paas' (1992) CL rating scale. Participants' CL ratings on Paas' (1992) CL scale varied from 1 to 8 (M = 5, SD = 2).

Subjective Cognitive Load Scale

The subjective cognitive load scale utilized in the present study was developed by the researcher prior to the time of testing (see Appendix G). The scale was mostly adapted from NASA-TLX work load scale (see Appendix H). The scale consists of eight items on a 7- point scale ranging from "not at all" to "extremely much/difficult", which are represented by 1 and 7 respectively. Hence, the maximum score in a case of maximum CL rating was 56. The scale had been explained and illustrated to the participants before they took it by going through both the explanation section of the scale (above the questions) and the questions, which aimed at minimizing any misunderstanding on the part of the participants. Each item on the scale was aimed at measuring one aspect of the construct of CL: mental effort, mental load, performance. Besides, the items also tried to gauge different types of CL: intrinsic CL, ECL and germane load. Therefore, the scale was multidimensional in that it tries to measure more than one dimension of CL, thus measuring overall CL of the instructional design of the present study not just ECL. The CL score of each participant was calculated by adding each numerical value of the all eight items as based on their ratings in the first administration of the scale. "Item 4" on the scale was a reverse item in terms of its wording, and it was recoded/reversed into another item "Ritem 4" (= item four reversed) before calculating the total CL score of the participants. Participants' ratings on the subjective CL scale ranged from 17 to 41 (M = 27.72, SD = 6.4).

Annett (2002) strongly suggests justification of the use of subjective or objective measures in a scientific study. In this respect, convergent validity, internal consistency and test retest reliability of the CL scale were tested: The convergent validity of the scale was tested by correlating each participant's total CL score on the immediate administration of the subjective CL scale with their total cognitive score on Paas' (1992) CL rating scale. Preliminary analyses were performed to ensure no violations of the assumptions of normality, linearity, and homoscedasticity and to watch out for outliers. The assumption of normality and linearity were violated on Paas' (1992) scale. Consequently, Spearman's rho was used as the correlation coefficient index. There was a moderate correlation ($r_s = .376$, $n = 29$, $p < .05$), with high scores on the subjective CL scale associated with high scores on Paas' (1992) scale. According to Cohen's (1988) guideline (Pallant, 2001, p.120), the correlation coefficient of .376 reflects a medium-strength relationship and indicates an 14.13 % of shared variance between Paas' (1992) scale and the scale used in the present study. The statistically significant r_s between the subjective CL scale used in the current study and Paas' (1992) scale means that the convergent validity of the subjective CL scale is established. It is worthy at this point to state that the range of participants' scores on Paas' (1992) was quite small (7) as compared to that of the subjective CL scale (24).

In order to test the test re-test reliability of the CL scale, the scale was distributed to the participants on two occasions: 1) right after the treatment (after the immediate delivery of the retention and transfer tests); 2) right after the delayed delivery of the retention and transfer tests. Hence, the same CL scale was distributed to the participants second time two weeks after the first delivery. Participants' CL ratings in the delayed administration of the scale varied from 14 to 40 (M = 28.7, SD = 1.42). Preliminary analyses were conducted to ensure that there was no violation of normality, linearity and homoscedasticity, and to check for outliers. Then, participants' scores on the first distribution were correlated with their scores on the second one. There was a strong positive correlation (r = .559, n = 29, p < .01), with high scores on the first administration of the scale associated with high scores on the second one. According to Cohen's (1988) guideline (Pallant, 2001, p. 120), the Pearson's correlation coefficient of .559 reflects a large correlation between the first and second administrations of the scale, which suggests quite a strong relationship between the two. The correlation coefficient of .559 indicates 31.24 % shared variance between the two administrations of the scale. All these mean that the test-retest reliability of the subjective CL scale holds strong.

The internal consistency of the scale was indexed through Cronbach's alpha. The Cronbach alpha coefficient of the scale was .749, which indicates that the scale has sufficient homogeneity. In other words, the items on the scale are sufficiently good at gauging the same construct. Pallant (2001) states that "Ideally, the Cronbach Alpha coefficient of a scale should be above .7" (p.85). However, she warns that the Cronbach alpha coefficient is sensitive to the number of items in a scale and that with scales having less than ten items it is not uncommon to find quite low Cronbach alpha values. In this respect, the subjective CL scale is a short scale since it is made of eight items. Hence, it is reasonable to argue that the Cronbach alpha value of .749 of the subjective CL scale shows that the items in the scale sufficiently measure the same construct. In cases of short scales, Pallant (2001) suggests that "it may be more appropriate to report the mean inter-item correlation for the items" (p. 85). The mean inter-item correlation was .266 for the scale. This value is within Briggs and Cheek's (1986) recommendation for an optimal range for the mean inter-item correlation of .2 to .4 (as cited in Pallant, 2001, p. 85).

Data Collection Procedures

At the beginning of the experiment, the participants were told about the procedure of the experiment and presented with an explanation sheet included in the consent form. The participants were also informed about all the tests including the subjective CL scale. An American native speaker of English volunteered to read and check English of the reading comprehension tests, prior knowledge test, subjective CL scale, and participant profile. Pilot study of a pen-and-paper version of the text, tests, and the participant profile was done with thirteen senior ELT students who did not participate in the original treatment (May 22, 2008). Another pilot study was conducted on the audio version of the text with nine advanced learners of English (June 2, 2008). The participants in the pilot study were asked to report any cases of misunderstanding and ambiguity regarding wording, grammar, and vocabulary. The overall data collection was actualized in three sections. In the first section, participants took the English verbal WM test in the computerized Italian language laboratory of the Department of Foreign Language Education (June 30 - July 4, 2008). In the second section (July 7-

11, 2008), the participants took the treatment and the other tests in the following order: participant profile, prior knowledge test, treatment, retention test, transfer test, Paas' (1992) CL scale, and subjective CL scale developed by the researcher. Three weeks after the WM test, namely two weeks after the treatment and immediate tests, participants took the delayed tests (retention test, transfer test, and subjective CL scale; July 21-25, 2008). In every session, the participants were asked to read or listen to the text in order to understand the content, because they were informed that they would be given comprehension tests later on.

Data Analysis

Although the participants were randomly assigned to four treatment groups, comparison on several control groups were also conducted. Thus, the groups were compared using two way ANOVA with ECL (high-low) and L2 verbal WM capacity (high-low) as between groups factors on prior knowledge, age, hours of internet use per week, hours of use of Word, FrontPage , PowerPoint, Excel, Outlook, Publisher, Dreamweaver, number of undergraduate courses taken before the time of testing, and the number of semesters spent at undergraduate study before the time of testing in each group in order to see whether the groups were different from each other at a probability value of .05 or below. Since two-way ANOVA is a parametric test, the assumptions of ANOVA were checked out by means of the following tools before further running the analyses and these tools were also used to check the assumptions in all parametric tests conducted:

1. Assumptions of independence of observations and random assignment were met with implementing an appropriate design.

2. Normality assumption: This assumption was checked by means of several instruments:

- Skewness and kurtosis coefficients: In the case of a prefect normal distribution, skewness and kurtosis coefficients are equal to zero. Theoretically, skewness is not considered to be extreme if the coefficient of skewness is between -1.0 and +1.0. In the same line of logic, kurtosis is not considered to be extreme if the coefficient of kurtosis turns out to be between -1.0 and +2.0.
- Normal Quartile-Quartile Plots (Q-Q plots): These plots compare observed values against expected values. The fit line in these plots represents the expected values. If the dots representing observed values fall more on the fit line, it can be safe to conclude that the observed values come from a normally distributed data. In other words, the more congruence between the fit line of the Q-Q plot and the line constituted by the observed values, the more normally distributed the data is.
- The Kolmogorov-Smirnov (30 or more participants) /Shapiro-Wilk(less than 30 participants) test of normality: The null hypothesis that the data are normally distributed was tested: The null hypothesis is retained in case of a normally distributed data or rejected if the distribution is not normal. Shapiro-Wilk test of normality was used in the presented study because the number of participants was lower than 30.

3. Homogeneity assumption: Homogeneity of variance between groups was gauged by means of Levene's test.

Since level of measurement for BUEPT scores, rate of WWW use, and rate of computer use were ordinal, while gender and year of study at undergraduate study were categorical variables, Kruskal Wallis test (non-parametric counterpart of one-way ANOVA) with treatment groups as the between-groups variable was implemented on these variables. In order to run the Kruskal Wallis test, a grouping variable named "treatment groups" with four levels (N+P_HWM, T+P_HWM, N+P_LWM, T+P_LWM) was created. The tests mentioned above showed that groups were equal on all variables, p's > .05. Finally, while conducting correlation analyses the assumptions of normality, linearity, homoscedasticity and possible cases of any extreme scores and outliers were checked. In the cases of non-linear relationships the Spearmen rho (r_s) was used as the correlation coefficient index. Variables involved in the correlation analyses were also checked in terms of restricted range. Together with results, coefficient of determination or shared variance was also presented in order to examine the practical significance of statistically significant correlation coefficients. Cohen's (1969) *d* was implemented to gauge effect size for each test conducted (as cited in Coe, 2002).

While determining main effects and combined effects, Bonferroni adjustment was applied in order to avoid inflated Type 1 error rate. Therefore, the adjusted p-value of .025 was the significance level for further analyses as otherwise indicated. Effect size and power indexes were also presented in order to examine the practical significance of statistically significant results. However, since the sample size is small, thus making the population of each experimental group less than ten a p-value of .05 was chosen to escape Type 2 error risk wherever possible by following what Pallant (2001) states:

> (…) Stevens (1996) suggests that when small group sizes are involved, it may be necessary to adjust the alpha level to compensate (e.g., set a cut off of .10 or .15, rather than the traditional .05 level) (p. 173)

Analysis of Research Questions

In the present study, prior knowledge, L2 proficiency and intrinsic CL were constants. In other words, the effects of independent variables were analyzed under the conditions of low prior knowledge, advanced L2 proficiency and high intrinsic CL. Of the three, prior knowledge, and proficiency are learner characteristics while intrinsic CL is related to the nature of the expository L2 text used in the current investigation.

Research Question 1

What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on retention of information from an expository multimedia text in the L2?

In order to investigate possible effects of ECL and English verbal WM capacity on retention of knowledge over time in a multimedia environment under low prior knowledge and high intrinsic CL conditions, a three-way mixed ANOVA with ECL (high-low) and English verbal WM capacity (high-low) as between-groups factors, and time (after treatment, two weeks after) as within-groups factor was conducted on immediate and delayed retention test scores.

Research Question 2

What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on transfer of information from an expository multimedia text in the L2?

Any possible statistically significant difference in overall transfer scores were tested via a 2 (ECL: high, low) X 2 (WM: high, low) X 2 (Time: immediate, delayed) mixed-design ANOVA. ECL and English verbal WM capacity were between-groups factor while time was repeated measures independent variable.

Research Question 3

What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on participants' perceptions of cognitive load involved in the presentation of an expository multimedia text in the L2?

Whether there were any effects of ECL, English WM capacity and time on participants' perception of overall CL created by the presentation mode of the treatment text, a 2 (ECL: high, low) X 2 (WM: high, low) X 2 (Time: immediate, delayed) was implemented on CL ratings of the participants collected by means of a multidimensional subjective CL scale developed by the researcher.

In all these mixed design ANOVA analyses, the assumption of sphericity was ignored since the within subjects variable, time, had only two levels. The assumptions of normality, homogeneity of variance, homogeneity of intercorrelations, and extreme scores including outliers were checked before interpreting the results.

Finally, the assumptions of any parametric test applied were checked in order to determine whether they violated the assumptions. The cases of violations of the assumptions are stated. Otherwise, test results are presented directly.

Summary

In the present study, the learning material “how tornadoes form” is chosen not because of the length of the reading text but because of the high element interactivity it has in its nature. Moreover, no learner had prior knowledge of how a tornado forms before the time of testing, which increases intrinsic CL. There are seven elements that interact with one another in the 254-word passage. As can be guessed from the number of words constituting the text, the text is not long. However, because of the seven elements that interact with each other and because of no prior knowledge on the part of the learners, the text is assumed to have high intrinsic load. For a tornado to form, from the beginning to the end, all the seven elements should occur in a way related to the previous one, which entails knowing a previous element to understand the occurrence of the next element.

The reason why a short passage was chosen is that the researcher is not interested in the difficulty that may be caused by the total number of elements but interested in the amount of element interactivity despite a relatively small number of information elements. Moreover, there were four experimental groups (N+P_HWM, T+P_HWM, N+P_LWM, and T+P_LWM) in the present study. Finally, even though there was the time limitation of ten minutes to read or listen to the text, participants were allowed to read or listen to the text more than once in the overall ten minutes. A summary of the variable information, the research questions and the corresponding procedures can be found in the following tables:

Table 1. Summary of the Variable Information

Variables	Type of variables	Definition	Operationalization	Level
ECL (extraneous cognitive load)	Independent (between groups) Categorical	Cognitive load imposed by presentation mode	Text + picture presentation mode (high load) vs. narration + picture mode (low load)	Two levels: high and low
L2 WM (working memory) capacity	Independent (between groups) Categorical	Ability to deal with incoming verbal information in the form of sentence judgment and keeping last words in memory	Mean z scores of participants on the English verbal working memory test	Two levels: high and low
Time of measurement	Independent (repeated measures) Categorical	Period of time passed between the immediate and delayed comprehension tests	Time interval between immediate and delayed tests (two weeks)	Two levels: immediate and delayed
Retention of information	Dependent Continuous	Ability to reproduce or recognize presented information in a text	Total participant scores on the immediate and delayed retention tests	
Transfer of information	Dependent Continuous	Ability to use presented information in a text in new situations	Total participant scores on the immediate and delayed transfer tests	
Perception of cognitive load	Dependent Continuous	Perception of cognitive load in the presentation of an L2 text as rated on a subjective cognitive load scale	Total participant scores on the subjective cognitive load scale	

Table 2. Overview of the Research Questions and Corresponding Procedures

Research Questions	Data Collection Instruments	Data Analysis Instruments	Expected Results
1. What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on retention of information from an expository multimedia text in the L2?	1. Two web presentations of the L2 text: text + pictures, narration + pictures 2. Computerized English verbal WM test (reading span test) 3. Immediate and delayed retention tests	A 2 (extraneous cognitive load: high and low) X 2 (working memory: high and low) X 2 (time: immediate and delayed) mixed design ANOVA	1. The N+P groups were expected to achieve higher retention scores over time as suggested by Cognitive Theory of Multimedia Learning (Mayer, 2001) 2. Participants with high verbal WM capacity were expected to achieve statistically higher retention scores than the ones with low verbal WM capacity over time in low and high ECL conditions based on Harrington & Sawyer, 1992; Hambrick & Engle, 2002; Hambrick & Oswald, 2005; Daneman & Carpenter, 1980; Mayer & Moreno, 1998; Moreno & Mayer, 1999.
2. What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on transfer of information from an expository multimedia text in the L2?	1. Two web presentations of the L2 text: text + pictures, narration + pictures 2. Computerized English verbal WM test 3. Immediate and delayed transfer tests	A 2 (extraneous cognitive load: high and low) X 2 (working memory: high and low) X 2 (time: immediate and delayed) mixed design ANOVA	1. The N+P groups were expected to achieve higher transfer scores over time as suggested by Cognitive Theory of Multimedia Learning (Mayer, 2001) 2. Participants with high verbal WM capacity were expected to achieve statistically higher transfer scores than the ones with low verbal WM capacity over time in low and high ECL conditions based on Harrington & Sawyer, 1992; Hambrick & Engle, 2002; Hambrick & Oswald, 2005; Daneman & Carpenter, 1980; Mayer & Moreno, 1998; Moreno & Mayer, 1999.

Table 2. Continued

Research Questions	Data Collection Instruments	Data Analysis Instruments	Expected Results
3. What are the immediate and delayed effects of extraneous cognitive load and L2 verbal working memory capacity on participants' perceptions of cognitive load involved in the presentation of an expository multimedia text in the L2?	Multidimensional subjective cognitive load scale	A 2 (extraneous cognitive load: high and low) X 2 (working memory: high and low) X 2 (time: immediate and delayed) mixed design ANOVA	1. The N+P groups were expected to report lower cognitive load scores over time based on Cognitive Theory of Multimedia Learning (Mayer, 2001) and Cognitive Load Theory (Sweller, 1988, 1994) and on Paas, 1992; Tabbers, 2002; Burkes, 2007. 2. Participants with high verbal WM capacity were expected to report lower cognitive load scores over time based on Cognitive Theory of Multimedia Learning (Mayer, 2001) and Cognitive Load Theory (Sweller, 1988, 1994) and on Paas, 1992; Tabbers, 2002; Burkes, 2007.

CHAPTER 4

RESULTS

The current chapter provides the results of the analyses utilized to answer the research questions. In order to examine the effects of working memory (WM) capacity and extraneous cognitive load (ECL) on L2 text comprehension in the short and long terms in a computerized multimedia environment, quantitative analyses and related follow up statistical analyses were conducted based on the procedures described in the previous chapter. The first research question examines the immediate and delayed effects of WM and ECL on retention of information, and it was tested through a retention test given to the participants immediately after the treatment and two weeks after the treatment. The second research question investigates the immediate and delayed effects of WM and ECL on transfer of information, and it was examined through a transfer test given to the participants immediately after the treatment and two weeks after the treatment. The third research question examines the short-term and long-term effects of ECL and WM on participants' perceptions of the overall cognitive load (CL), which was gauged through a subjective CL scale distributed to the participants immediately after the treatment and two weeks after the treatment. Findings are presented by research questions in the following sections.

Effects of Extraneous Cognitive Load and English Verbal Working Memory Capacity on Retention of Knowledge over Time (immediate, delayed)

The first research question investigates whether the level of ECL involved in the presentation mode of an L2 text in a multimedia environment and L2 verbal WM capacity had an effect on participants' retention of information from an expository L2 text over time. The treatment groups were compared by the means of their average retention scores on the immediate and delayed retention tests. The overall results of the immediate and delayed retention tests are described in Table 3:

Table 3. Means and Standard Deviations for Retention Tests

Groups		Retention Test			
Presentation mode (ECL)	English verbal WM capacity	Immediate *M*	Immediate *SD*	Delayed *M*	Delayed *SD*
N+P (low ECL)	Low	7.17	2.317	4.67	1.506
	High	6.44	1.667	6.22	2.224
T+P (high ECL)	Low	7.25	2.866	5.25	1.909
	High	7.83	2.137	4.67	1.751

As can be seen in Table 3 above, the delayed retention test scores were lower than the immediate retention test scores. In the immediate retention test performance T+P_HWM (high working memory group exposed to high extraneous cognitive load) group had the highest score while the participants in N+P_HWM (high working memory group exposed to low extraneous cognitive load) group had the lowest score. Overall, participants in the T+P (high ECL) condition got higher scores than the participants in the N+P (low ECL) condition. As for the delayed retention test, N+P_HWM earned the highest score while the T+P_HWM and the N+P_LWM (low working memory group exposed to low extraneous cognitive load) had the lowest scores. Actually these two groups had the same average retention score on the delayed retention test. The table also shows that N+P_HWM group had similar scores on the retention tests, while the difference between immediate retention and delayed retention scores were the highest for T+P_HWM group. Mean immediate and delayed retention scores can also be seen in Figure 8:

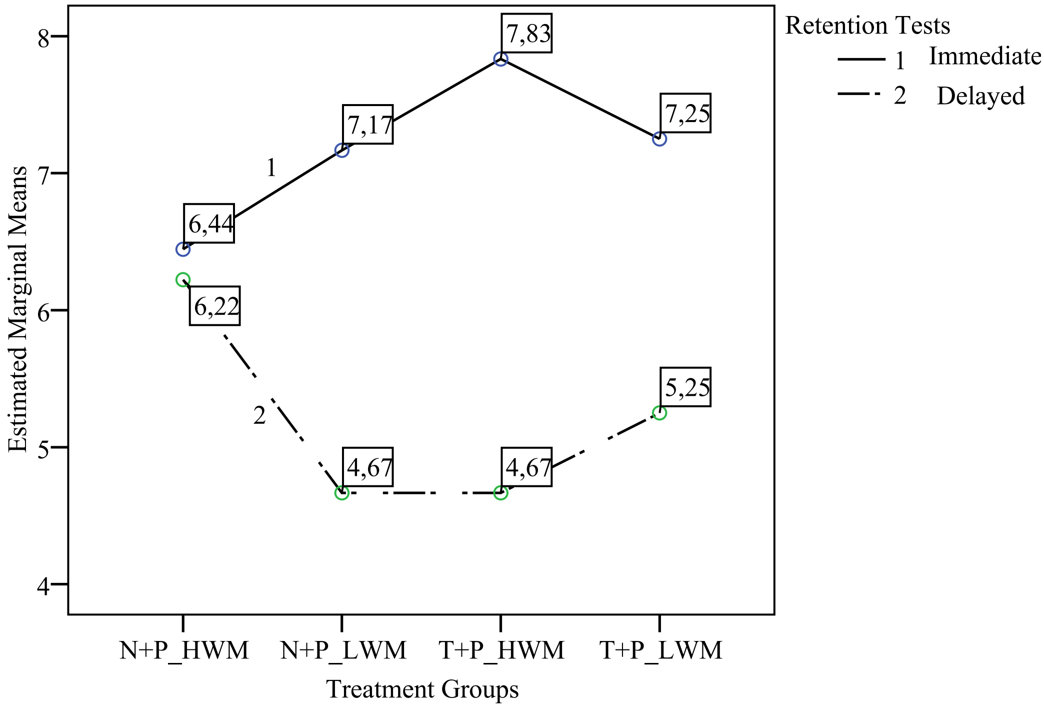


Fig. 8. Mean immediate and delayed retention scores for each treatment group

In order to determine whether these differences in scores were statistically significant, immediate and delayed retention test scores were analyzed through a 2 (WM: high and low) X 2 (ECL: high and low) X 2 (Time: immediate and delayed retention tests) mixed design ANOVA. Table 4 provides the summary table for the mixed ANOVA:

Table 4. Summary of Mixed ANOVA for Immediate and Delayed Retention Tests

Source	SS	df	MS	F
Between Subjects		28		
ECL	.220	1	.220	.031
WM	.610	1	.610	.087
ECL*WM	.610	1	.610	.087
Error between	175.167	25	7.007	
Within Subjects		29		
Time	54.645	1	54.645	29.734**
Time*ECL	5247	1	5.247	2.855
Time*WM	1.084	1	1.084	.590
Time*ECL*WM	10.417	1	10.417	5.668*
Error within	45.944	25	1.838	
Total		57		

** $p < .001$ *$p = .025$

Table 4 refers to two statistically significant effects. However, the significant results should be approached with caution because of the low power of the test for the non-significant main and combined effects (.37 for time*ECL; .11 for time*WM; .059 for ECL*WM; .053 for ECL; .059 for WM). The power of the test for time effect was .99. Even the power for the significant triple interaction, .62, was lower than the desirable power value of .80. This implies that the sample size was not large enough to detect significant effects of independent variables and double interactions above. Because of the significant interaction, the significant main effect of time can be ignored. Table 4 also shows that the interaction of time, WM and ECL has a significant effect on retention performance of the participants, $F_{1,25} = 5{,}668$, $p = .025$, $\eta^2 = .18$ with a large effect size ($d > .8$). The significant combined effect of ECL, WM, and time can be seen in Figure 9 and Figure 10:

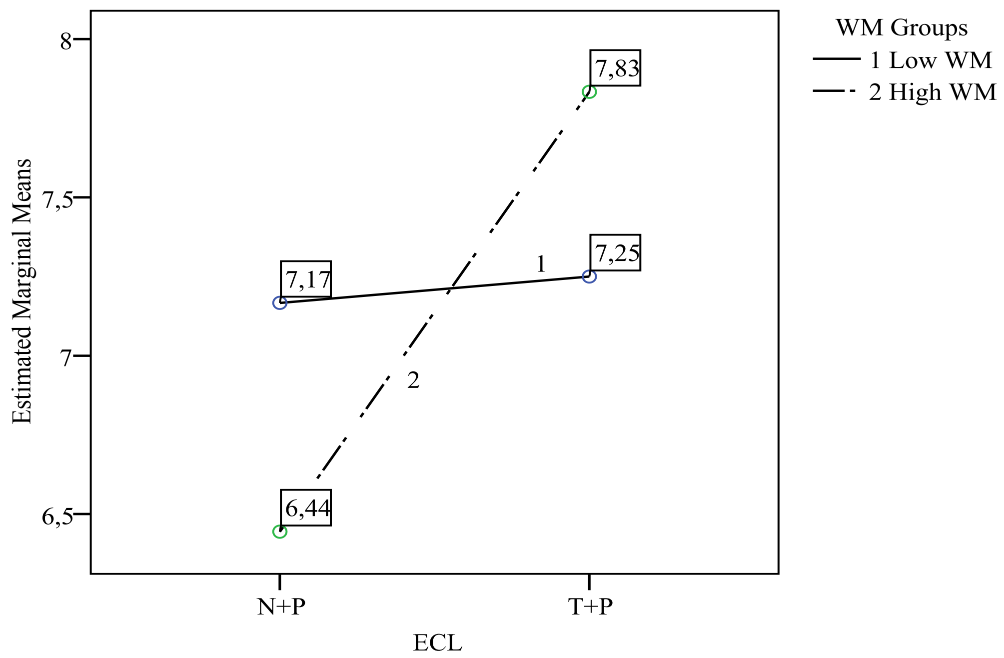


Fig. 9. ECL and WM effects on immediate retention test

A two-way ANOVA with ECL and WM as the independent variables and immediate retention scores as the dependent variable was conducted to determine whether the visible interaction of ECL and WM in Figure 9 above is significant and to see if there are any significant main effects. Results revealed a nonsignificant main effect for both ECL ($F_{1,25}$ = .735, p > .05) and WM ($F_{1,25}$ = .007, p > .05), and a nonsignificant combined effect of ECL and WM ($F_{1,25}$ = .578, p > .05). This shows that there were no retention performance differences among the treatment groups on the immediate retention test. In other words, there was no immediate ECL or WM effect on retention of information.

The nonsignificant interaction effect of ECL and WM capacity on immediate retention performance, on the basis of Figure 9, is not surprising, since time effect was eliminated. However, still, Figure 9 reveals an interaction between ECL and WM on retention performance that reaches statistical significance over time.

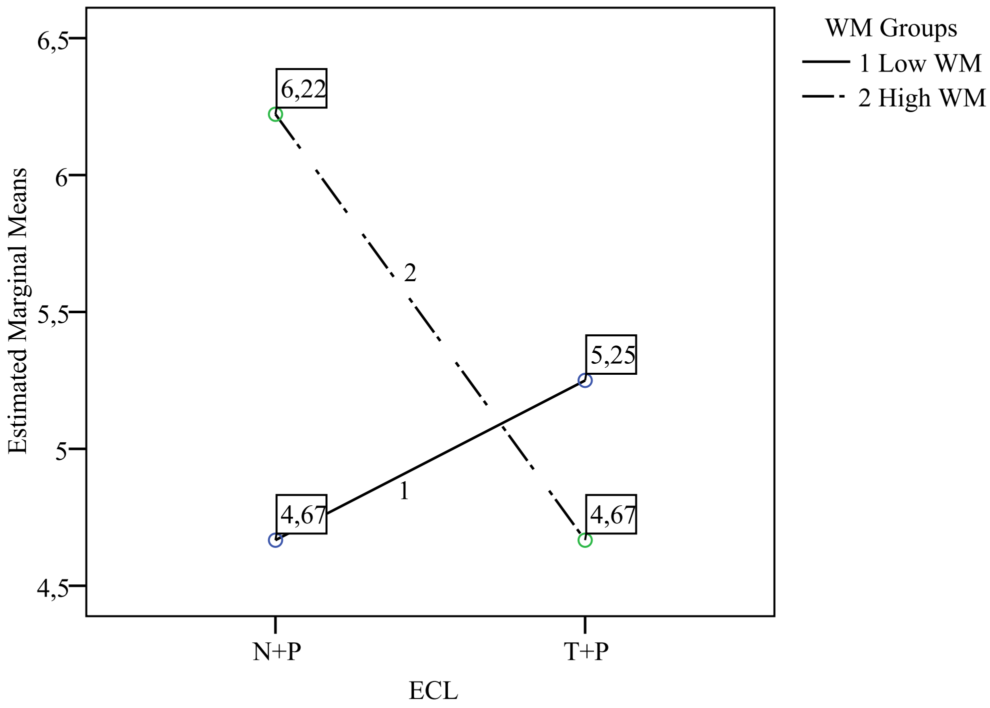


Fig. 10. ECL and WM effects on delayed retention test

A two-way ANOVA with ECL and WM as the independent variables and delayed retention scores as the dependent variable was performed to determine whether the visible interaction of ECL and WM on delayed retention test in Figure 10 is significant and to detect if there are any significant main effects. The results indicated no main effect either for ECL ($F_{1,25} = .452$, $p > .05$) or for WM ($F_{1,25} = .452$, $p > .05$), and no significant interaction of ECL and WM ($F_{1,25} = .2.190$, $p > .05$), which shows that none of the experimental groups performed better than the others on the delayed retention test.

However, Figure 10 still refers to a combined effect of WM and ECL observed in the significant triple interaction of WM, ECL, and time. Briefly, the non-significant interaction effect of WM and ECL on both immediate and delayed retention tests and the significant combined effect of WM, ECL, and time of testing reveal that time was an important factor affecting retention performance of the participants in the current study.

It is also clear from Table 4 that the combined effect of WM and ECL on retention scores of the participants was not significant regardless of the main effect of time of measurement at the p-value of .025. The statistically significant triple interaction means that the effect of time was not the same among the four experimental groups (N+P_HWM, T+P_HWM, N+P_LWM, T+P_LWM). In other words, there was not the same change in retention scores of the experimental groups over time due to the combined effects of time, WM, and CL. More specifically speaking, the pattern of differences in retention scores on the immediate retention test among treatment groups and the ones on the delayed retention test was not equal, which in turn means that pattern of retention score differences between the immediate and delayed tests is different among the four groups. Table 5 displays difference scores on the retention tests (immediate - delayed) for each group:

Table 5. Descriptive Statistics for Retention Difference Scores for Each Treatment Group

Group	N	Min.	Max.	M	SD	Possible Maximum
N+P_HWM	9	-2	4	.22	1.92	15
N+P_LWM	6	1	4	2.50	1.37	15
T+P_HWM	6	.00	6	3.17	2.56	15
T+P_LWM	8	.00	5	2	1.69	15

Furthermore, Figure 11 presents the pattern of retention difference scores among experimental groups:

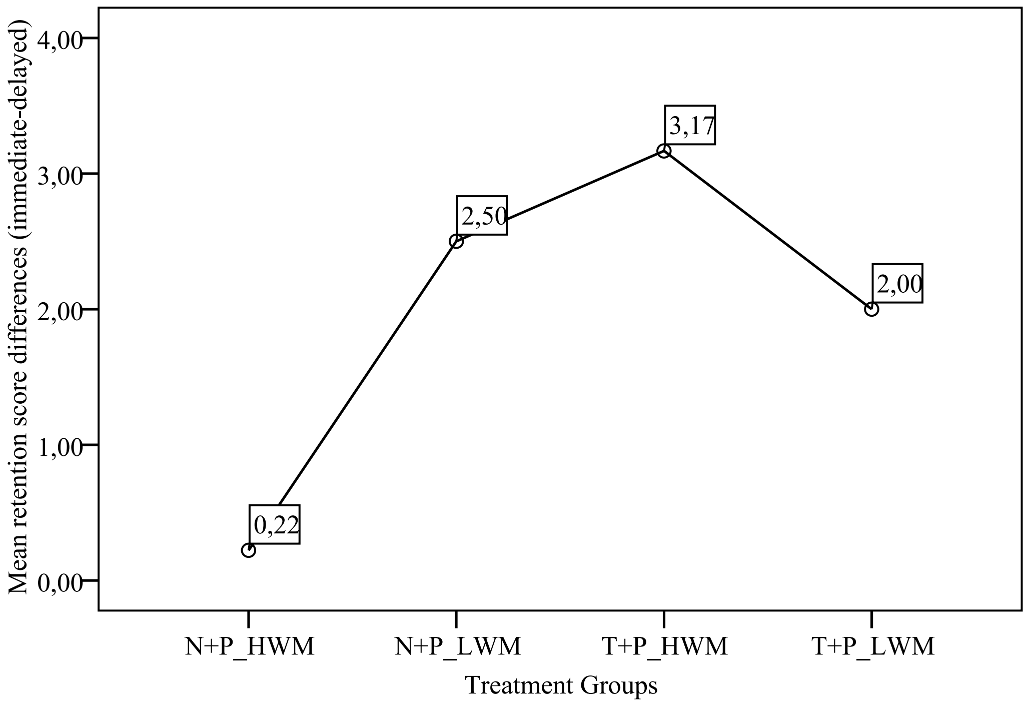


Fig. 11. Mean retention score differences for treatment groups

In order to determine the source of the significant triple interaction, post hoc pair-wise comparison tests were conducted. Following pair-wise comparisons showed that the three experimental groups T+P_HWM, N+P_LWM, and T+P_LWM lost significant retention scores on the delayed retention test, but not N+P_HWM group as shown by Table 6. Table 6 clearly indicates that changes over time in participants' responses on the retention tests differed by the level of ECL across the levels of L2 verbal WM capacity:

Table 6. Pair-wise Comparisons for Retention Difference Scores for the Treatment Groups

WM	ECL	Mean Difference (Immediate – delayed)	Std. Error	Sig.
Low	N+P	2.50*	.783	.004
Low	T+P	2.00*	.678	.007
High	N+P	.22	.639	.731
High	T+P	3.167*	.783	.000

*p < .01 with Bonferroni adjustment

Figure 12 depicts a comparison of the average retention difference scores among the four experimental groups:

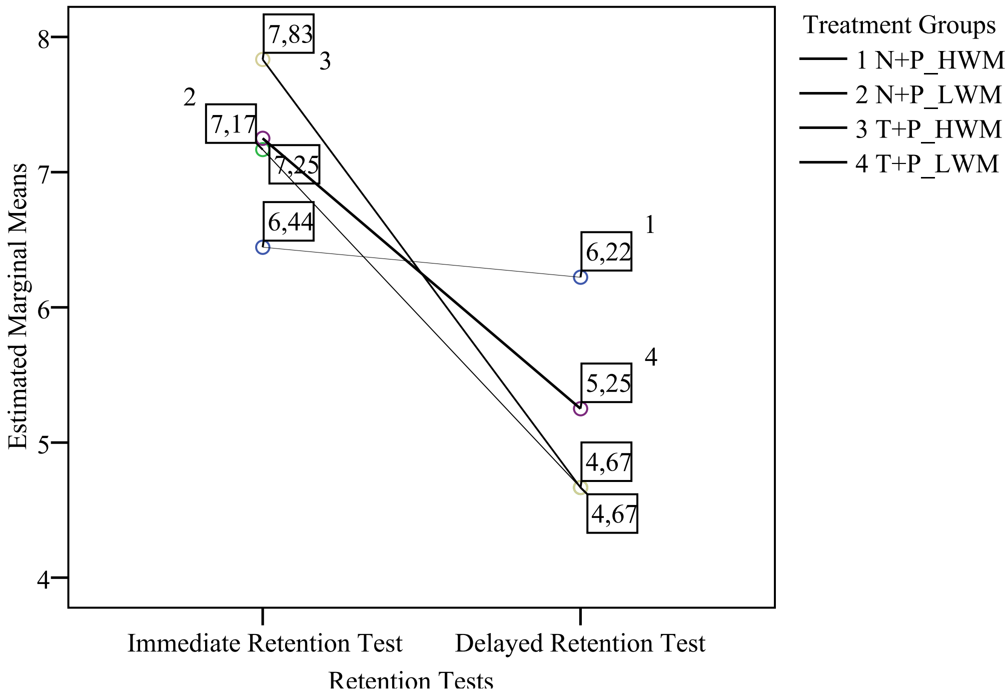


Fig. 12. Comparison of mean retention score differences for each treatment group over time

In order to get deeper insights into the nature of the significant combined effect of time, ECL, and WM, orthogonal contrasts were conducted on retention difference scores of the treatment groups. Field (2000) suggests using “one less contrast than the number of groups” (p. 260). Thus, three orthogonal contrast procedures with three contrasts and one procedure with one contrast were conducted to learn more about the significant interaction. The first contrast procedure firstly indicated that the difference between the combined scores of N+P_HWM and N+P_LWM groups, and that of T+P_HWM and T+P_LWM groups was not statistically different (T_{25} = -.1690, p > .05). This means that low ECL did not help participants to lose fewer scores on the delayed retention test compared to participants under high ECL regardless of WM capacity. Second, the difference between N+P_HWM and N+P_LWM groups was significant meaning that in low ECL group, higher WM led to less difference in retention scores (T_{25} = -2.254, p < .05). The result of this contrast can also be seen in Figure 13:

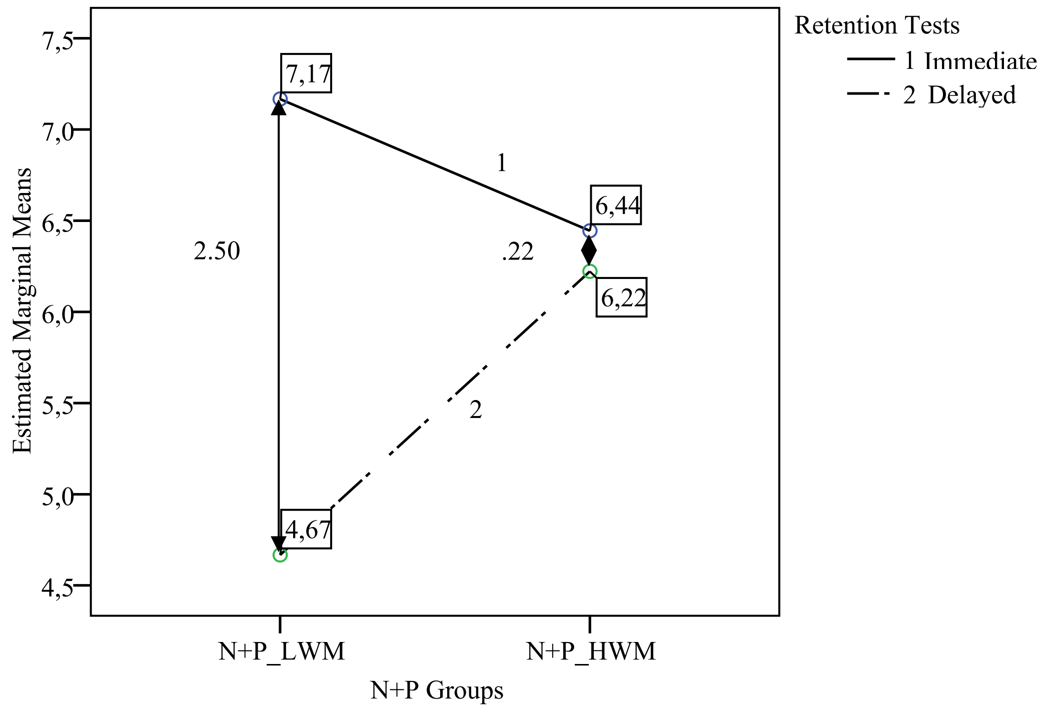


Fig. 13. WM and time effects on retention performance of N+P group

The significant orthogonal contrast between N+P_HWM and N+P_LWM indicates that retention score difference in N+P_LWM is statistically different from the difference in N+P_HWM, which suggests that there is a significant interaction of time and WM in N+P group. In other words, average retention score difference between immediate and delayed tests in N+P_LWM group is statistically higher than that in N+P_HWM group. Figure 13 above and the result of the orthogonal contrast analysis also suggest that L2 verbal WM capacity played some role in retention performance under low ECL depending upon time interval between immediate and delayed retention tests. Interestingly, however, there was not a significant retention score difference between T+P_HWM and T+P_LWM groups, which suggests that high WM participants under high ECL did not lose fewer retention scores than their counterparts with low WM capacity ($T_{25} = 1.127$, $p > .05$) as shown by Figure 14:

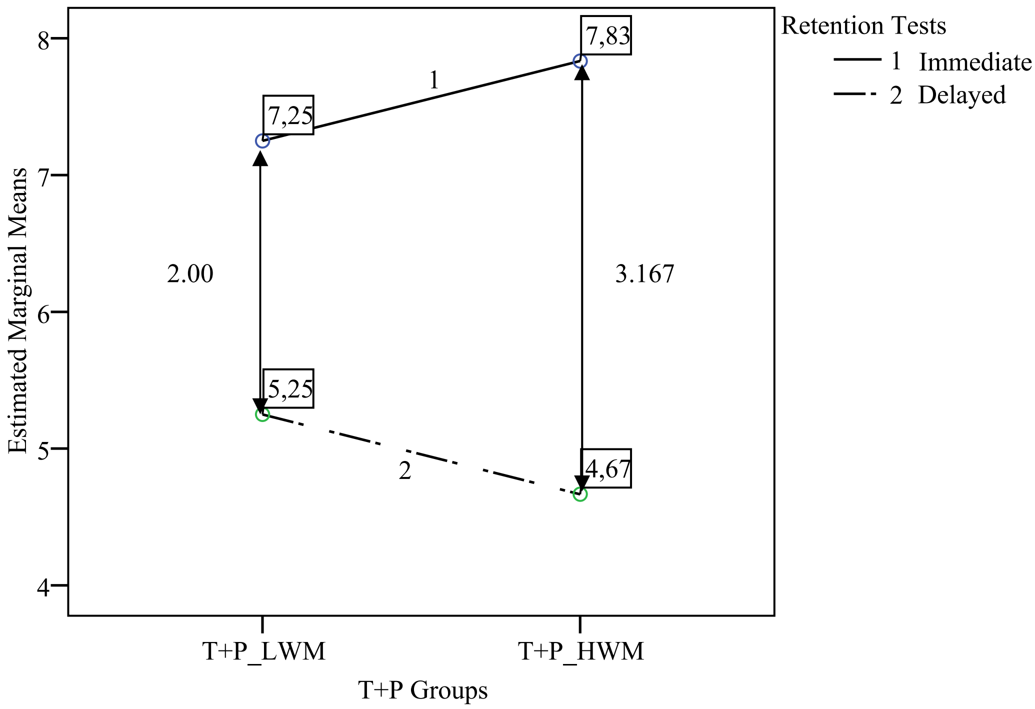


Fig. 14. WM and time effects on retention performance of T+P group

The non-significant orthogonal contrast between T+P_HWM and T+P_LWM indicates that retention score difference in T+P_LWM is not statistically different from the difference in T+P_HWM, which suggests that there is not a significant interaction of time and WM in T+P group. In other words, average retention score difference between immediate and delayed tests in T+P_LWM group is the same as that in T+P_HWM condition.

The second orthogonal contrast procedure revealed firstly that the mean retention score difference of N+P_HWM group was statistically lower than the combined mean retention score differences of T+P_HWM, T+P_LWM, and N+P_LWM groups ($T_{25} = -3.023$, $p < .025$). Moreover, there was not a significant difference between the combined mean retention score difference of N+P_LWM and T+P_HWM, and mean retention score difference of T+P_LWM ($T_{25} = .952$, $p > .05$). Finally, there was not a significant retention score difference between N+P_LWM and T+P_HWM groups ($T_{25} = -.602$, $p > .05$).

According to the third orthogonal contrast procedure, there was not a significant difference between the combined mean retention score differences of N+P_HWM and T+P_HWM and the total mean of N+P_LWM and T+P_LWM groups ($T_{25} = -.768$, $p > .05$). There was a significant difference between N+P_HWM and T+P_HWM groups, indicating that high WM participants in the N+P condition lost significantly fewer retention scores than the high WM participants in T+P condition ($T_{25} = -2.914$, $p < .025$) as can be seen in Figure 15:

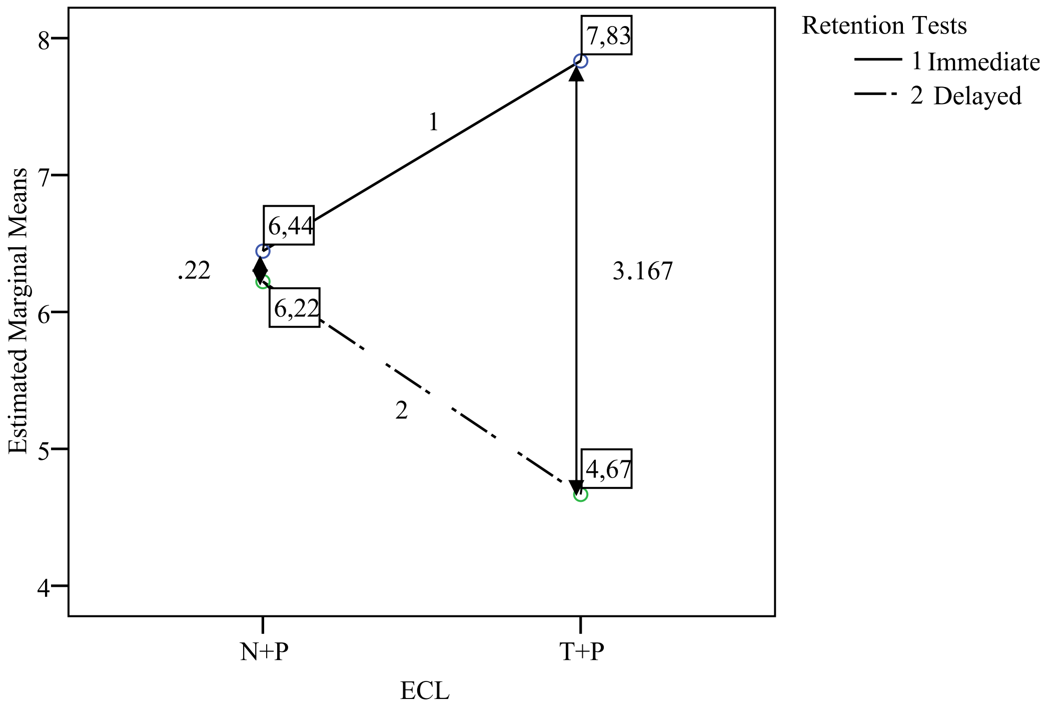


Fig. 15. ECL and time effects on retention performance of high WM group

The significant orthogonal contrast between N+P_HWM and T+P_HWM indicates that retention score difference in N+P_HWM is statistically different from the difference in T+P_HWM, which suggests that there is a significant interaction of time and ECL in high WM group. In other words, the difference in average retention scores between immediate and delayed tests in N+P_HWM group is statistically lower than that in T+P_HWM condition. Figure 15 above and the result of the orthogonal contrast analysis also suggest that ECL had some effect on retention performance in high WM group depending upon time interval between immediate and delayed retention tests. Finally, there was not a significant difference between N+P_LWM and T+P_LWM groups ($T_{25} = .483$, $p > .05$) as shown in Figure 16:

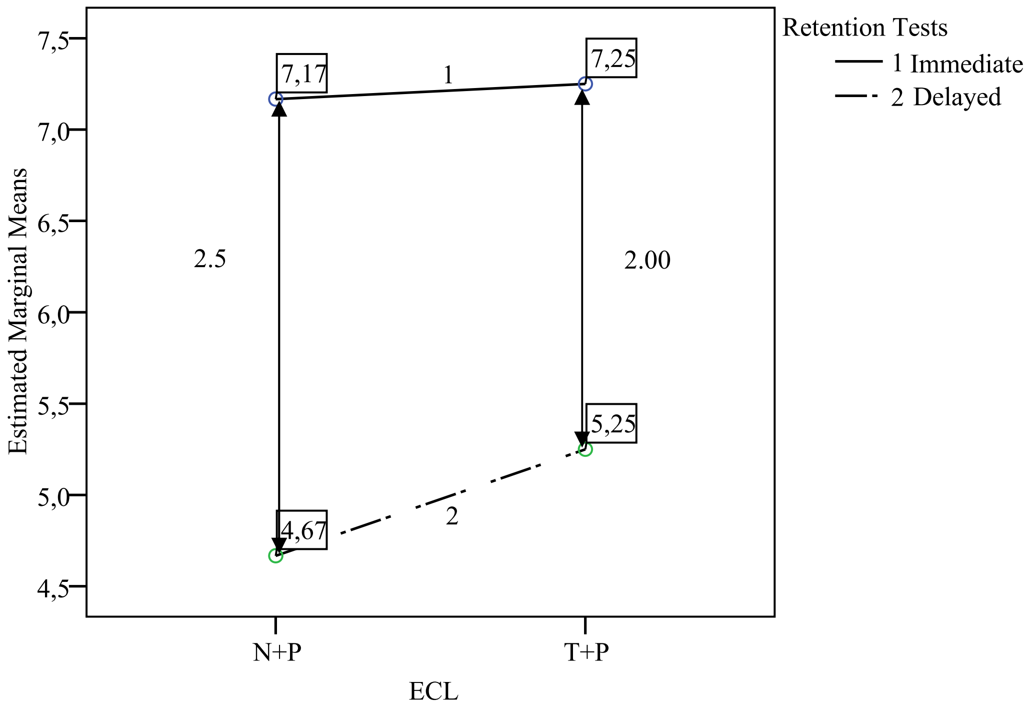


Fig. 16. ECL and time effects on retention performance of low WM group

The non-significant orthogonal contrast between N+P_LWM and T+P_LWM indicates that retention score difference in N+P_LWM is not statistically different from the difference in T+P_LWM, which suggests that there is not a significant interaction of time and ECL in low WM group. In other words, average retention score difference between immediate and delayed tests in T+P_LWM group is the same as that in N+P_LWM group. That said, there remains only one contrast not conducted: the one between N+P_HWM and T+P_LWM. The orthogonal contrast between these two groups yielded that the groups were not significantly different from each other in terms of average retention score differences ($T_{25} = 1.908$, $p > .05$).

Finally, orthogonal contrast analyses above showed that high WM participants in N+P group achieved higher retention performance than high WM participants in T+P group but not than low WM participants in T+P condition over time. Likewise, low WM participants in N+P group did not achieve higher retention scores than both the high WM and low WM participants in T+P condition. As a result, Hypothesis 1-a, which assumes that participants in N+P condition will achieve higher retention scores than those in T+P condition, is partially rejected in the sense that only under the condition of high WM capacity N+P participants performed better on retention tests. Moreover, the same analyses indicated that only under low ECL (N+P) high WM capacity participants outperformed those with low WM, which partially retains Hypothesis 2-a (Participants with high verbal WM capacity will achieve statistically higher retention scores than the ones with low verbal WM capacity over time). The former indicates that WM capacity may mediate the modality principle of Mayer's (2001) cognitive theory of multimedia learning in that the modality effect works more for high WM individuals.

To sum up, the results of the orthogonal contrast analyses above show that there are multiple sources of the significant combined effect of time, ECL, and WM on retention performance: N+P_HWM group lost significantly less retention scores than the total retention score difference of T+P_HWM, T+P_LWM, and N+P_LWM groups on average. Furthermore, there was a significant average retention score difference between N+P_HWM and N+P_LWM; N+P_HWM and T+P_HWM groups but not between N+P_HWM and T+P_LWM. The significant orthogonal contrast between N+P_HWM and N+P_LWM in terms of average retention score difference indicates that higher WM capacity leads to better performance under low ECL while the significant contrast between N+P_HWM and T+P_HWM implies that modality effect applies more for high WM L2 learners. The results also showed that time of testing was an important factor that affects retention performance of the participants moderating the effects of ECL and WM on retention performance.

Effects of Extraneous Cognitive Load and English Verbal Working Memory Capacity on Transfer of Knowledge over Time (immediate, delayed)

The second research question investigates whether level of ECL imposed by the presentation mode of an L2 text in a computerized multimedia environment and L2 verbal WM capacity had an effect on participants' transfer of information over time. The treatment groups were compared by the means of their average transfer scores on the immediate and delayed transfer tests. The overall results of the immediate and delayed transfer tests are described in Table 7:

Table 7. Means and Standard Deviations for Transfer Tests

Groups		Transfer Test			
Presentation mode (ECL)	English verbal WM capacity	Immediate M	Immediate SD	Delayed M	Delayed SD
N+P (low ECL)	Low	7.00	.894	6.33	2.733
	High	7.56	2.744	7.22	2.224
T+P (high ECL)	Low	8.38	2.326	6.75	1.982
	High	7.67	2.582	6.83	2.13

As can be seen in Table 7, the delayed transfer test scores were lower than the immediate transfer test scores. In the immediate transfer test T+P_LWM had the highest score while the N+P_LWM had the lowest score. Overall, participants in the T+P condition got higher scores than the participants in the N+P condition. As for the delayed transfer test, N+P_HWM group had the highest scores while the N+P_LWM had the lowest scores. The table also shows that N+P_HWM group had similar scores on the transfer tests, while the difference between immediate and delayed transfer scores were the highest for T+P_LWM. Figure 17 displays mean immediate and delayed transfer scores of each treatment group:

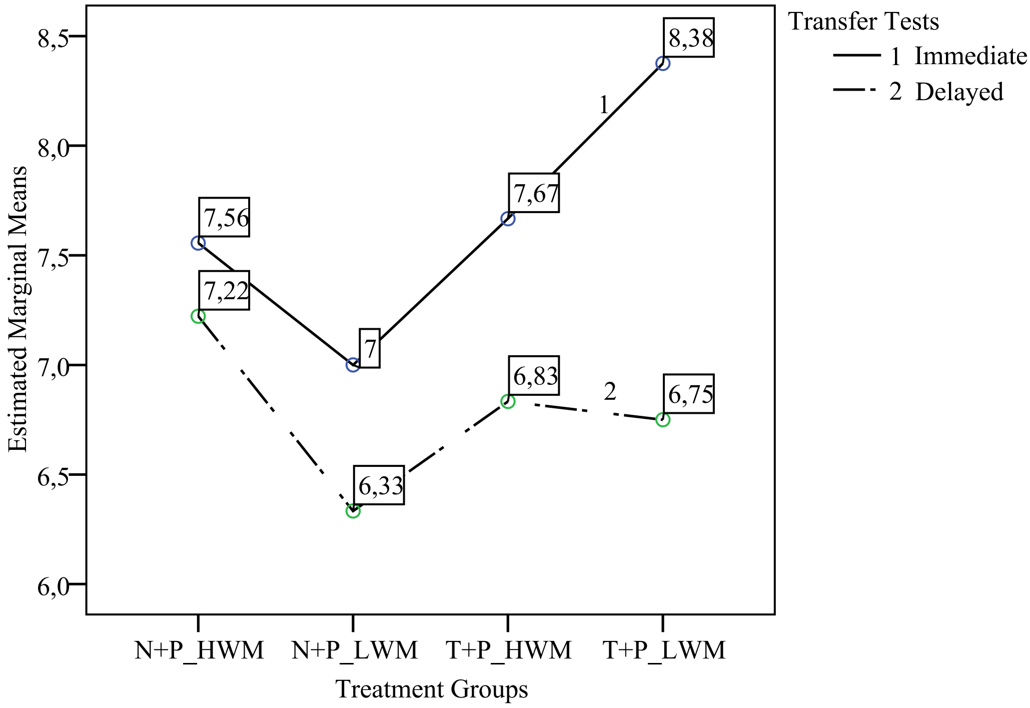


Fig. 17. Mean immediate and delayed transfer scores for each treatment group

In order to determine whether these differences in scores were statistically significant, immediate and delayed transfer test scores were analyzed through a 2 (WM: high and low) X 2 (ECL: high and low) X 2 (Time: immediate and delayed transfer tests) mixed design ANOVA. Table 8 provides the summary table for the mixed ANOVA:

Table 8. Summary of Mixed ANOVA for Immediate and Delayed Transfer Tests

Source	SS	df	MS	F
Between Subjects		28		
ECL	2.012	1	2.012	.245
WM	.590	1	.590	.072
ECL*WM	3.760	1	3.760	.457
Error between	205.632	25	8.225	
Within Subjects		29		
Time	10.502	1	10.502	4.604*
Time*ECL	1.867	1	1.867	.819
Time*WM	1.111	1	1.111	.487
Time*ECL*WM	.184	1	.184	.081
Error within	57.021	25	2.281	
Total		57		

Note: None of the main effects and interaction effects was statistically significant at the adjusted p-value of .025. *$p < .05$

Remember that in the present study, the p-value of .025 was chosen through Bonferroni adjustment because of the number of dependent variables involved in each research question to minimize Type 1 error risk. The power indexes of the 3-way mixed design ANOVA conducted on immediate and delayed transfer tests (.076 for ECL; .058 for WM; .100 for ECL*WM; .541 for time; .140 for time*ECL, .103 for time*WM; .059 for time*ECL*WM) are quite less than the desirable power value of .80, which indicates that the present study have a sample size problem. As such, if we remove the Bonferroni adjustment on our critical alpha level and increase it to .05, and conduct the mixed ANOVA again, by following Stevens' (1996) suggestions stated in the previous chapter (see page 85), the ANOVA summary table above suggests that time had a statistically significant main effect on transfer performance of the participants $F_{1,25} = 4.604$, $p < .05$, $\eta^2 = .15$ with a large effect size ($d > .8$). The main effect of ECL ($F_{1,25} = .245$, $p > .05$) and L2 WM capacity ($F_{1,25} = .072$, $p > .05$)

turned out to be nonsignificant. Following pair-wise comparisons showed that participants had significantly higher scores on the immediate transfer test. Table 9 shows the results of pair-wise comparisons:

Table 9. Pair-wise Comparisons for Time Effect on Transfer Scores of the Participants

Transfer tests	Mean difference	Standard Error
Immediate – Delayed	.865*	.403
Delayed – Immediate	-.865*	.403

* $p < .05$

Table 9 above suggests that delayed transfer scores of the participants were negative because of the transfer score differences on the delayed transfer test. In other words, all treatment groups lost transfer scores on the delayed transfer test on average. In order to interpret more precisely, it may be helpful to examine the mean transfer score difference of each group (immediate – delayed). Table 10 represents these scores:

Table 10. Descriptive Statistics for Transfer Difference Scores for Each Treatment Group

Group	N	Min.	Max.	M	SD	Possible Maximum
N+P_HWM	9	-3	2	.33	1.41	20
N+P_LWM	6	-3	4	.67	2.58	20
T+P_HWM	6	-.5	5	.83	3.31	20
T+P_LWM	8	.00	4	1.62	1.18	20

Following pair-wise comparisons with Bonferroni adjustment showed that only T+P_LWM group lost significant transfer scores, as shown by Table 11:

Table 11. Pair-wise Comparisons for Transfer Difference Scores for the Treatment Groups

WM	ECL	Mean Difference (Immediate – delayed)	Std. Error	Sig.
Low	N+P	.67	.783	.004
Low	T+P	1.62*	.678	.007
High	N+P	.33	.639	.731
High	T+P	.83	.783	.000

*p < .05 with Bonferroni adjustment

Figure 18 presents the pattern of transfer score differences among treatment groups:

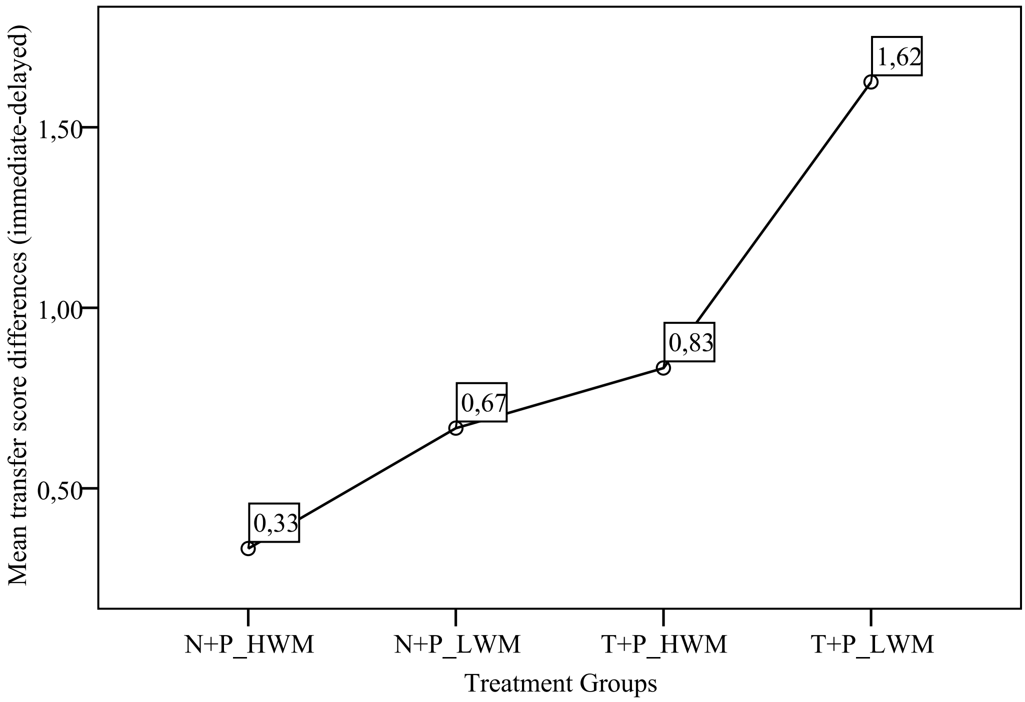


Fig. 18. Mean transfer score differences for treatment groups

Consequently, both Hypothesis 1-b, which predicts that participants under low ECL will perform significantly better on the transfer tests than the ones under high ECL, and Hypothesis 2-b, which assumes that high WM participants will achieve a better transfer performance than those with low WM capacity are rejected. All of these groups achieved the same amount of transfer performance over time. These results are in contrast with cognitive theory of multimedia learning (Mayer, 2001).

In sum, it was found that time of testing has a significant effect on L2 learners' transfer of information performance in the sense that transfer performance decreases significantly over time independent of ECL and WM capacity.

Effects of Extraneous Cognitive Load and English Verbal Working Memory Capacity on Perception of Cognitive Load over Time (immediate, delayed)

The third research question examines whether the level of ECL imposed by the presentation mode of the L2 text in a computer-based multimedia environment and L2 verbal WM capacity had an effect on participants' perception of CL over time. The treatment groups were compared by the means of their average CL ratings on both the first and second administrations of the subjective CL scale developed by the researcher. The overall results of the immediate and delayed administrations of the subjective CL scale are described in Table 12:

Table 12. Means and Standard Deviations for Subjective CL Scales

Groups		Administration of CL Scale			
Presentation mode (ECL)	English verbal WM capacity	Immediate M	Immediate SD	Delayed M	Delayed SD
N+P (low ECL)	Low	32.00	4.336	31.33	6.346
	High	27.67	5.148	31.00	6.928
T+P (high ECL)	Low	27.25	7.704	29.12	6.686
	High	24.17	7.026	21.83	8.727

As Table 12 displays, overall, participants in N+P condition reported higher CL ratings than the ones in the T+P condition on both immediate and delayed administration of the subjective CL scale. In the immediate ratings, N+P_LWM group had the average highest rating score while the T+P_HWM group had the lowest rating score. As for the delayed administration of the scale, it was again N+P_LWM group that had the highest average rating while the T+P_HWM had the lowest rating. The table also shows that N+P_LWM group had similar rating scores on the subjective CL scales, while the difference between average immediate rating and delayed rating scores were the highest for N+P_HWM group. Furthermore, average delayed CL rating scores decreased in N+P_LWM and T+P_HWM groups. Figure 19 presents mean CL ratings of each treatment group:

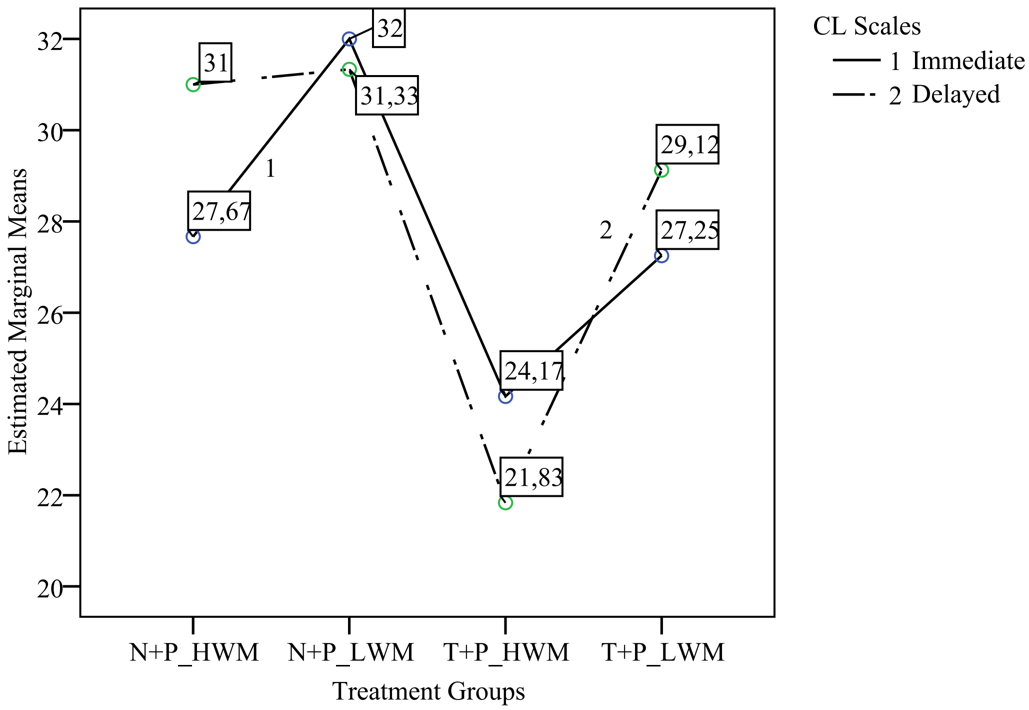


Fig. 19. Mean CL ratings for each treatment group

In order to determine whether these differences in average CL rating scores were statistically significant, immediate and delayed rating scores were analyzed through a 2 (WM: High and Low) X 2 (ECL: High and Low) X 2 (Time: immediate and delayed retention tests) mixed design ANOVA. Table 13 presents the summary table for the mixed ANOVA:

Table 13. Summary of Mixed ANOVA for Immediate and Delayed Subjective Cognitive Load Scales

Source	SS	df	MS	F
Between Subjects		28		
ECL	338.172	1	338.172	5.019*
WM	198.660	1	198.660	2.948
ECL*WM	28.611	1	28.611	.425
Error between	1684.604	25	67.384	
Within Subjects		29		
Time	4.282	1	4.282	.190
Time*ECL	8.575	1	8.575	.381
Time*WM	.038	1	.038	.002
Time*ECL*WM	59.160	1	59.160	2.628
Error within	562.771	25	22.511	
Total		57		

Note: None of the main effects and interaction effects was statistically significant at the adjusted p-value of .025. * $p < .05$

The ANOVA summary table above indicates that none of the main effects and interaction effects was statistically significant at the adjusted p-value of .025. However, the power indexes of the 3-way mixed design ANOVA conducted on immediate and delayed administrations of the subjective CL scale (.577 for ECL; .379 for WM; .096 for CL*WM; .070 for time; .091 for time*ECL; .050 for time*WM; .344 for time*ECL*WM) are quite less than the desirable power value of .80, which implies that sample size in the present study was not large enough to detect any possible significant effects. As such, if we remove the Bonferroni adjustment on our critical alpha level and increase it to .05, and conduct the ANOVA again, by following Stevens' (1996) suggestions (see page 85), the ANOVA summary table above suggests that ECL had a significant main effect on perception of CL over time ($F_{1,25} = 5.019$, $p < .05$, $\eta^2 = .16$) with a large effect size ($d > .8$). As Table 13 shows, there was no significant effect of WM capacity ($F_{1,25} = 2.948$, $p > .05$). Following pair-wise comparisons showed that participants in the N+P condition reported significantly higher CL rating scores than the ones in the T+P condition over time. Table 14 below shows the results of the pair-wise comparisons:

Table 14. Pair-wise Comparisons for ECL Effect on Cognitive Load Ratings of the Participants

Extraneous Cognitive Load	Mean Difference	Standard Error
Low (N+P) – High (T+P)	4.906*	2.190
High (T+P) – Low (N+P)	-4.906*	2.190

* $p < .05$

Table 14 shows that participants exposed to low ECL (N+P) perceived significantly higher CL than the ones exposed to high ECL (T+P). Correlation analyses were conducted to further test this finding. Before running correlation analyses, comprehension scores and subjective CL rating scores were turned into standardized values (*z* scores) since reading span scores were calculated in the form of standardized values beforehand. Because of the small sample size and range of scores are very small and assumptions of homoscedasticity and linearity were violated, Spearman's rank order correlation coefficient was used. There was a significant, negative, and large correlation between WM and subjective CL ratings in the N+P group ($r_s = -.518$, $n = 15$, $p < .05$) with a shared variance of 26.83 %. This negative correlation coefficient suggests that N+P participants spent a significant amount of their WM resources on dealing with CL, which explains the finding that N+P group reported significantly higher CL ratings than T+P group on average. On the other hand, there was a small negative correlation between WM capacity and subjective CL ratings of the T+P group (r_s = -280, n = 14, p > .05).

The findings of ANOVA and correlation analyses are inconsistent with Mayer's (2001) modality principle. However, the CL scale used in the present study gauged overall CL not just ECL. CL is the additive sum of intrinsic CL, ECL, and germane load. Moreover, the text was intrinsically high loaded in both N+P and T+P conditions. Therefore, significant CL rating scores may stem from ECL differences, differences between the amount of mental effort to understand the text, and both but not from ECL only. In order to test these possibilities follow up mixed ANOVA analyses were conducted:

First, a mixed ANOVA with WM (high versus low), presentation mode (N+P versus T+P) as the between-groups independent variables and time (immediate versus delayed) as the within-groups independent variable was conducted on the fifth item [How much effort did you invest to navigate through the text (scrolling up and downward, controlling winamp)?] of the subjective CL scale. Before running the analysis, immediate rating of one of the subjects on the fifth item was removed, since it was an extreme score. Moreover, normality assumption was violated by both immediate and delayed ratings of the fifth item. "Square root" and "logarithm" transformations conducted on immediate and delayed ratings of the fifth item did not work. Thus, the data were left as they were. No effects including the main effect of presentation mode was significant (p's > .05). This suggests that statistically higher CL ratings of the participants exposed to low ECL does not stem from any navigational difficulty participants in the N+P presentation might have encountered including the separate control of the webpage and the winamp panel.

Second, an analogous ANOVA was also run on the first item [(How much mental effort was required (e.g. thinking, deciding, remembering etc.)?] to determine whether participants exposed to low ECL spent more effort than the ones exposed to high ECL. Immediate and delayed ratings on the first item were not normally distributed. "Reflect and square root" and "reflect and logarithm" transformations done on the immediate ratings did not work. As a result, immediate rating data on the first item were left as they were. On the other hand, "reflect and square root" transformation of the delayed ratings on the first item worked, and transformed data were used in the ANOVA analysis. The results indicated no significant main effect of presentation mode and no other significant effects (p's $> .05$), which eliminates the option that N+P groups might have invested more effort than T+P groups.

Finally, when the significant correlations found between reading spans and CL ratings of the participants in N+P condition come together with the finding that high WM participants exposed to low ECL achieved more retention performance over time, this raises the question whether participants in this group spent any amount of WM capacity on L2 comprehension. In order to answer this question, partial correlation analyses in which CL ratings were controlled for were used to detect whether there was any relationship between N+P participants' reading spans and L2 listening comprehension, since there was a significant correlation between WM and subjective CL as indicated above. No significant correlations were found between L2 reading span and any of the L2 listening comprehension tests: 1) between WM and immediate retention test ($r = -.287$, $n = 15$, $p > .05$); 2) between WM and delayed retention test ($r = -.109$, $n = 15$, $p > .05$); 3) between WM and immediate transfer test ($r = .502$, $n = 15$, $p = .067$); and 4) between WM and delayed transfer test ($r = .209$, $n = 15$, $p > .05$). Moreover, there were no significant correlations between CL and comprehension tests in the N+P group when WM is controlled for: 1) between CL and immediate retention test ($r = -.352$, $n = 15$, $p > .05$): 2) between CL and delayed retention test ($r = -.363$, $n = 15$, $p > .05$); 3) between CL and immediate transfer test ($r = .349$, $n = 15$, $p > .05$); and 4) between CL and delayed transfer test ($r = -.171$, $n = 15$, $p > .05$). These results indicate that under low ECL, WM resources were used to deal with CL, which did not directly promote comprehension.

However, N+P group consisted of both high WM and low WM participants, which could have affected the results of the correlation analyses stated above because of the WM capacity differences existing among the participants. Table 15 separately presents Spearman's rank order correlations between WM and L2 comprehension in high WM and low WM groups under low ECL:

Table 15. Relationship between WM and L2 Comprehension in N+P Groups

N+P Groups	Immediate Retention Test	Delayed Retention Test	Immediate Transfer Test	Delayed Transfer Test	Subjective Cognitive Load Scale
HWM	.295	-.351	.740*	.866**	-.494
LWM	-.441	-.500	-.239	-.116	-.058

N+P_HWM (N) = 9; N+P_LWM (N) = 6
*$p < .05$, **$p < .01$

As can be seen in Table 15, reading spans of low WM group exposed to low ECL do not correlate significantly with comprehension tests and CL ratings, which may be because of their lack of WM resources. On the other hand, reading span scores of high WM group in N+P condition significantly correlate with both immediate transfer test ($r_s = .740$, $n = 9$, $p < .05$) with a shared variance of 54.80 % and delayed transfer test ($r_s = .866$, $n = 6$, $p < .01$) with 75 % shared variance. As Table 15 displays, there is no significant correlation between reading span and retention tests even in the high WM group exposed to low ECL. These suggest that the relationship between cognitive capacity measures and comprehension measures may depend on the cognitive resources of the participants and the type of comprehension. Moreover, the correlation analyses support Turner and Engle's (1989) claim that the secondary task and the primary task do not need to be the same, which seems to depend upon what kind of a comprehension test is used in a study. Therefore, this study suggests that L2 reading span can be used as a WM index where L2 listening comprehension is under investigation, which depends on type of comprehension.

As a result, hypothesis 3-a, which assumes that participants under low ECL will report statistically lower levels of CL than the ones under high ECL is rejected. In particular, the reverse of this assumption turned out to be true in that participants in T+P (high ECL) condition reported significantly less CL ratings than those in N+P (low ECL) condition. In addition, hypothesis 3-b, which hypothesizes that high WM participants will report statistically lower CL ratings than the ones with low WM, is rejected as well. Specifically, these two groups reported the same amount of CL on average.

To sum up, the participants under low ECL perceived a statistically higher amount of CL than the ones under high ECL in the present study. Following statistical analyses showed that this finding may not be based on either presentation mode or mental effort differences between participants exposed to high ECL and those exposed to low ECL, and that it is safe to use L2 reading span test as an index of WM capacity when L2 listening comprehension is under investigation.

Conclusion

The current chapter presented the results of quantitative analyses conducted in accordance with the research questions and hypotheses stated in the previous chapter.

As for short and long term effects of ECL and English verbal WM capacity on retention of information over time, it was found that the combined effect of time, ECL and WM capacity was statistically significant. Orthogonal contrasts showed that the significant triple interaction had multiple sources: First of all, N+P_HWM group lost significantly less average retention scores between the delayed and immediate retention tests than the combined average retention score difference of the other three groups. N+P_HWM group lost significantly less retention scores than both T+P_HWM group and N+P LWM group on average. Besides, it was found that WM affected retention performance only in N+P condition depending upon time. Likewise, effect of ECL on retention performance was detected in high WM group depending upon the time interval between immediate and delayed tests. Overall, treatment groups' retention scores decreased on the delayed retention test, which shows that time had a deteriorating effect on retention performance of the participants. Finally, only N+P_HWM group did not lose a significant amount of retention scores between the immediate and delayed retention tests.

In terms of immediate and delayed effects of ECL and English verbal WM capacity on transfer of knowledge across time, only time has a significant main effect. This significant main effect indicated that participants performed significantly better on the immediate transfer test on average. Generally, all groups lost transfer scores on the delayed transfer test, which means that time had a deteriorating effect on transfer performance of the participants as well. Moreover, it was the T+P_LWM that lost a statistically significant amount of transfer score between the immediate and delayed transfer tests.

It was ECL whose main effect turned out to be significant on participants' perception of overall CL imposed by the learning environment. N+P groups reported significantly higher CL ratings than the T+P groups, a finding that is in contrast to modality principle of Mayer's (2001) cognitive theory of multimedia learning.

Furthermore, it is worth to highlight a fact about the small sample size of the current study, which led to small power indexes: Leahy et al. (2003) states that "obtaining significant effects using small sample sizes is difficult and only possible with large effects" (p. 414). What Leahy et al. (2003) pinpoints implies that the significant effects found in the present investigation are "large" in the sense that, despite small sample size, the significant combined effect of time, ECL, and L2 verbal WM on retention performance; the significant main effect of time on transfer performance; and the significant effect of ECL on participants' perceptions of CL are reliable, which was confirmed by both Cohen's *d* and eta squared values.

Moreover, one can conceptualize retention and transfer loss scores as forgetting, which is unique to the present investigation. In other words, in terms of retention performance only N+P_HWM did not forget a statistically significant amount of scores over a two-week period of time while it was only T+P_LWM group that did forget a significant amount of scores during the two-week interval between immediate and delayed transfer tests. These results are partly in line with Mayer's (2001) cognitive theory of multimedia learning whereas the finding that N+P group reported a significantly higher CL rating than T+P group is not so. The later implies that in L2 learning contexts; learners' perception of CL may be at odds with existing theoretical assumptions, which warrants further research with larger sample sizes in

L2 learning contexts.

CHAPTER 5

DISCUSSION AND CONCLUSION

The main objective of this study was to examine the modality principle of Mayer's (2001) cognitive theory of multimedia learning in relation to L2 verbal working memory (WM) capacity of advanced English speakers over time. More specifically, the presented study tests modality effect in a university context where English is the medium of instruction. The learning environment consisted of multimedia presentation of an expository L2 passage on the computer. In addition, the participants had very low or no prior knowledge about the topic before the investigation. Another important aspect of the study is that the participants were allowed to read or listen to the treatment text as many times as possible in the total allotted time period of ten minutes before taking the comprehension tests.

Based on the analyses described in detail in the previous chapter, the current chapter will first discuss the findings of the study by research questions. Second, implications will be presented. Finally, the limitations and delimitations of the study will be discussed and suggestions for further research will be provided.

Discussion

Findings of the study in relation to the first research question, which investigated the immediate and delayed effects of extraneous cognitive load (ECL) and WM on retention of information revealed a significant triple interaction of ECL, WM, and time (immediate versus delayed). This triple interaction is discussed below.

The findings indicate no main effect of presentation mode on retention of information from an expository L2 text. On the other hand, the effect of presentation mode becomes significant in relation to WM capacity and time of testing. More specifically, it was found that high WM group in the high ECL (T+P) condition lost significantly more retention scores than high WM group in the low ECL (N+P) condition, which may indicate that the effect of ECL comes up under high WM condition. The theoretical rationale behind this finding would be that low WM participants spend so much of their WM capacity on holding information in memory that their remaining cognitive capacity is not sufficient enough to integrate verbal and pictorial models. On the other hand, high WM participants can more easily hold presented information in memory and have enough cognitive resources to integrate verbal and visual information representations. Considering that the participants in the present study had no prior knowledge of the topic, design effects may be stronger for learners with high WM capacity and no prior knowledge. It is also important to remember at this point that N+P groups rated significantly higher cognitive load (CL) levels than T+P groups, which raises the possibility that audio L2 input causes more CL on even advanced L2 learners.

As for the effect of WM capacity, high WM participants in the low ECL condition lost significantly less retention scores than their low WM counterparts on average. This suggests that WM plays a role in low rather than the high ECL condition. However, there was no significant retention score difference between high WM and low WM participants in the high ECL condition, which indicates that WM capacity did not help under high ECL.

These findings related to the first research question provide additive information on Mayer’s (2001) cognitive theory of multimedia learning; particularly, on the modality assumption which suggests that it is better to present a text in narration and visuals than in words and visuals. It should be remembered that it was the high WM group under low ECL that lost the lowest retention scores on average. Besides, the retention performance difference (immediate- delayed) of the high WM participants under low ECL was significantly different from that of the high WM group under high ECL. These findings suggest that the modality effect is observed in relation to L2 WM capacity and time of testing in terms of retention of knowledge from an L2 passage that has high element interactivity (high intrinsic cognitive load).

When it comes to the findings related to the second question investigating the immediate and delayed effects of ECL and WM on transfer of information, transfer scores decreased significantly in time regardless of L2 WM capacity and of the amount of ECL caused by the presentation mode, which is not in line with the modality principle. In addition, as on the retention scores, all treatment groups lost scores on the delayed transfer test. This suggests that independent of the level of ECL and L2 verbal WM capacity, participants in this study performed worse on the delayed transfer test and that time has a significant deteriorating effect on transfer of information.

The results of the present study in regard to both retention and transfer do not completely support the findings of several other studies investigating the modality effect (e.g., Leahy et al., 2003). These studies suggest that learning is more efficient when multiple sensory pathways are used to present information to the learners. A number of studies showed that spoken text improved learning from different perspectives: 1) lower mental effort and higher scores on tests (Tindall-Ford, Chandler, & Sweller, 1997); 2) less time spent on the following problem solving (Jeung, Chandler, & Sweller, 1997; Mousavi et al., 1995); and 3) better scores on retention, transfer, and matching tests (Mayer & Moreno, 1998; Moreno & Mayer, 1999; Kalyuga, Chandler, & Sweller, 1999, 2000). Mousavi et al. (1995) found that participants performed better when presented with a dual modality: visual diagram and auditory explanations. Likewise, Velayo and Quirk (2000) found that participants receiving visual-auditory information outperformed those exposed to other mixed-modality presentation modes. As an explanation to these findings, Mayer and Moreno (1998) claim that when information is presented in different modalities, the participants had more room in their cognitive architecture to retain and process the information presented. On the other hand, claiming that the modality effect found in previous studies can be attributed to the fact that in visual-only presentation the learners have to split-attention between pictorial information and textual information, Tabbers (2002) argues that modality effect holds true in system-paced instructions whereas it does not or even there can be a reverse effect in learner-paced learning environments. The current investigation led to additional insights into the findings of these studies: it suggests that even though modality effect may disappear in learner-paced instructions in the short-term, there may be a recovery of the effect in the long term, both confirming and expanding Tabbers' (2002) finding. Moreover, Tabbers' (2002) study included instructions on an instructional design strategy in L1 Dutch that were presented for at least twenty minutes even in the fastest tasks. Therefore, the results of the present study also expands Tabbers' (2002) finding to L2 learning, to a different subject matter (tornado formation), and to a slower presentation pace (at most ten minutes).

As for the effect of WM, the results do not support those of Leeser (2007) who found that the effects of WM depend upon prior knowledge. More specifically, Leeser (2007) argues that WM capacity differences affect reading comprehension only when learners have a reasonable amount of prior knowledge about a topic in question. However, the current study pinpoints that L2 verbal WM capacity may still affect retention performance in the case of novel information processing. This means that regardless of prior knowledge L2 verbal WM capacity may affect L2 comprehension. Similarly, both Hambrick and Engle (2002) and Hambrick and Oswald (2005) suggest that WM capacity may affect higher-level cognitive task performance independent of domain knowledge.

While reading all the contrasts between the present study and previous research, it should be kept in mind that most of the previous research tested the effects of presentation mode or WM capacity on immediate learning outcomes. On the other hand, the present study examines both short-term and long-term effects of L2 verbal WM capacity and modality. It should be noted that a two way ANOVA with ECL and WM as the independent variables conducted on the immediate retention test revealed no significant results, all p's > .05. An analogous ANOVA was also conducted on the immediate transfer test as well. Results indicated no significant effects either: all p's > .05. These results are in contrast with those of previous research most of which is either system-paced or with paper-based instructions where time-on-task was limited to the total time of the narration (e.g., Kalyuga et al., 1999, Mayer & Moreno, 1998, Moreno & Mayer, 1999). However, the results of the present study are in line with Tabbers (2002) claiming that modality effect does not hold in learner-paced instructional designs.

An interesting finding was obtained with regard to the third research question that sought to examine the effects of presentation mode and WM on advanced L2 learners' perception of CL over time. Participants were tested by immediate and delayed administrations of a subjective CL scale. Interestingly enough, participants in the N+P condition that is assumed to impose less ECL on WM reported statistically higher CL ratings than the ones in the T+P condition.

One possible explanation for this finding is that it may stem from the nature of the N+P design instead of the presentation mode of the instruction: The audio material consisted of a website and a winamp sound player. The audio record was not embedded into the webpage, meaning that participants had to control both the website to view static pictures of tornado formation and the winamp panel to listen to the text (see Appendix I). In other words, the synchronization of the listening text with the corresponding pictures had to be done by the participants, which could have increased the CL imposed because participants had to go back and forth between the web screen and the winamp control panel. However, statistical analyses conducted on the fifth item of the CL scale that asked participants to rate how much effort they invested to navigate through the text showed that this explanation does not hold true because participants in the N+P group did not spend more effort on navigating through the text compared to those in the T+P group.

Another possibility is that participants in the N+P groups spent more mental effort in order to meet the higher demands of L2 listening compared to L2 reading in order to understand the L2 text better. After all, all participants took the treatment without any schema on tornado formation, which requires use of more cognitive capacity on all participants as Paas (1992) states:

> (…) the component processes of any task might individually require less cognitive capacity if better schemata are acquired or if more rules are automated. (…) As a result of limited cognitive capacity less effort could be invested in more relevant processes, such as abstracting appropriate schemata (p. 433).

On the other hand, statistical analyses conducted on the first item of the CL scale that asked participants to rate how much mental effort was required revealed that the amount of mental effort N+P and T+P participants reported did not differ significantly from each other, which shows that this option cannot explain significantly higher CL ratings of the participants in N+P group.

Therefore, the remaining possible explanation for the finding that participants exposed to low ECL reported higher CL than those exposed to high ECL is that listening or audio information processing in an L2, rather than L2 reading, may be more difficult and challenging for L2 learners, which is in line, to some extent, with L2 studies conducted on the effects of text mode (reading versus listening) where readers comprehended more of the represented information than listeners (Greenslade, Bouden, & Sanz, 1999, as cited in Leeser, 2004; Lund, 1991; Wong, 2001; Leeser, 2004). Likewise, L1 studies investigating differences between listeners and readers pinpoint that the advantages readers have over listeners stem from “the greater control over rate of processing as well as their ability to notice word, sentence, and paragraph boundaries” (Anderson, 1980; Danks, 1980; Ferreira & Anes, 1994, as cited in Leeser, 2004, p. 590). As such, the higher CL rating of N+P participants could be due to the assumption that they needed to review the text more than the participants in T+P group because of the qualitative differences between L2 reading and L2 listening. Leeser (2004) argues that when greater control over processing resources comes together with the claim that L2 learners more easily detect word and sentence boundaries in reading since reading texts are more clearly segmented, this might explain the effect of text mode on comprehension. All L2 researchers cited above based their findings on the assumption that processing resources are less constrained in the written mode than in the audio mode. By contrast, Taub and Kline (1976, 1978) reported that younger and older adults who spoke English as a primary language remembered more information from visually presented prose than from aural presentation of prose, and the researchers attributed this finding to the fact that in visual presentation the passage was available for reviews, and when they eliminated the chance of review, there were no differences between the two modalities (as cited in Kim, 2006, p.22). This further suggests that since both the participants who listened to and read the L2 text in the present study had the chance of reviewing, participants in N+P condition might have needed to spend more time on reviewing because of the difficulties of L2 listening, which could explain their higher CL ratings. The ten-minute instructional time and learner-paced nature of the instruction could have allowed this.

Related to the qualitative differences between reading and listening, another option is that there could be another dimension of text presentation that could have led to ECL on learners: Dillon (1992) argues that splitting sentences across screens may lead to extra burden on the limited WM capacity, by forcing learners to keep previous information in their WM while processing upcoming information. Likewise, McCrudden et al. (2004) claim that sentence-by-sentence presentation of a text body on a computer screen imposes more storage demands, thus making text learning more difficult. In the same line of logic, in the T+P condition of the present study, the whole text body is shown on a screen not splitting sentences across screens. On the other hand, it is important to note that even though sentences of the text in T+P presentation were presented on one webpage, participants still had to scroll down and up if they needed because every sentence had (except for damage scale levels and the first sentence) its own corresponding picture between itself and the previous one together with a black line placed above the pictures. In other words, every sentence and its corresponding picture in T+P condition was presented in a way that readers could only see one sentence and its corresponding picture on the computer screen without being able to see the next sentence or its picture. The rationale behind such a design was to ensure equivalency between T+P and N+P conditions as much as possible because, in listening, there is a sentence-by-sentence presentation of a text because of the nature of the listening task, thus listeners cannot listen to all of the sentences at the same time. In such a presentation, when learners finish listening to a sentence than the presentation mode urge them to hear next upcoming sentence. Thus, in order to build coherence among sentences listeners have to keep the previous sentences active in their WM while paying attention to next sentences. However, in reading, all of the sentences could be in the visual field of the learners or more easily accessible, so a reader could go back more easily and review a previous sentence upon need. Even though all of the sentences were not in the visual field of the participants as a whole in the T+P condition of the current study, it seems that it was easier in the T+P condition to review than in the N+P condition, because sentences and pictures were presented on the same webpage in the T+P condition, which was not the case in the N+P condition. For this reason, rather than sentence-by-sentence presentation, the higher CL rating of N+P group might have stemmed from the inner characteristics of L2 listening that may increase ECL, thus leading participants in N+P condition to review the text more. Table 16 represents the possible sources of ECL for T+P and N+P groups in the current study as based on previous research:

Table 16. Possible Sources of ECL for T+P and N+P Groups

ECL Variable	T+P	N+P
Visual Channel Load (Mayer, 2001)	1	0
Text Organization (McCrudden et al., 2004)	0	0
Text Presentation (Whole-text vs. sentence-by-sentence) (McCrudden et al., 2004)	1	1
Segmentation among word, sentence, and paragraph boundaries (in terms of noticing and perception) (Anderson, 1980; Danks, 1980; Ferreira & Anes, 1994, as cited in Leeser, 2004, p. 590)	0	1
Control over rate of processing (Anderson, 1980; Danks, 1980; Ferreira & Anes, 1994, as cited in Leeser, 2004, p. 590)	0	1
Control of Presentation (computer vs. learner-paced) (Mayer & Chandler, 2001)	0	0
Total	2	3

In Table 16, each possible source of ECL was given 1 point if it is considered to be imposing load on T+P or N+P groups, and if it is not, no point was given (0). The small-scale instrument shown in Table 16 was developed by the researcher. As the total load scores suggest, L2 listening seems to include more text variables that may contribute to increased ECL. Therefore, as based on previous research (Leeser, 2004; Lund, 1991; Wong, 2001), it is not unreasonable to argue that L2 listening imposes more load than L2 reading.

In conclusion, the current study yielded that it is the combined effect of the presentation mode of an L2 text, L2 verbal WM capacity and time, which has impacts upon retention of information from an expository L2 text. As for transfer of knowledge, this study showed no effects of either ECL or WM and provided evidence that the time interval between immediate and delayed transfer tests has negative affects. Finally, the present study showed that advanced L2 learners' perception of the overall CL imposed by an L2 reading or listening passage depends upon the presentation mode of the whole text body. Specifically speaking, advanced L2 learners in the low ECL condition reported significantly higher overall CL ratings than the ones in high ECL. This may stem from the qualitative differences between L2 reading and L2 listening: Listening exists in time whereas reading occurs in space, which seems to make it easier for readers to look for contextual clues, go back and forth, and detect units of ideas. All these imply that listening may become more problematic even for advanced L2 learners than reading, even though they can achieve more permanent success in listening tasks that are supported by visual aids.

Implications

This study suggests important implications for measuring L2 comprehension and for the design and choice of multimedia presentations of L2 texts and materials particularly in computer assisted language learning and L2 education in general. The implications of the study also refer to the subjective measurement of overall CL, and to the relationship between L2 reading span and L2 comprehension. All the implications will be stated with respect to multimedia learning environments on the computer.

In multimedia environments, it is reasonable to divide reading and listening comprehension into retention and transfer of information, since this study showed that under certain conditions the factors affecting retention and transfer of information from reading or listening to an L2 text are different. In other words, in terms of retention, even though presenting L2 texts as a listening task with pictorial information may not promote better comprehension in the short-term and involve more ECL, considering that when given the chance of reviewing, high WM participants in N+P group lost fewer retention scores than high WM participants in T+P group, L2 listeners can outperform L2 readers in the long-term. This design benefit does not appear to promote transfer of knowledge that was found to be affected negatively by time of testing. All these suggest a thorough reflection on question types employed in L2 comprehension tests.

As for the design and choice of L2 texts and materials to be used in L2 education, cognitive load theory (Sweller, 1988) and cognitive theory of multimedia learning (Mayer, 2001) may provide the practitioners with a well-structured guidance on when and how to use which material with a certain population of learners. For instance, the current study implies that L2 texts do not have to be quite long to be difficult for learners: A short L2 text consisting of a fair number of interdependent sequences of information may also be challenging for learners. Therefore, any amount of cognitive load that can be caused by not only the presentation mode but also inner characteristics of an L2 material should be taken into account before using it.

Moreover, in terms of subjective CL rating or measurement, the present study shows that time does not appear to affect perception of CL, which suggests that people are able to remember a certain amount of CL to which they were previously exposed, at least two weeks after the first test and three weeks after reading or listening to the text. All these imply that subjective CL measurement in an L2 is reliable and valid when the presentation mode of learning materials is taken into account, thus they are safe to use and to index CL.

Finally, the current results indicated that L2 reading span can be used as a WM index when L2 listening comprehension is investigated, which highly depends on cognitive capacity of L2 learners and type of comprehension. In other words, it is better to take WM capacity and type of comprehension into account while investigating the relationship between cognitive capacity and language comprehension in general.

Limitations and Recommendations for Future Research

The major limitation of the study was, as indicated by the low observed power values, sample size. All the observed power indexes were lower than the desirable value of .80. Moreover, the participants were 29 advanced level EFL speakers attending an English-medium university in Turkey. Thus, in order to generalize the findings of the study to a larger target population, this study should be replicated with more participants having different levels of English proficiency and prior knowledge in other learning contexts encompassing different levels of intrinsic CL through both cross-sectional and longitudinal studies.

Second, the N+P presentation in this study did not embed audio material (winamp). More specifically speaking, the winamp control panel was not embedded into the N+P webpage. So, the participants in this group had to control two existing materials: winamp panel and N+P webpage including corresponding pictures. Inevitably, this design made the participants synchronize what they were listening with its corresponding pictures, and go back and forth between the winamp panel and the website. Future research should take this into account and test L2 comprehension under N+P conditions through more intricately designed multimedia materials.

Third, since it was summer term, and because of the space limitations in the computer laboratory, the suitable time schedule of every each participant changed from one to another. So, the participants took both the immediate and the delayed tests on three successive sessions. In order not to let them share the content of the text and the questions, the participants were asked not to talk about both the treatment and test questions with other people enrolling in the study. In order to better escape such a threat to the internal validity of the study, the participants should either be taken all together into the experiment, or all participants should participate in one day in successive groups and the experimenter should make sure that no previous group have the chance to get in contact with the upcoming groups. For example, a group that took the treatment could be moved into another place where they take the questions and the next group is taken into the laboratory and receives the treatment and so on.

Fourth, the subjective CL scale was applied after the treatment and comprehension tests, and it also included CL rating on questions included in the tests. In other words, subjective CL ratings of the participants do not purely reflect their perceived amount of ECL involved only in the presentation mode of the L2 text but the whole multimedia learning environment. Therefore, we need further research that investigates whether L2 learners exposed to narrated text with corresponding pictorial information report higher CL ratings than the ones exposed to written text with corresponding pictures. To do so, participants should be given a CL scale right after the presentation of the text body.

Fifth, even though the instructional time of reading or listening to the treatment was limited to ten minutes, participants could have read or listened as many times as needed simply because the treatment was quite a short text (a 254-word text). So, the results of the study are constrained to a certain amount of instructional time not to reading or listening time, which further lessens the generalizability of the study. We need future research that compares the situations of constrained reading or listening time (e.g., just once).

Finally, comprehension tests used in the present study are not general comprehension indexes but specific comprehension tests including retention and transfer questions. Therefore, whether the significant correlations found between transfer tests and reading span in the high WM group exposed to low ECL can be found between general comprehension and reading span warrants further research.

Delimitations

The current study was delimited to intact groups of 29 advanced L2 English teaching university students who have English as the medium of education at their home university and operated under the assumption that the students participating could read or listen to and comprehend the tornado formation text and the reading comprehension questions and answer them as honestly and accurately as possible and that they participated in the study willingly without lacking the motivation to read or listen to how a tornado forms, and that the learning environment was comfortable with no high noise or heat that could increase CL.

Conclusion

The results of the present study do not defer to the modality principle of Mayer's (2001) cognitive theory of multimedia learning theory to the full extent. However, the finding that high WM group in N+P condition lost significantly less retention scores than high WM group in T+P condition and the other three treatment groups, and that only low WM participants in T+P condition lost a significant amount of transfer of information extends modality effect in a self-paced L2 learning environment. In other words, this study indicates that not only the presentation mode, namely, the level of ECL but also individual differences like L2 verbal WM capacity and time of testing affect retention of knowledge while it is only time that has effects on transfer of information.

In addition, the present results indicate that the modality effect may show up even in learner-paced multimedia environments over time, which seems to be directly important for forgetting of information. All these imply that time of testing should be considered as an important factor for research design since it appears to be a highly relevant factor interacting with WM capacity and modality effect.

This study also shows that subjective CL scales are easy and reliable to implement. Furthermore, it was shown that the qualitative differences between L2 reading and L2 listening may lead to more reviewing, which affects L2 learners' perception of CL. These suggest that qualitative differences between L2 reading and listening should be taken into account while investigating design guidelines.

Overall, this study implies that the assumptions of the cognitive theory of multimedia learning (Mayer, 2001) and CL theory (Sweller, 1988) can be limited to native speakers and to system-paced research designs on immediate learning outcomes to a certain extent, and that we may be in need of different approaches for non-native learners in learner-paced L2 learning contexts that examine not only immediate but also delayed comprehension results. In other words, the modality effect can even be traced in self-paced L2 multimedia learning environments depending on time of testing, type of comprehension, and WM capacity.

Appendix A

Participant Profile

Name:
Age:
Gender:
First Language:
Department:

1) How many years/semesters have you spent at Boğaziçi University School of Foreign Languages (YADYOK)?

2) How would you rate your use of WWW?

1) not at all effective 2) slightly effective 3) effective 4) quite effective 5) very effective

3) How would you rate your use of a computer?

1) not at all effective 2) slightly effective 3) effective 4) quite effective 5) very effective

4) For how many hours per week do you use the following computer programs & applications?

	Hours of use per week
a) Internet	__________
b) Microsoft Word	__________
c) Microsoft FrontPage	__________
d) Microsoft PowerPoint	__________
e) Microsoft Excel	__________
f) Microsoft Outlook	__________
g) Microsoft Publisher	__________
h) Dreamweaver	__________

5) What was your BUEPT score? (Please circle your score)

A B C

6) How many undergraduate courses have you taken so far?

____

7) You are a ______ :

a) freshman b) sophomore c) junior d) senior

8) How many semesters have you spent at your undergraduate study so far? (Please circle your answer)

1 2 3 4 5 6 7 8 9 10

Appendix B

Pre-Reading/Listening

HOW A TORNADO FORMS[1]

SOME EXTREME FACTS ABOUT TORNADOES:

*Although tornadoes have been observed on every continent except Antarctica, most occur in the United States.

*The most extreme tornado in recorded history was the Tri-State Tornado which roared through parts of Missouri, Illinois, and Indiana on March 18, 1925.

*The deadliest tornado in world history was the Daultipur-Salturia Tornado in Bangladesh on April 26, 1989, which killed approximately 1300 people.

*The highest wind speed ever measured in a tornado, which is also the highest wind speed ever recorded on the planet, is 301 ± 20 mph (484 ± 32 km/h) in the Moore, Oklahoma tornado.

*Although studied for about 140 years and intensively for around 60 years, there are still aspects of tornadoes which remain a mystery.

*Reliably predicting tornado intensity and longevity remains a problem, as do details affecting characteristics of a tornado during its life cycle and tornadolysis.

*Scientists still do not know the exact mechanisms by which most tornadoes form, and occasional tornadoes still strike without a tornado warning being issued.

MYTHS and MISCONCEPTIONS:

*One of the most persistent myths associated with tornadoes is that opening windows will lessen the damage caused by the tornado.

*Another commonly held belief is that highway overpasses provide adequate shelter from tornadoes. On the contrary, a highway overpass is a dangerous place during a tornado.

*An old belief is that the southwest corner of a basement provides the most protection during a tornado.

DO NOT FORGET:

*As a general rule, no area is "safe" from tornadoes, though some areas are more susceptible than others.

[1] *Source: wikipedia*

Appendix C

A Screenshot from Video Shown Prior to Treatment

Appendix D - Comprehension Tests

Retention Test

Name:

Please answer the questions below. You may answer in either English or Turkish or both. You have 20 minutes to answer all the questions.

1) Please write down an explanation of how a tornado forms. Pretend that you are writing to someone who does not know much about tornadoes.

2) Match the example damages on the left with their corresponding damage scale.

1) Some trees blown over	_________	F3
2) Strong buildings swept away	_________	F0
3) Heavy cars thrown	_________	F5
4) Well constructed buildings destroyed	_________	F1
5) Moving cars blown off roads	_________	F4
6) Mobile homes demolished	_________	F5
		F6

3) What is the damage scale of a tornado moving with a speed of <u>117-180kms?</u>

a) Considerable damage
b) Incredible damage
c) Moderate damage
d) Severe damage

4) What is the damage scale of a tornado that has an F4 damage scale level?

a) Incredible damage
b) Devastating damage
c) Severe damage
d) Considerable damage

Appendix D.01

Answer Key to the Retention Test

Retention Test:

1) Please write down an explanation of how a tornado forms. Pretend that you are writing to someone who does not know much about tornadoes.

2) Intense or unseasonable heat warms up the ground
3) ground temperature increases
4) moist air heats and starts to rise
5) moist air meets cold dry air and it explodes upwards
6) a storm quickly develops
7) Winds from different directions cause rising air to rotate
8) A visible cone or funnel drops out of the cloud towards the ground

2) Match the example damages on the right with their corresponding damage scale.

1) Some trees blown over	F0	F3
2) Strong buildings swept away	F5	F0
3) Heavy cars thrown	F3	F2
4) Well constructed buildings destroyed	F4	F1
5) Moving cars blown off roads	F1	F4
6) Mobile homes demolished	F2	F5
		F6

3) What is the damage scale of a tornado moving with a speed of 117-180kms ?

a) Considerable damage
b) Incredible damage
c) Moderate damage
d) Severe damage

4) What is the damage scale of a tornado that has an F4 damage scale level?

e) Incredible damage
f) Devastating damage
g) Severe damage
h) Considerable damage

Appendix D.02

Transfer Test

Name:

Please answer the questions below. You may answer in either English or Turkish or both. You have 30 minutes to answer all the questions.

1) What cause(s) a tornado?

2) Suppose you see a storm together with rain, thunder and lightning but no tornado. Why/How can this happen?

3) What could you do to be safe at home during a tornado with a level of severe damage?

4) How is air temperature related with a tornado?

5) What determine(s) the duration and the strength of a tornado?

6) List any number of environmental clues to watch out for a tornado?

Appendix D.03

Answer Key to the Transfer Test

1) What cause(s) a tornado?
Ans: unseasonable, intense heat
Ans: rising warm air rotated by strong winds blowing off in opposite directions
Ans: direction of strong winds

2) Suppose you see a storm together with rain, thunder and lightning but no tornado. Why/How can this happen?
Ans: lack of strong winds
Ans: winds blowing off in the same directions

3) What could you do to be safe at home during a tornado with a level of severe damage?
Ans: go to the shelter/storm cellar
Ans: being far away from the walls and windows.

4) How is air temperature related with a tornado?
Ans: It heats up the ground
Ans: The earth becomes hot
Ans: Moist warm air heats and starts to rise rapidly
Ans: Cold dry air is punctured by the rising moist warm air

5) What determine(s) the duration and the strength of a tornado?
Ans: The size of the vortex of winds
Ans: The speed of the winds

6) List any number of environmental clues to watch out for a tornado?
Ans: unseasonable, intense heat
Ans: rain
Ans: storm
Ans: strong winds
Ans: thunder
Ans: lightning
Ans: warm, moist (rising) air

Appendix E

Prior Knowledge Test

Name:

You are going to read or listen to an English passage on how tornadoes form. The researcher would like to know what you already know about this topic before you read or listen to the passage. Please answer the following questions. You may answer in either English or Turkish or both.

Part I.

1) Have you ever experienced a tornado? If yes, please indicate where and when?

2) Please write down anything you know about tornadoes or how a tornado forms?

Part II.

1) Under what condition(s) do tornadoes form?

2) In what way(s) are tornadoes related to winds?

3) What damage scale is used to estimate the speed of a tornado?

4) How does ground temperature contribute to tornado formation?

Appendix E.01

Answer Key to the Prior Knowledge Test

Part I.

1) Have you ever experienced a tornado? If yes, please indicate where and when?

2) Please write down anything you know about tornadoes or how a tornado forms?

1) Intense or unseasonable heat warms up the ground
2) ground temperature increases
3) moist air heats and starts to rise
4) moist air meets cold dry air and it explodes upwards
5) a storm quickly develops
6) Winds from different directions cause rising air to rotate
7) A visible cone or funnel drops out of the cloud towards the ground

Part II.

1) Under what condition(s) do tornadoes form?

Intense or unseasonable heat

2) In what way(s) are tornadoes related to winds?

1) Strong and rapid winds blowing off in opposite directions
2) Such winds rotate the rising moist and warm air

3) What damage scale is used to estimate the speed of a tornado?

Fujita damage scale

4) How does ground temperature contribute to tornado formation?

As the ground temperature increases, moist air heats and starts to rise (Moist air near the ground heats because of the high ground temperature)

Appendix F

Paas' (1992) Cognitive Load Scale

The *"Rating Scale"* below is the instrument used to gauge the cognitive load imposed by the multimedia condition in question. No information got from the questionnaire will be shared with a third party. If you are not comfortable with this or with the study in general, please do not fill in the questionnaire. You are free to do so. Whether you choose to do it or not, thanks in advance for taking your time.

For each statement below, you are asked to indicate on a 9 point scale, the degree to which you agree or disagree with the item. Please mark the appropriate number which most accurately reflects your level of agreement or disagreement. Please answer the questions as naturally and honestly as possible, in a way that shows how the task really was, not how you think the task should have been. The first answer that pops into your mind is what is looked for, so try not to spend too much time thinking about your answers.

Name:

1. In reading or listening to the preceding text I invested :

1 Very very low mental effort	2	3	4	5 Neither low nor high mental effort	6	7	8	9 Very very high mental effort

Rating Scale based on Paas (1992)

Appendix G

Subjective Cognitive Load Scale

The *"Rating Scale"* below is the instrument used to gauge the cognitive load imposed by the multimedia condition in question. No information got from the questionnaire will be shared with a third party. If you are not comfortable with this or with the study in general, please do not fill in the questionnaire. You are free to do so. Whether you choose to do it or not, thanks in advance for taking your time.

For each statement below, you are asked to indicate on a 7 point scale, the degree to which you agree or disagree with the item. Please mark the appropriate number which most accurately reflects your level of agreement or disagreement. Please answer the questions as naturally and honestly as possible, in a way that shows how the task really was, not how you think the task should have been. The first answer that pops into your mind is what is looked for, so try not to spend too much time thinking about your answers.

Name:

1. How much mental effort was required (e.g. thinking, deciding, remembering etc.)?

1	2	3	4	5	6	7
not at all						very much

2. Was the reading/listening text demanding?

1	2	3	4	5	6	7
not at all						extremely demanding

3. How hard was it to understand the content of the text?

1	2	3	4	5	6	7
not at all						extremely hard

4. How successful do you think you were in your attempt to understand the content of the text?

1	2	3	4	5	6	7
not successful						very successful

5. How much effort did you invest to navigate through the text (scrolling up and downward, controlling winamp)?

1 2 3 4 5 6 7
not extremely high
at all

6. How insecure, discouraged, irritated and annoyed did you feel while reading/listening to the text?

1 2 3 4 5 6 7
not extremely
at all

7. While reading/listening to the text, how difficult was it to understand the tornado formation process?

1 2 3 4 5 6 7
not extremely difficult
at all

8. How difficult was it for you to answer the questions that followed?

extremely extremely
easy difficult

1 2 3 4 5 6 7

Appendix H

NASA-TLX Work Load Scale

NAME	TASK	DATE

TASK LOADING INDEX

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

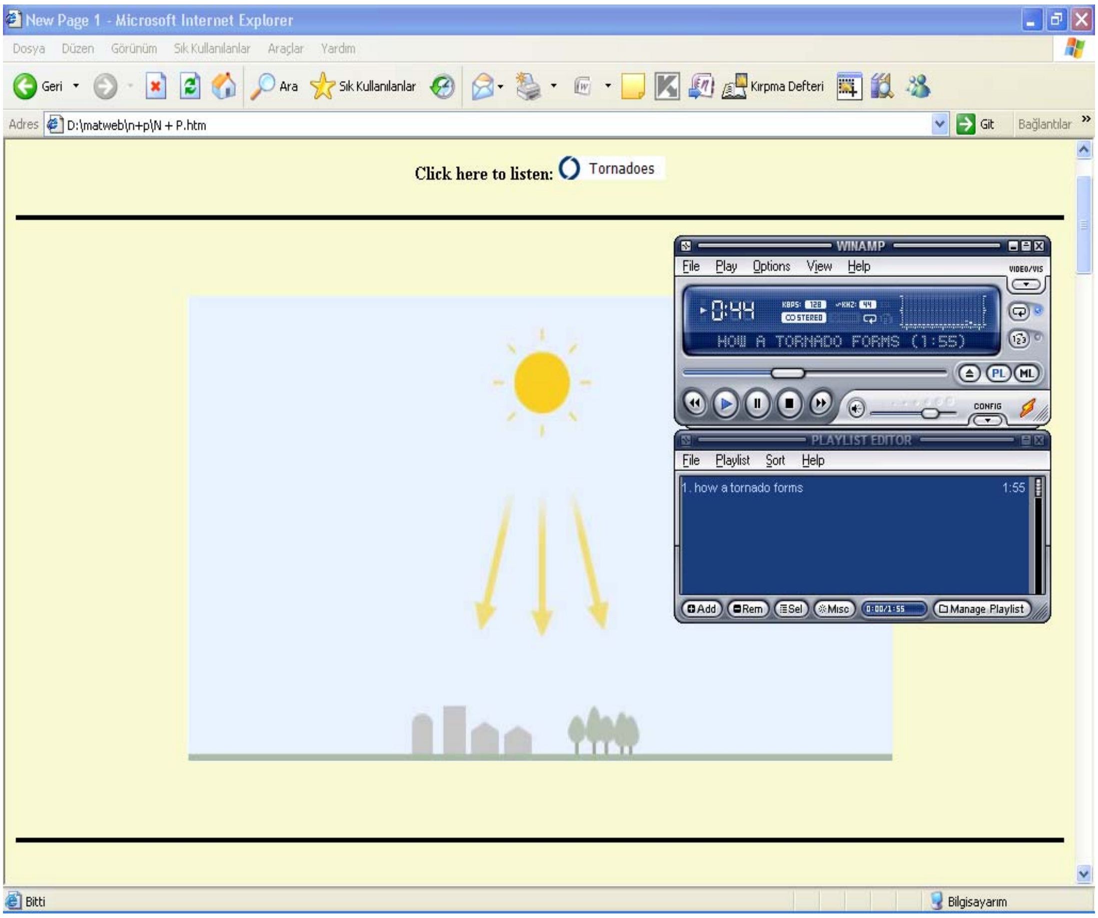
Very Low — Very High

Frustration — How insecure, discouraged, irritated and annoyed were you?

Very Low — Very High

Appendix I

A Screenshot from Audio-visual Presentation



Appendix J

The Text

HOW A TORNADO FORMS

Although no two tornadoes are the same, they need certain conditions to form - particularly intense or unseasonable heat. As the ground temperature increases, moist air heats and starts to rise. When the warm, moist air meets cold dry air, it explodes upwards, puncturing the layer above. A thunder cloud may begin to build. A storm quickly develops - there may be rain, thunder and lightning. Upward movement of air can become very rapid. Winds from different directions cause it to rotate. A visible cone or funnel drops out of the cloud towards the ground. The vortex of winds varies in size and shape, and can be hundreds of meters wide. A tornado can last from several seconds to more than an hour and may travel dozens of miles. Winds within the tornado may be so fast they cannot be properly measured. Instead, the Fujita damage scale is used to estimate speed:

F0 (0-117kms) - Light damage: Some damage to chimneys. Branches broken from trees and some trees blown over.
F1 (117-180kms) - Moderate damage: Moving cars blown off roads, mobile homes overturned, or pushed off their foundations.
F2 (181-253kms) - Considerable damage: Mobile homes demolished, large trees snapped or uprooted, cars lifted off the ground.
F3 (254-332kms) - Severe damage: Trains overturned, most trees uprooted, heavy cars thrown, walls of homes destroyed.
F4 (333-419kms) - Devastating damage: Well constructed buildings destroyed, large objects thrown.
F5 (420-512kms) - Incredible damage: Cars thrown more than 100 meters, strong buildings swept away.

Total words in passage: 254

Total sentences: 26

Average words per sentence: 10 (approximately)

Appendix K

Letter of Permission

-------- Original Message --------
Date: Wed, 21 May, 2008 16:58
From: "Paas, Fred" < Fred.Paas@ou.nl>
To: kadir.kozan@boun.edu.tr
Subject: RE: Asking for Permission to use your 1992 Cognitive Load Scale
Part(s): 2 mental effort rating scale2.doc

Dear Kadir,
Sure, in the attachment you can find the scale.
Good luck with your studies and I look forward to seeing the results.
Best,
Fred

Appendix L

Committee on Ethical Conduct in Human Research (İNAREK) Approval

Boğaziçi Üniversitesi
İnsan Araştırmaları Etik Kurulu

13 Mayıs 2008

Sn. Kadir Kozan
Boğaziçi Üniversitesi
Yabancı Diller Eğitimi Bölümü
Bebek - İstanbul

Sn. Kozan,

"Harici bilişsel yükün yoğun içsel-bilişsel yük taşıyan çoklu bir öğrenme ortamında ikinci dil okuduğunu anlama başarısı üzerine etkileri." başlıklı projeniz ile ilgili olarak Boğaziçi Üniversitesi İnsan Araştırmaları Etik Kurulu'na yapmış olduğunuz başvuru (Protokol no: 2008/21) kurulumuzun 13 Mayıs 2008 tarih ve 2008/02 sayılı toplantısında değerlendirilerek uygun bulunmuştur. Bilgilerinize sunarız.

Doç. Dr. Ali İ. Tekcan
Başkan

REFERENCES

Annett, J. (2002). Subjective rating scales: science or art? *Ergonomics, 45*(14), 966-987.

Ariew, R., & Erçetin, G. (2004). Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning Journal, 17*(2), 237-259.

Baddeley, A. (1992). Working memory. *Science, 255,* 556-559.

Baddeley, A., & Gathercole, S. E. (1993). *Working Memory and Language.* East Sussex: Lawrence Earlbaum.

Bannert, M. (2002). Managing cognitive load-recent trends in cognitive load theory. *Learning and Instruction, 12,* 139-146.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science 32,* 115-132.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*(2), 109-119.

Burkes, K. M. E. (2007). *Applying cognitive load theory to the design of on-line learning.* Unpublished doctoral dissertation, University of North Texas. Retrieved August 12, 2008, from http://digital.library.unt.edu/permalink/meta-dc-3698:1

Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. *Journal of Educational Psychology, 95*(3), 629-640.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chen, S. Y., & Macredie, R. D. (2002). Cognitive styles and hypermedia navigation: Development of a learning model. *Journal of the American Society for Information Science and Technology, 53,* 3-15.

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

Clark, R. C., Nguyen, F., & Sweller, J. (2005). *Efficiency in learning: Evidence-based guidelines to manage cognitive load.* San Diego, CA: Pfeiffer.

Coe, R. (2002). *It's the Effect Size, Stupid What effect size is and why it is Important*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 12-14 September 2002. Retrieved May 12, 2009, from http://www.leeds.ac.uk/ educol/documents/ 00002182.htm

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

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

Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics, 35*(10), 1297-1326.

Dixon, P. (1991). From research to theory to practice: Commentary on Chandler and Sweller. *Cognition and Instruction, 8*(4), 343-350.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19-23.

Field, A. (2000). *Discovering Statistics using SPSS for Windows: Advanced techniques for the beginner.* London: Sage.

Field, A., & Hole, G. (2003). *How to design and report experiments.* London: Sage.

Gellevij, M., van der Meij, H., de Jong, T., & Pieters, J. (2002). Multimodal versus unimodal instruction in a complex learning context. *The Journal of Experimental Education, 70*(3), 215-239.

Ginns, P. (2005) Meta-analysis of the modality effect. *Learning and Instruction, 15,* 313-331.

Grace-Martin, M. (2001). How to design educational multimedia: A "loaded" question. *Journal of Educational Multimedia and Hypermedia, 10*(4), 397-409.

Grimley, M. (2007). Learning from multimedia materials: The relative impact of individual Differences. *Educational Psychology, 27*(4), 465-485.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44,* 339-387.

Hambrick, D. Z., & Oswald, F. L. (2005). Does domain knowledge moderate involvement of working memory capacity in higher-level cognition? A test of three models. *Journal of Memory and Language, 52,* 377-397.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14,* 25-38.

Heo, M., & Chow, A. (2005). The impact of computer augmented online learning and assessment tool. *Educational Technology and Society, 8*(1), 113-125.

How a tornado is formed. (n.d.). Retrieved March 4, 2008, from BBC site: http: //downloads.bbc.co.uk/news/nol/shared/spl/hi/sci-nat/06/tornado/pdf/how_ tornado_is_formed.pdf

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17,* 329-343.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*(3), 558-568.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Cognitive Psychology, 13,* 351–371.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92,* 126-136.

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(3), 567-581.

Kane, M. J., Conway, A. R. A., & Engle, R. (1999). What do working memory tests really measure? *Behavioral and Brain Sciences, 22*(1), 101-102.

Kim, E. S. (2006). *Effects of presentation modality on language processing in young and older adults.* Retrieved from ProQuest Dissertations & Theses (PQDT). (AAT 3216322)

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12,* 1-10.

Kramsch, C., & Andersen, R. W. (1999). Teaching text and context through multimedia. *Language Learning and Technology, 2,* 31-42.

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology,17,* 401-418.

Lee, H., Plass, J. L., & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology, 98*(4), 902-913.

Leeser, M. J. (2004). The effects of topic familiarity, mode, and pausing on second Language learners' comprehension and focus on form. *Studies in Second Acquisition, 26,* 587-615.

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: topic familiarity and working memory. *Language Learning, 52*(2), 229-270.

Lund, R. (1991). A comparison of second language listening and reading comprehension. *Modern Language Journal, 75,* 196-204.

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics, 44,* 229-243.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88,* 49-63.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

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

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

Mayer, R. E., & Moreno, R. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117-125.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638-643.

McCrudden M., Schraw, G., Hartley, K., & Kiewra, K. A. (2004). The influence of presentation, organization, and example context on text learning. *The Journal of Experimental Education, 72*(4), 289-306.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22,* 149-158.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156-163.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87,* 319-334.

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society, 30*(4), 287-289.

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*(2), 117-118.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84,* 429-434.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32,* 1-8.

Paas, F., Tuovinen, J. E., Tabbers, H., &Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Paas, F., van Merriënboer, J. J. & Adam, J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79,* 419-430.

Pallant, J. (2001). *SPSS Survival manual.* UK: Open University Press.

Payne, T. W., Kalibatseva, Z., & Jungers, M. K. (2009). Does domain experience compensate for working memory capacity in second language reading comprehension? *Learning and Individual Differences, 19,* 119-123.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19,* 221-243.

Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. In R.E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 467-488). New York: Cambridge University Press.

Purdue University Online Writing Lab (OWL) (2009). *APA formatting and style guide*. Retrieved April 29, 2009, from http://owl.english.purdue.edu/owl/resource/560/01/

Rikers, R. M. J. P., Van Gerven, P. W. M., & Schmidt, H. G. (2004). Cognitive load theory as a tool for expertise development. *Instructional Science, 32,* 173-182.

Schnotz, W. (1993). Adaptive construction of mental representations in understanding expository texts. *Contemporary Educational Psychology, 18,* 114-120.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4,* 295-312.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32,* 9-31.

Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn? *Cognition and Instruction, 12*(3), 185-233.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tabbers, H. K. (2002). *The modality of texts in multimedia instructions: refining the design guidelines.* Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen. Retrieved August 12, 2008, from Open University of the Netherlands site: http://www.ou.nl/Docs/Expertise/OTEC/Publicaties/huib%20tabbers/doctoral%20dissertation%20Huib%20Tabbers%20-%20web%20version.pdf

Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2001). The modality effect in multimedia instructions. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 1024-1029). Mahwah, NJ: Erlbaum. Retrieved August 12, 2008, from http://www.hcrc.ed.ac.uk/cogsci2001/pdf-files/1024.pdf

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3,* 257-287.

Turner, M. L., & Engle, R. W. (1989) Is working memory task dependent? *Journal of Memory and Language, 28,* 127-154.

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5-13.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics, 25*(3), 315-339.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, 49A* (1), 51-79.

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2,* 1-12.

Wong, W. (2001). Modality and attention to form and meaning in the input. *Studies in Second Language Acquisition, 23,* 345-368.

Velayo, R. S., & Quirk, C. (2000). How do presentation modality and strategy use influence memory for paired concepts? *Journal of Instructional Psychology, 27,* 126-135.

Yeung, A. S., Jin, P., & Sweller, J. (1997). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology, 23,* 1-21.

Zhang, Y., & Luximon, A. (2005). Subjective workload measures. *Ergonomia, 27*(3), 199-206.