FACTORS EFFECTING EYE TRACKING MEASURES AND ACHIEVEMENT IN MULTIMEDIA LEARNING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SERKAN ALKAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER EDUCATION AND INSTRUCTIONAL TECHNOLOGY

JANUARY 2013

Approval of the thesis

FACTORS EFFECTING EYE TRACKING MEASURES AND ACHIEVEMENT IN MULTIMEDIA LEARNING

submitted by SERKAN ALKAN in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Education and Instructional Technology Department, Middle East Technical University** by,

Prof.Dr. Canan Özgen ___________
Dean,Graduate School of**Natural and App. Sciences**

Prof.Dr. Soner Yıldırım ___________
Head of Department, **Computer Education and Instructional Technology**

Prof.Dr. Kürşat Çağıltay ___________
Supervisor, **Computer Education &Instructional Tech. Dept., METU**

Examining Committee Members:
Prof.Dr. Safure Bulut ___________
Secondary Science Mathematics Education Dept., METU

Prof.Dr. Kürşat Çağıltay ___________
ComputerEducation &Instructional Tech. Dept., METU

Prof.Dr. Zahide Yıldırım ___________
ComputerEducation &Instructional Tech. Dept., METU

Assist.Prof.Dr. Tuğba Tokel ___________
ComputerEducation &Instructional Tech. Dept., METU

Assist.Prof.Dr. Erol Özçelik ___________
Computer Engineering Dept.,AtılımUniversity

Date: January 29, 2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Serkan Alkan
Signature:

ABSTRACT

FACTORS EFFECTING EYE TRACKING MEASURES AND ACHIEVEMENT IN MULTIMEDIA LEARNING

Alkan, Serkan

Ph.D., Department of Computer Education and Instructional Technology
Supervisor: Prof.Dr. Kürşat Çağıltay

January 2013, 97 pages

In this study, factors affecting eye tracking measures and achievement in multimedia learning were explored. Familiarity, redundancy, and control are three important factors, which affect the levels of achievement in multimedia learning. In this study, three experiments were conducted in which the main effects and interactions of familiarity, redundancy, and pace investigated. In Experiment 1, a chemistry lesson with narration were studied twice as multimedia lesson. In Experiment 2, different group of participants studied two versions of chemistry lessons. Both versions had subtitles; however, in one version in Experiment 2, narration was removed from background during the experiment. In Experiment 3, different group of participants studied two versions of mechanism lessons. One of the mechanism lessons was system-paced. The other one was also self-paced; however, in this one, the learner decided to proceed to next slide as his or her own choice. After studying lessons, participants completed an achievement test, which consisted of recognition, recall, and transfer questions. The results showed that fixation count, fixation duration, total fixation duration, and total visit duration showed significant differences as well as interactions as per fragments, familiarity, areas of interest, and type of images, varying in accordance with the lesson type. The correlations among scores of achievement tests and eye tracking metrics were also reported. The results are discussed within the scope of cognitive theory of multimedia learning design principles and cognitive load theory in the conclusion chapter.

Keywords: Multimedia Learning, Cognitive Theory of Multimedia Learning, Cognitive Load Theory, Eye Movements, Eye Tracking

ÖZ

ÇOKLU ÖĞRENME ORTAMINDA BAŞARIYI VE GÖZ HAREKETİ ÖLÇÜMLERİNİ ETKİLEYEN FAKTÖRLER

Alkan, Serkan

Doktora, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü
Tez Danışmanı: Prof.Dr. Kürşat Çağıltay

Ocak 2013, 97 sayfa

Bu çalışmada, çoklu öğrenme ortamında başarıyı ve göz hareketi ölçümlerini etkileyen faktörler çalışılmıştır. Çoklu öğrenme ortamındaki fazlalıklar, ön bilgi ve kontol seçeneği öğrenme düzeyini ve göz hareketlerini etkileyebilecek faktörlerdir. Çalışmada sunuma aşinalık, sunumdaki fazlalıklar ve sunum hızının kontrol edilebilirliği için üç deney yapılmıştır. Bu deneylerde değişkenlerin göz hareketleri ve başarı üzerine etkileri incelenmiştir. Deney 1'de fonda sözel anlatım içeren kimya dersi kullanılmıştır. Deney 2'de, deney 1'de kullanılan kimya dersine altyazı eklenmiş bir sürüm ve fonda sesin olmadığı altyazılı diğer bir sürüm kullanılmıştır. Deney 3'de ise akışın bilgisayar tarafından kontrol edildiği ve akışın katılımcı tarafından kontrol edildiği iki farklı mekanizma sürümü kullanılmıştır. Her deneyin sonunda katılımcılar tanıma, anımsama ve transfer sorularından oluşan başarı testine katılmışlardır. Sonuçlar, sabitlenme süresi, sayısı, toplam sabitlenme süresi ve toplam ziyaret süreleri açısından incelenmiştir. Göz hareketlerinin ilgi alanları, resim biçimleri, ders bölümleri, göz aşinalık değişkenlerine göre anlamlı farklar ve etkileşim gösterdiği bulunmuştur. Başarı testi sonuçları ve göz hareketleri ölçütleri arasındaki korelasyon katsayıları verilmiştir. Sonuç bölümünde bulgularçoklu öğrenme bilişsel teorisi prensipleri ve bilişsel yük teorileri kapsamında tartışılmıştır.

Anahtar sözcükler: Çoklu Öğrenme, Bilişsel Yük Teorisi, Çoklu Öğrenmenin Bilişsel Teorisi, Göz Hareketleri

To My Family,

ACKNOWLEDGMENTS

First of all, I would thank to Prof.Dr. Kürşat Çağıltay, my advisor, for his unobtrusive, continuous, and constructive support. Second, I would like to extend my gratitude to the jury members, Prof. Dr. Safure Bulut, Prof.Dr. Zahide Yıldırım, Y.Doç.Dr. Tuğba Toker, and Y.Doç.Dr. Erol Özçelik for their constructive feedbacks about my thesis. Their critics and feedbacks made my draft a thesis.

I am grateful to the subjects from Middle East Technical University and Atılım University for their participation in the experiments with no expectations. The fellows and colleagues at the CEIT department and RCILO always encouraged me to work on my dissertation and showed tolerance for my lengthy working hours.

My family merits the greatest portion of indebtedness; Can, Duru, Neş'e, Father, and Brother. It is wonderful to be a part of such a family.

The laboratories used in this study established by Faculty Development Program supported by Republic of Turkey Ministry of Development, and Atılım University. The experimental equipment for data collection was provided by a TÜBİTAK grant (SOBAG 104K098).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter presents brief information about the background, purpose, and the significance of the study as well as hypotheses. The theories and relevant literature reviewed here forms the basis of the eye-tracking methodology of current study, reported in Chapter 2.

1.1. Background of the Study

Multimedia learning acquired a common consent in education practice in the last few decades. There are some theories coexist in this area, and two of them studied at great extent. These two contemporary theories of multimedia learning are Cognitive Load Theory (Sweller, 1988; Sweller & Chandler, 1994) and Cognitive Theory of Multimedia Learning (Mayer, 2001, 2009).

Why do these two theories have supremacy in this domain? A possible answer is given by Reed (2006) in his review article, although also he added several further questions. Reed reviews architectures in terms of implications on interaction between multimedia and learning. He grouped theories based on multimodality and instructional implications. The cognitive architectures which form the theoretical foundation for multimedia learning are (i) Paivio's (1969, 1975) Dual Coding Theory; (ii) Baddeley's (1999) working memory model, (iii) Engelkamp's (1998) multimodal theory; (iv) Sweller's (Sweller, Merriënboer, & Paas, 1998) cognitive load theory; (v) Mayer's (1997i 2001) multimedia theory; and (vi) Nathan's (Nathan, Kintsch, & Young, 1992; Nathan, Resnick, 1994) ANIMATE theory. The comparison of these architectures in terms of their typical inputs, coding aspects, related memory structures and the most significant contributions is given Table 1.

Table 1: Cognitive Architectures for Multimedia Learning (Reed, 2006).

Theorist	Typical Input	Coding	Memory	Contribution
Paivio	Words Pictures	Semantic associations Visual images	Long term	Dual coding theory
Baddaley	Words Spatial Materials	Phonological Visual/Spatial	Short term	Working memory model
Engelkamp	Action phrases	Motor programs Semantic concepts?	Long term	Multimodal theory
Sweller	Mathematics problems Diagrams	Schema construction Schema construction	Short term	Cognitive load theory
Mayer	Science text Animation	Verbal model Pictorial model	Short term/ long term	Multimedia design principles
Nathan	Word problems Animation	Problem model Situation model	Short term	Constructivist feedback

Reed (2006) suggests that the first three models, the multimodal theories of mind, act as the theoretical foundations for the latter three models for the instructional theories. Sweller's Cognitive Load Theory (CLT) has used the limited capacity of working memory as theoretical base, whereas, Mayer's Cognitive Theory of Multimedia Learning (CTML) has used Paivio's dual coding, Baddeley's Limited Working Memory Model/Theory (see Baddeley, 2007; Baddeley, Eysenck, & Anderson, 2009) and Sweller's cognitive load theory. Reed said that Mayer's theory is a specific theory for "multimedia learning" and has a practical orientation that made this model useful for instructional design. The model can produce easy-to-follow outputs for designers and practitioners. Therefore, it can be accounted for a reason why Mayer's and Sweller's theories are studied extensively by researchers.

Cognitive Load Theory was originally conceptualized by Sweller (1988). CLT is elucidated further in Chapter 2. In general, CLT emphasizes human cognitive system's bottleneck, namely working memory (WM). WM can be easily overloaded if it has to manage chunks of information simultaneously, which exceed its capacity. According to CLT, the two main factors influencing the load are architecture of human cognitive system and prior domain knowledge. To conceptualize the types of load in multimedia environment essential, extraneous and germane load types are defined. Intrinsic cognitive load produced by the interactivity of the elements that should be processed simultaneously. Prior knowledge has an effect on the perceived level of intrinsic cognitive load. The more knowledge learners have about a given domain, the less amount of load they become exposed to. Extraneous cognitive load refers to information, which is not directly related to information to be processed such as design factors. Germane load, a subsequent addition to the theory (Sweller et.al, 1998), covers the extra cognitive activities associated with multimedia materials such as self-explanation or imaging requirements designed purposefully to foster learning (Kalyuga, 2009). The interaction of these load types and instructional implications are discussed in Chapter 2 as mentioned earlier.

The current state of research studies about CLT can be grouped into three topics, which are (i) learning in complex environments, (ii) learner control and choice, and (iii) animated and multimedia instruction (Kirschner, Ayres, & Chandler, 2011). Currently, multimedia research topics in CLT focus on dynamic and static learning environments, task characteristics and learner engagement (Kirscher, Kester, & Corbalan, 2011). The aim of these groups is to optimize the cognitive load to achieve maximum learning.

Second dominant theory in multimedia learning is Mayer's (2009) Cognitive Theory of Multimedia Learning (CTML). It is not surprising that, as the CTML anticipates, the research emerged in multimedia learning based on CTML generally have practical considerations in design. Mayer's (2009) "use-inspired basic research" approach regarding multimedia learning takes more central position for the aims of the study. As an introduction to the concepts of CTML, Mayer's (2009) recent book on multimedia learning can be taken into account as the most significant resource. At first, recent publication date of the book gives state-of-art information about CTML. Secondly, Mayer's "use-inspired basic research" orientation and theoretical stability, located CTML at the center of the multimedia studies as a one of the major theories among instructional designers. And, the last merit of the book is that its structure. After listing principles with favoring research data, Mayer explains limitations with boundary conditions. Furthermore, Mayer is generous in suggesting further research, describing what should be done and what is needed next at the end of every section of his book.

The main structure of this study was formed in 2008 while the basis of the book dates back to 2001. Since the second edition of the book in 2009 before the data collection process, I had the chance to revise and update my original hypotheses according to set out in this latter edition. Mayer (2009) suggests that the main difference from 2001 to the 2009 is the change in the number of multimedia learning principles. He argues that the principles have matured but not saturated because of many boundary conditions related to principles. In the last decade, the number of principles increased from seven to twelve which are grouped into three main groups: (i) principles for reducing extraneous processing, (ii) principles for managing essential processing, and (iii) principles for fostering general processing. The principles and their intentions are listed in Table 2. The reasons behind this modification in the number of principles are discussed in the literature review section.

Table 2: Principles of Cognitive Theory of Multimedia Learning (Mayer, 2009).

Intention of Principle	Multimedia Design Principle
• principles for reduce extraneous process	• coherence • signaling • redundancy • spatial contiguity • temporal contiguity
• principles for managing essential processing	• segmenting • pre-training • modality
• principles for fostering generative processing	• multimedia • personalization • voice • imaging

The principles of CTML are mapped with Cognitive Load Theory's load assumptions; extraneous, intrinsic, and germane, respectively (Kalyuga, 2011). Kalyuga (2011) highlights that CTML's elegance is a result of its logical consistency. Although CLT and CTML share common assumptions such as limited capacity, dual coding, and the role of Long Term Memory (LTM), they differ in terms of their focuses. While, CTML focuses on the *different processing routes* during learning, CLT focuses on the *load* c*reated by the processes* in working memory. Since three types of cognitive loads are defined by CLT is encompassed by CTML's principles and intentions, using only CTML's clear theoretical structure seems to cover both theories, at least, to some extent. For this reason, CTML's clear theoretical structure considered as more appropriate for the aims of the current study. On the other hand, the outcomes of the experiments conducted in the study are discussed separately according to CLT and CTML, because these theories differ on their position regarding the role of LTM. Somehow, CTML seems to underestimate the value of LTM during processing; however, CLT places it in a more central position in terms of load. This difference gains importance when participants studied material more than ones, which violate novelty of the content. The next section briefly explains the rationales behind selecting CTML and the methods employed.

1.2. Background of the Problems and the Purpose of the Study

The CTML has three components (See Figure 1); Model accepts modal approach as a basis for its theoretical position for memory and learning. Mayer (2009) collates Atkinson and Shiffrin's (1968) multi-store memory model and Baddaley's working memory model to depict CTML. In his model, Mayer prefers Baddeley's (2007) revised Working Memory Model in lieu of Atkinson and Shiffrin's (1968) Short Term Memory (STM) structure. Originally, Atkinson and Shiffrin's (1968) model has three structures; Sensory Memory (SM), Short Term Memory (STM), and Long Term Memory (LTM). In this model, the first stage is SM, which lasts for only a few seconds and acts as a door to receive input from environment, transmitting the raw, unprocessed sensory information. The SM receives information from environment via its registers. The registers can be in visual, auditory, olfactory, taste or haptic. Despite this, the CTML takes into account only the most widely studied registers, namely auditory and visual. The study of other sensory registers, such as haptic registers, calls for the attention of learning and cognitive researchers. Short Term Memory (STM) is the conscious part of the memory, which is temporary and requires rehearsal to remember its content. In CTML, Mayer replaces STM with the more sophisticated WM to reach a more complete model. The last stage in the modal memory approach is the LTM where is the place in which all memories stored, but, the core issues such as capacity of LTM or type of storage in the LTM have not been decided yet. Additionally, CTML is inattentive to the functions or structures of LTM and requires further research.

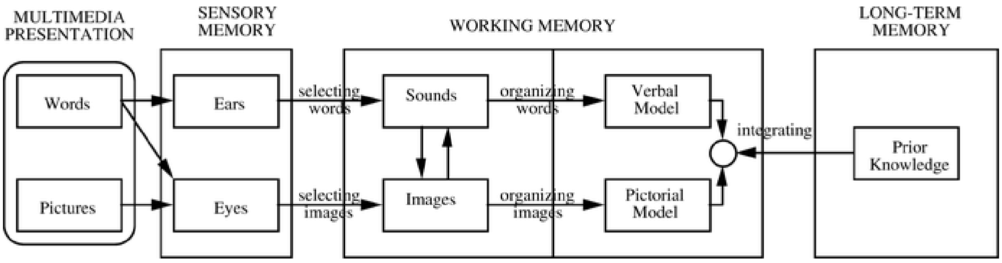


Figure 1: Cognitive Theory of Multimedia Learning (CTML) by Mayer (2001, 2009).

Mayer's model suggests different routes for multimedia processing for verbal and pictorial models in SM and WM. Both structures of SM and WM models are congruent with previous models. The interconnectedness of LTM with WM is ignored in CTML, and conceptualized as "prior knowledge" and connected with a one-way arrow to the WM. However, narrowing down all memory types to the "prior knowledge" and ignoring all other memory types such as semantic, procedural, episodic etc. can be considered as a drawback for the current status of CTML. Therefore, Mayer's CTML model depicts multimedia learning's interaction with auditory and visual channels of working memory falling short on the other aspects of the memory process such that in this model the interaction of multimedia learning with prior knowledge is not detailed in the structure.

The aim of this study is to investigate the theoretical structure of multimedia learning by focusing on CTML via new eye tracking technology. The novelty of the is technique studying exactly same content more than once and collecting data about learning process via eye-tracking tools which is supported by paper based recognition-recall-and-transfer test conducted just after completion of studying session. These multi-faceted data from experiments will be used for reconsidering principles and their relations to the model. It is assumed that studying the same content more than once and online monitoring of eye movements through learning process will yield comparable results with previous literature of multimedia learning. At the end of the study, a more comprehensive version of multimedia learning model is expected, in which prior knowledge, pace and domains other than science are experimented. Prior knowledge, control, and domain invariability are the examples for boundary conditions of CTML, which are described below.

The skeleton of the CTML is constituted by computing effects sizes – differences of means divided by SD – individual articles for all principles. Results are generally quite satisfactory for novices and but all principles have various boundary conditions or limitations. Consequently, Mayer (2009) suggests further research to resolve these uncertainties –or increase saturation of principles- by including subjects other than novices and domains other than science. Fortunately, the current status of the theory strengthens my hypotheses in 2008 by emphasizing the importance of (i) prior knowledge, (ii) domain variability, and (iii) focusing processes rather than output, which is, focusing only on the transfer of knowledge. In this study, the research questions are designed to test three boundary conditions of multimedia learning by measuring eye movements during study a multimedia content and supported by an achievement test.

1.3. The Research Questions

The research questions in the light of literature about CTML are given below. Instead of listing questions in an itemized fashion, the associated discussions and question are given together. The questions are given under three subtopics.

1.3.1. What is the Interrelation between Multimedia Learning and Prior Knowledge?

Prior knowledge is one of the topics emphasized by Mayer to attract researchers' attention frequently. The quotations below are barrowed from Mayer's (2009) Multimedia Learning book.

In these, Mayer mentions the need for further research to reveal the function(s) of prior knowledge on multimedia learning process or output.

> "Research is needed to determine whether the coherence principle-applies-mainly to **low-knowledge learners** rather than **high knowledge learners**." (p.105)
>
> "Future research is needed to pinpoint the boundary conditions of the spatial contiguity principle, particularly the role of the learner's **prior knowledge**. In particular, it would be useful to know how the learner's prior knowledge mitigates poor instructional design. *Unobtrusive* [emphasis added] techniques for measuring prior knowledge would also be helpful." (p.150)
>
> "Thus, there is preliminary evidence that the pre-training principle is most likely to apply when learners **lack domain knowledge**, although further research is warranted." (p.199)
>
> "Overall, more work is needed on how best to create effective **pre-training experiences**." (p.199)
>
> "Research is needed on how to embed knowledge assessments within a lesson so that the appropriate level of **pre-training** can be provided for each individual learner." (p.199)
>
> "Additional research is needed to determine the boundary conditions of the personalization principle, particularly whether it applies better early in training rather than later and for students who are **less experienced working with online tutors rather than more experienced**." (p.253)

As seen from the above quotations, the operational definition of prior knowledge could not be extracted from the way that Mayer uses the relevant terms. The definition of knowledge that person has before testing defined by using high-low, previous, or pre-training terms. Since, Mayer prefers stay unclear about the sub-components or types of LTM, the definitions of these terms cannot be isolated from the studies conducted by Mayer so far. Although the meaning and function of *prior knowledge* is vogue in the area of multimedia learning, it is clear that it should be explored by further studies. On the other hand, the conceptualization of prior knowledge in CTML can be drawn from the methods of measurement of learning. Mayer uses retention tests to measure prior knowledge. Retention tests consist of two main parts. One is a set of recall questions, and the other consists of transfer questions (Mayer, 2009). The retention tests used in CTML covers the conscious processes such as recall or recollection. But there are also unconscious processes of memory which have effect on task performance such as recognition or familiarity. Both types of memory cover the automatic use of memory unlike intentional use of such as recollection (Jacoby, 1991). The difference between intentional or automatic types of memory indicates a dissociation that might be observed by processes in addition to task performances also. Scott and Dienes (2008) argue that familiarity, which supports dual processing model, is an essential source for creating knowledge for specific types of learning. Additionally, Khosrowjerdi and Iranshahi (2011) conceptualize the differences between expertise and prior knowledge, which have so far been studied extensively by multimedia researchers as depicted in the following part.

Familiarity is the degree of awareness, which does not result from actual experience; whereas expertise covers the power of analysis to solve problems, which is related to success directly. In familiarity, the time spent on the task is not related to processes or content, but in expertise, the processes are connected to success. The past experience, on the other hand, refers to the ability acquired in previous uses of similar contents which is akin to prior knowledge (Khosrowjerdi & Iranshahi, 2011).

One way to conduct experiments, which used prior knowledge as a dependent variable, is embedding it as a factor of repeated measures in the design. The categories of prior knowledge about the domain can be constructed by re-studying the same content of the multimedia learning process. The difference between first and second re-study sessions could be counted as the familiarity factor on multimedia learning. The details about design and variables are given in the methodology section.

The research questions about the interaction between multimedia learning and familiarity are restated below.

1. It is known that, multimedia heuristics was varied according to learner's prior knowledge. The advantages of the multimedia content are lost when the learners' knowledge about the domain increases. In this study, it is expected to be observed that, the changes in subjects' studying strategies when learners have familiarity about the material, which was created by studying same content repeatedly? It is assumed that the online and unobtrusive measures of eye tracking could reveal data about "how the responses of the participants will change in the case of multiple readings of the same content".

   To test the question above, it is hypothesized that the familiarity created by multiple studies of the same multimedia content will not produce significant differences in terms of related eye tracking metrics. .

2. Does the eye tracking data yield comparable results with the traditional measurement methods of multimedia learning such as recognition, recall, and transfer tests in the presence of familiarity about content?

   Based on the this research question it is hypothesized that there is no relationship between paper-pencil based achievement tests applied after studying multimedia content and eye tracking measures recorded during multiple studies of same content.

1.3.2. What is the Relationship between Multimedia Learning and Pace of Presentation?

The speed of learning is highly dependent on the personal characteristics of learners. The interaction between presented material and the learners is emerged the best when the presented material complex for the learner. This perceived complexity is best predicted by the learners' prior knowledge and the pace of the presented materials. If the learner has prior knowledge or is competent about the material, the advantages of the multimedia presentation over other methods are lost. Another method to overcome complexity is the presentation pace of the material. If the material is relatively complex and the learner has no control over the presentation, the multimedia advantage is maximized. There are many studies about self-paced versus system-paced designs to investigate multimedia principles. Mayer (2009) suggests that speed is one of the key factors that controls cognitive load during experiments and affects the transfer performances of the learners. Mayer suggests further research to explore effect of the pace on the principles of multimedia learning. Some of these suggestions are given below (Mayer, 2009).

> "The negative effects of redundancy may be eliminated when the presentation is **slow-paced or under learner control**. These are research questions that warrant further study." (p.134)
>
> "Research on temporal contiguity yields two important boundary conditions: … the lesson is **learner-paced rather than system-paced** … Further research is needed to determine what constitutes an ideal segment size." (p.168)
>
> "Further research is needed in which the same simultaneous and successive lessons are presented under **learner control and under system control**. … Based on the cognitive theory of multimedia learning, segmenting is most likely to have its strongest impact when the material is complex, the presentation is **fast-paced**, and the learners are inexperienced with material. More research is needed to pinpoint the boundary conditions of the segmenting principle."(p.185)
>
> "Research is needed to determine the relative effectiveness of continue buttons based on instructor-determined segments (as used in these experiments) versus pause/ continue buttons and slider bars without instructor-determined segments. … Similarly, research is needed to determine how the complexity of the material and the **pace of presentation** affect the segmenting principle."(p.188)
>
> "Research on the modality principle suggests boundary conditions involving the complexity of the material, **the pacing of presentation**, and the learner's familiarity with the words. Further research is needed to pinpoint the boundary conditions of the modality principle, and to determine the implications for a cognitive theory of multimedia learning." (p.219)

In the light of the statements above, about the interrelation between multimedia learning and pace of presentation, the third question can be formulized as follows.

3. Since the pace of the presentation has an effect on multimedia principles, how do responses of learners change when they re-study the same content in a pace-controlled environment? It is assumed that re-study of the same content and pace of the study will interact or will be affected by learner's perceived complexity of the material, increased familiarity with revisits and the cognitive load resulting from the ability to control speed. Eye tracking will provide online data about "what might be going on in the learners mind" during self-study.

   Therefore it was hypothesized that having control over multimedia lesson should not produce significant differences eye tracking metrics in multiple study of the same content".

1.3.3. What is the Interrelation between Multimedia Learning and Domain of Content?

Another research question investigated (Sweller & Chandler, 1994) in this study is the domain of materials used in multimedia learning. In multimedia studies, the materials are generally developed by the researchers for the experiments' sake by themselves. And these materials are chosen generally from science domains since these are more appropriate for designing multimedia environments. Consequently, considerable data on science domain accumulates as time passes that shows multimedia principles work quite well. But there are some objections to the frequent use of science materials among multimedia learning studies. Multimedia learning cannot be constrained only to science domain because science is just one field of education in general. It is suggested that other neglected domains should be studied to show their congruity with the known multimedia principles. Mayer (2009) states that "*Again, additional research is needed to pinpoint the role of the nature of the material in multimedia learning.*" Ainsworth (2008) also proposes that one of the drawbacks of these studies is the sole focus on science materials, which narrow its ecological validity.

4. Does an example of effective and well-designed multimedia material proposed by Mayer work for different science domains such as chemistry or mechanism? It is known that multimedia advantage appears most "*when the material is complex and presented at a rapid pace for the learner"* (p.275) in science domains. In this study it will be investigated that what happens when a parallel design is used with other topics.

   As a last hypothesis in the current study, it is predicted that new domains will produce differences in terms of working principles in multimedia learning.

The questions about prior knowledge, pace, and domain variability are related and they enclose most of the principles defined in Mayer's CTML. The boundaries and limitations of the principles will be redefined with the help of an eye tracking technology, which monitors real-time processing during studying and the re-study methodology which controls the prior knowledge. The ways of operationalizing these terms are given in this section and their integration to the experimental design are explained in the methods section.

To sum up, in this study a repeated measure of familiarity should yield information about the principles of coherence, spatial contiguity, pre-training, and personalization. Pace is important for redundancy, temporal contiguity, segmenting, and modality principles in multimedia learning. Testing multimedia learning in different contexts creates an opportunity to clarify show learners behave in the condition of concise narrative animation during repeated study of the same content.

1.4. Significance of the Study

Empirical multimedia studies generally gather the attention of researchers' from a wide range of disciplines such as psychology, cognitive/computer science, educational/instructional science and human computer interaction. Psychologists and cognitive scientists focused on converging and diverging processes of pictures and texts at the theoretical or computational level. The Human Computer Interaction groups' concern can be outlined as how multimedia elements should be designed to maximize usability. On the other hand, educationalists need practical and applicable results. Thus, they have to utilize the results of both theoretical and applied multimedia studies to optimize learning. Additionally, the multimedia learning in an educational content cannot be limited only to the visual inputs, i.e., pictures, diagrams, animations or text.

Appropriate use of audio channels can also enhance learning if they are used with visual input effectively (Mayer, 2001).

Formulating processes of mind at the presence of pictures or words and adapting them into the limitations of cognition is a more crucial question for instructional designers (Schnotz & Kürschner, 2007). Finding out the best method to present information, and answering whether designer should use pictures, diagrams are the core considerations for designers. At the same time, should learners read or listen to the presented content for achieving best learning (Grimley, 2005)?

Principles or design implications imposed by multimedia theories can be used by designers without help of any theoretical base. But lack of theoretical base –even in use inspired basic research (Mayer, 2009) - might limit the quality of the output. Therefore, it can be said that both theoretical and practical designs in the area contribute equally to an effective learning process.

The present study aims to make both theoretical and practical contributions to the multimedia literature by introducing a new methodology. It is assumed that, data based on an unobtrusive online eye-tracking methodology with repeated measure, will provide usable data for both theory and practice. In the study, boundary conditions of the multimedia will be investigated by not only measuring with transfer questions but also by recognition and recall questions. Adding contents other than classical domains and restudying the same content will bring multimedia research closer to real-life situations. Consequently, as researches approximate to real-life situations, their external and ecological validities will increase.

1.5. Terms and Abbreviations

In this section frequently used terms in this study are alphabetically listed. Their definitions and limitations of use if exists are also given.

- **Areas of Interest (AOI):**

Areas of Interest (AOI) defines the area on which fixation occurs on screen. User defined AOI's can take any shape, and give researchers the opportunity to study with names instead of coordinates.

- **Cognitive Load Theory (CLT):**

Cognitive Load Theory predicts that the working memory load is proportionally directly to the learning outcome. The three types of load should be balanced to maximize the learning. These are intrinsic, extraneous, and germane load.

- **Cognitive Theory of Multimedia Learning (CTML):**

Cognitive Theory of Multimedia Learning proposes that dual processing of verbal and visual cues can foster learning. The use of a dual channel, with visual and verbal cues embedded, is better than using a visual channel only. CTML's boundary conditions and relation to memory structures are still a matter of debate.

- **Concise Narrated Animation (CNA):**

Concise Narrated Animation is the best practice of CTML. Here the verbal and visual processes are used together to transmit information. CNA indicates that using narration and images is more profitable than using on-screen text and images for learners.

- **Fixation Count:**

Fixation count is the number of individual fixations on a predefined area such as AOI or an AOI group.

- **Fixation Duration:**

Fixation duration is the time devoted to a specific location measured by reflections from fovea. The durations are generally measured in milliseconds.

- **Long Term Memory (LTM):**

Long Term Memory refers to the storage in which information stored unconsciously or pre-consciously. The information from LTM can be retrieved from LTM to WM or STM if needed.

- **Multimedia Learning:**

The use of multimedia learning can be twofold. One is the short version of CTML and the other is the learning from both images and word. Words can be in either written or verbal form.

- **Recall:**

Recall refers to the retrieval of information from LTM. The term recall used in several studies as an equal term for memory historically. There are several types of recall such as cued recall or free form.

- **Recognition:**

Recognition is the remembering of a stimulus or event, whether it was seen before or not. Recognition includes a response and a cue, and accepted to be easier than a recall task.

- **Sensory Memory (SM):**

Sensory Memory is the temporary storage between external world and Short Term Memory. The physical stimulus from outer world is transferred to STM with sensory receptors within the SM system. The SM has a very short duration.

- **Short Term Memory (STM):**

Short Term Memory is the active, conscious part of the memory. The information from LTM and inputs from SM are integrated in STM. The average duration of STM is about 30 seconds unless it is rehearsed continuously.

- **Transfer:**

Transfer is the activity in that a learnt material can be used to solve a new problem in a new situation. Transfer is a sign of better learning when compared to remembering of something measured with retention test.

- **Visit Count:**

Visit Count is the number of visits to a specified AOI. Consequent fixations in that given area does not count, but if the fixation takes place out of AOI and then on the AOI, this is counted as a visit.

- **Visit Duration:**

Visit Duration is the time interval between the first and last fixation on a specified AOI. Visit duration can include many fixations if they are consequent. Fixations out of the AOI are not counted.

- **Working Memory (WM):**

Working memory is a transitory process where all information is kept actively to be integrated. The information can be transferred to WM from LTM and SM. However, WM keeps information active, and it is not merely a store, but it is a place where information is handled for further processing if needed. The older concept of STM is used in lieu of WM; however, WM is, in fact, a more comprehensive structure than STM, which apparently is interpreted to be a relatively static structure.

CHAPTER 2

LITERATURE REVIEW

It is known that the integrated use of both visual (pictures and texts) and audio input somehow helps comprehension and enhances learning. The idea of implementing multimedia into instruction as pictures, graphs, or images with text or audio is not new; the coherent usage of text and pictures had been recommended even in medieval era by Comenius (Schnotz, 2005) whose name was given to the lifelong learning program of the European Union. Although we have been accustomed with diagrams, graphs or pictures nowadays, the visual representation of data did not become common until late nineteenth century in printed media (Kennedy, 1983). In addition to printed media, Crow (2006) argues that the spread of visual information was exploded by television broadcasts. He adds that the effect of television on culture was followed by a paradigmatic shift from left to right hemisphere assumed having the characteristics of image based, non-verbal, and holistic orientation unlike left hemisphere's linguistic, verbal and analytical characteristics. Despite this, the rich production of the visual-media communication rules by the non-educationalists, the way of using visual information in an educational context has not justified yet. Crow's point is a little bit speculative for education professionals, but it should not be ignored, because it addresses possible qualitative changes in addition to quantitative ones by using pictures instead of words in the future. The accumulation of knowledge about multimedia learning might change the methods of education in the future, but the road map of the change has not been drawn up yet.

Two current major theories about multimedia learning have been reviewed in the following sections: Mayer's (2009) Cognitive Theory of Multimedia Learning and Sweller's (2010) Cognitive Load Theory.

2.1. Cognitive Theory of Multimedia Learning

Mayer (2001) defines Multimedia Learning as learning from words and pictures. In the same manner, Multimedia Instructional Message or Multimedia Instructional Presentation is the use of word and pictures to enhance learning. Mayer (2001) emphasizes these definitions because the verbal-pictorial research in cognitive science does not always overlap with his dual-code or dual-learning considerations.

Cognitive Theory of Multimedia Learning (CTML) is the first theory specific to multimedia learning. Mayer's theory is based on information processing model. CTML has three assumptions about processes of human cognition. The first is dual channels assumption in which humans process visual and auditory information separately (Clark & Paivio, 1991). The limited capacity of processing information in WM (Baddeley, 1999) is another model embedded in the set of assumptions by CTML. Lastly, information organization and integration with previous mental representations is taken for granted by CTML (Mayer, 2005). CTML takes three memory structures into account; sensory memory, working memory and long-term memory. In CTML, sensory memory's function is just to receive information via ears or eyes. The central focus of CTML is WM. As stated in preceding sections, the main function of WM is to process and manipulate information actively from both SM and LTM. The LTM in CTML is depicted as storage for knowledge and additionally functions to recall materials when required. The up-to-date model of multimedia learning is given in Figure 2: Each figure depicts different routes for processing of pictures (top), spoken words (middle), and printed words (bottom).

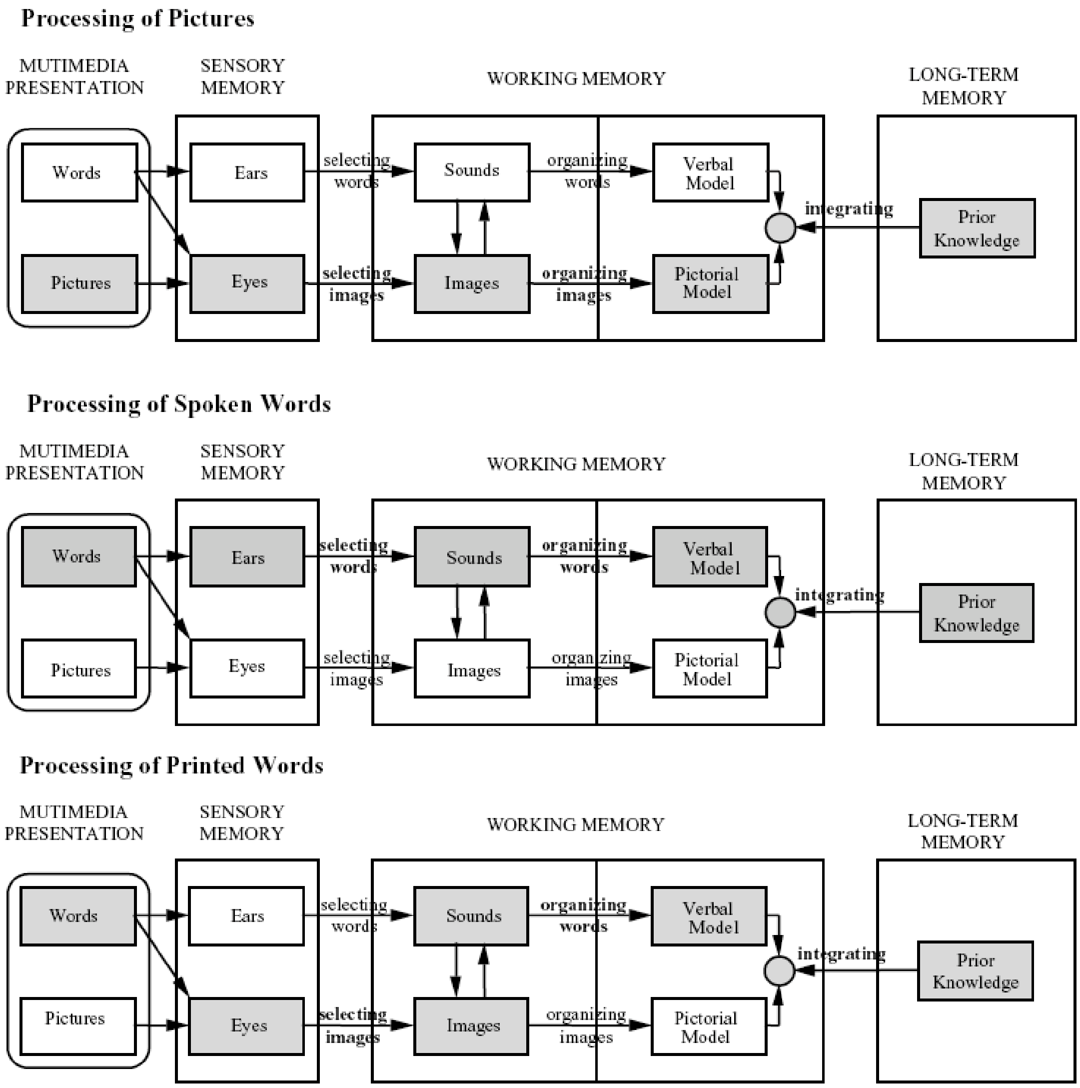


Figure 2: The up-to-date version of Cognitive Theory of Multimedia Learning. Processing of pictures (top), spoken words (middle), and printed words (bottom) (Mayer, 2009).

The Multimedia studies by Mayer (1997, 2001, 2005, and 2009) and his colleagues showed that the components of the working memory play a significant role in the differentiation of verbal and pictorial stimuli. Mayer et al conducted a series of studies in which science lessons were presented in the form of text, graphics and combination of both. Results showed that pictorial representation help learners for the recall and recognition of material presented.

Learning depends on several successive cognitive processes in CTML. Selecting whether words or images, organizing selected words or images and integrating them in a meaningful whole is essential for successful learning. During these processes, information is converted into different representations. Initial step is the presentation of content as words or pictures to the learner. Afterwards, learner picks information via ears or eyes into sensory memory. The information in sensory memory stays there for a very short time. Selected sounds and images are processed in WM. Consecutively, verbal or pictorial models are re-constructed together with previous information recalled from LTM.

The principles of CTML propose practical guidelines for designers about how graphics and text can be used together during presentation of a specific topic. Mayer's classical scenarios such as

"How a strike occurs" or "How brakes work" were studied in detail. The rate of learning and transfer of the knowledge to the novel situation is measured by means of a summative evaluation. The transfer of the knowledge is the major sign of comprehension that is an alteration in long term-memory according to the Mayer in 2001. At the beginning of the millennium, placing transfer at the heart of educational system (Bransford & Swartz, 1999) had been acknowledged most; however, it subsequently changed a decade later by suggesting a focus on not only output but also on process (Mayer, 2009).

Mayer (2001) formulated seven research-based principles for the practical applications of CTML. For each principle, he has two different empirical evidence sources from retention and transfer tests. These principles are summarized in Table 3.

Table 3: Seven Research-based Principles for the Design of Multimedia Messages (Mayer, 2001).

1. Multimedia Principle: Students learn better from words and pictures that from word alone.
2. Spatial Contiguity Principle. Students learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.
3. Temporal Contiguity Principle: Students learn better when corresponding words and pictures are presented simultaneously rather that successively.
4. Coherence Principle: Students learn better when extraneous words, pictures, and sounds are excluded rather than included.
5. Modality Principle: Students learn better from animation and narration than animation and on-screen text.
6. Redundancy Principle: Students learn better from animation and narration than from animation, narration and, on-screen text.
7. Individual Differences Principle: Design effects are stronger for low-knowledge learners than for high-knowledge learners and for high-spatial learners than for low spatial learners.

Mayer and Moreno (2003) propose ways to handle cognitive load in multimedia learning. Their proposals are based on five overload scenarios and possible solutions to them. The list of overload scenarios and possible solutions with research effect is given Table 4. However, neither scenarios nor solutions are inclusive of all possibilities for multimedia learning. This attempt provides a guideline to study cognitive load in multimedia learning. They suggest future studies and additional direct measures for cognitive load be developed. In the same line of research, Moreno (2006a) proposes her own modified multimedia learning theory, called Cognitive Affective Theory of Learning with Media (CATLM). Recently, Moreno and Mayer (2010) published a chapter incorporating CTML and CATLM in a way that affect, motivation and metacognition are all taken into consideration. They compare the assumptions of CLT and CTML with other studies conducted outside their research group. They conclude that any inference from CTML studies about CLT would be a consequence of the above studies, CLT and CTML are currently the two main theories in the area of multimedia learning; however, they have still yet to provide us with a complete picture. Future studies in this area require new methods and questions to refine their assumptions and to provide outcomes that are more practical for designers.

Table 4: Load-Reduction Methods for Five Overload Scenarios in Multimedia Instruction (Mayer & Moreno, 2003).

Type of Overload Scenario	*Load-Reducing Method*	*Description of Research Effect*
Type 1: Essential processing in visual channel > cognitive capacity of visual channel		
Visual channel is overloaded by essential processing demands.	Off-loading: Move some essential processing from visual channel to auditory channel.	Modality effect: Better transfer when words are presented as narration rather than as on-screen text.
Type 2: Essential processing (in both channels) > cognitive capacity		
Both channels are overloaded by essential processing demands.	Segmenting: Allow time between successive bite-size segments.	Segmentation effect: Better transfer when lesson is presented in learner-controlled segments rather than as continuous unit.
	Pretraining: Provide pretraining in names and characteristics of components.	Pretraining effect: Better transfer when students know names and behaviors of system components.
Type 3: Essential processing + incidental processing (caused by extraneous material) > cognitive capacity		
One or both channels overloaded by essential and incidental processing (attributable to extraneous material).	Weeding: Eliminate interesting but extraneous material to reduce processing of extraneous material.	Coherence effect: Better transfer when extraneous material is excluded.
	Signaling: Provide cues for how to process the material to reduce processing of extraneous material.	Signaling effect: Better transfer when signals are included.
Type 4: Essential processing + incidental processing (caused by confusing presentation) > cognitive capacity		
One or both channels overloaded by essential and incidental processing (attributable to confusing presentation of essential material).	Aligning: Place printed words near corresponding parts of graphics to reduce need for visual scanning.	Spatial contiguity effect: Better transfer when printed words are placed near corresponding parts of graphics.
	Eliminating redundancy: Avoid presenting identical streams of printed and spoken words.	Redundancy effect: Better transfer when words are presented as narration rather narration and on-screen text.
Type 5: Essential processing + representational holding > cognitive capacity		
One or both channels overloaded by essential processing and representational holding.	Synchronizing: Present narration and corresponding animation simultaneously to minimize need to hold representations in memory.	Temporal contiguity effect: Better transfer when corresponding animation and narration are presented simultaneously rather than successively.
	Individualizing: Make sure learners possess skill at holding mental representations.	Spatial ability effect: High spatial learners benefit more from well-designed instruction than do low spatial learners.

2.2. Cognitive Load Theory

Initially, CLT was developed as a result of the studies of learning to solve problems in 70's. In the 80's Sweller introduced the "Cognitive Load" concept to explain the limited capacity of working memory and information processes in learning (Schnotz & Kürschner, 2007). Sweller (2005) argues that the laws of evolution in biology are also applicable to the development of cognitive processes. He claims that the lack of central executive process in working memory requires an organization of information processing as is the case in the evolution of species. If the receiver has no *a priori* experience about incoming information, his/her responses become completely random manner. To overcome this randomness there are two options. The first option is the information stored in LTM with which humans can handle huge information if they have appropriate schemas or experience. The second option is that organizing data before presenting them to the learner, i.e., the art of instructional design. Previous experience and knowledge stored in memory reduces the number of variables processed by working memory. Although the original number was seven chunks for variables (Miller, 1956) that working memory can process at one time, later this number was shown to be smaller, that is, 2 to 4-way-interactions. Halford, Baker, McCredden and Bain (2005) found that in 5-way-interactions human processing capacity performs at chance levels that humans can process maximum for interaction simultaneously. To explain this complexity, Sweller (2010) gives an example of buttons on the screen. For example, if there are four buttons to learn, and the task was to combine them at any given time, this combination can be thought as a transfer to the LTM or any existing schemata. The possible permutation is 4!=24, a hard figure to decide with, but still manageable. For instance, consider there are 10 buttons on the screen. In this case, the number of possible elements to work with is 3,628,800. Apparently, any kind of cognitive architecture that compares the effectiveness of 3,628,000 possible elements would not be efficient.

Fortunately, human mind has solutions to overcome this limitation with presumably unlimited long term memory storage and the use of clever instructions. Sweller (2010) argues that all communication means, e.g. writing, speaking, etc., are designed according to the limitations of our working memory.

CLT assumes that limited the capacity of the working memory is the bottleneck of the human processing system (Kirschner, 2002) and it defines earning as the construction of the schemata by using the limits of working memory. Instructional designers should be aware of cognitive load limits of working memory to achieve learning. The factors that affect cognitive load are depicted in Figure 3. These are causal and assessment factors. Causal factors are external ones such as cognitive characteristics of the subject, difficulty of the task or the environmental conditions. Assessment factors are mental load and effort with performance. According to the CLT, the interaction of these factors determines the resultant learning.

The cognitive load on WM can be affected from the inherent nature of the material. That is intrinsic cognitive load. The presentation styles of the material increases or decreases the extraneous cognitive load. The germane load describes the effort allocated to construct schemata in the long term memory. CLT asserts that instructional interventions should deal with extraneous and germane cognitive load. It is assumed that an appropriate instructional design should decrease extraneous load and increase germane load (Kirschner, 2002).

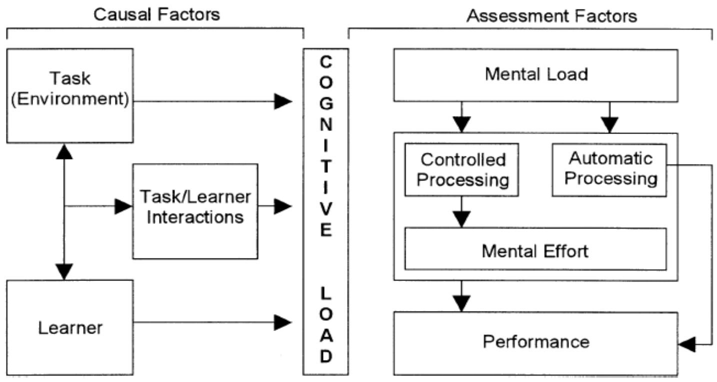


Figure 3: Factors Determining the Level of Cognitive Load (Kirschner, 2002).

Schnotz and Kürschner (2007) showed that some of the empirical data is incompatible with the assumptions of CLT. Although the load of working memory reduces learning for novices, in some situations reducing the load of working memory does not help learning; on the contrary, it leads to poorer the learning performance. They revised the theory by adding Vygotsky's Zone of Proximal Development (ZPD) to the picture. They argue that reducing cognitive load does not always produce desirable results. If the task difficulty exists in the area of ZPD, lesser task difficulties can generate negative effects. So the cognitive load is a function of expertise and task difficulty.

Attempts were made to render CLT a computational model. For example, Sawicka (2008) proposed a formal dynamic model to control cognitive load in instructional design. She claims that forming a computational dynamic model can create aid and simulation opportunities for designers. But she also suggests that her model should be extended other contents. Another computational architecture is Sorden's (2005) model. He reviews the basic concepts of Working Memory, Cognitive Load Theory and Cognitive Theory of Multimedia Learning. Afterwards he proposes a computational-cognitive architecture based on Adaptive Character of Thought – Rational (ACT-R) production system (Anderson, 2007) to act as a starting point of instructional design. As Sorden, Mayer's Cognitive Theory of Multimedia Learning has also assumptions similar to the Cognitive Load Theory as described in previous sections.

CLT has not a complete diagram as CTML. Its development has evolutionary characters. The reason for that can be as follows. The CLT does not have any assumptions in relation to learning theory; instead, CLT emphasizes on exploring the relationship between human cognitive architecture and instructional design. Moreno and Park (2010) suggest three stages to depict the development of the CLT (Figure 4). In the first stage, there are schema acquisition and automation activities with extraneous load. In the second stage, intrinsic load added to the model. Last stage contains all three load types; intrinsic, extraneous, and germane. The last model assumes that all three types of load can be affected by instructional design unlike previous one's assumptions.

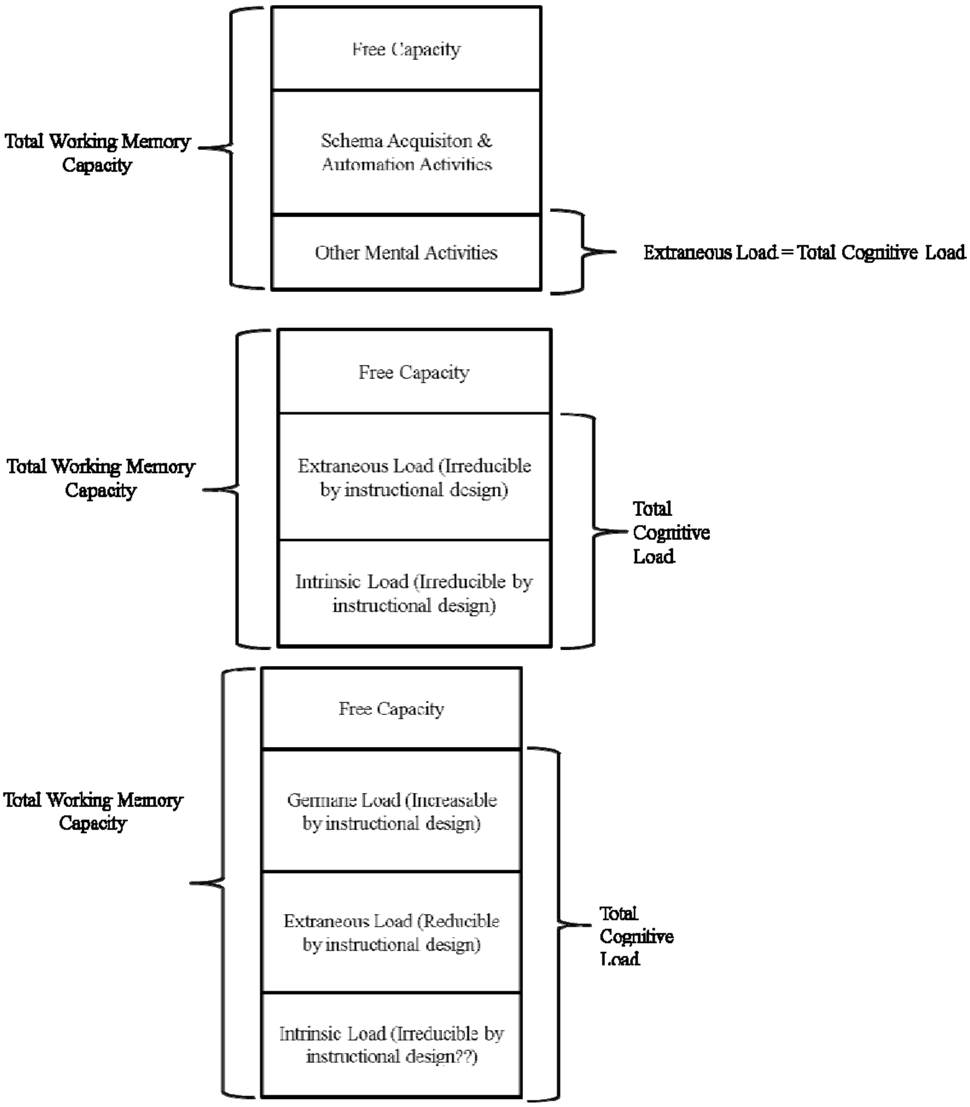


Figure 4: The Evaluation of Cognitive Load Theory (Moreno & Park, 2010).

Although CLT has a central role with CTML in this study, its contribution to the overall design is not as overt as that of CTML's. The CTL is used in the interpretation of the results in the last chapter.

2.3. Instrumenting Eye Tracking as a Tool in Multimedia Learning

The rationale behind the use of eye tracking method was chosen to study eye tracking will be explained in this section. Eye tracking methodology is a more direct measure for assessing cognitive load. Traditionally, eye tracking has been used as a method to study reading for a long time (See Rayner, 1998, for a review). More recently, Rayner, Chace, Slattery, and Asby (2006) showed that eye tracking methodology could be used for online assessment of comprehension processes in reading. They also argued that with the development of eye tracking methodology, it can be used to determine reading problems in school settings soon. Additionally, Schnotz and Kürschner (2007) mention eye responses -especially pupil dilation- as one of the possible methods of cognitive load measurement in addition to self report and dual task approach methodology (Brünken, Plass & Leutner, 2003).

Schnitzer and Kowler (2006) conducted a study about how eye movements change during multiple readings of the same text. They claim that recording eye movements give an opportunity to observe participants' real-time behaviors during the learning process. Online measurement of eye movements can entail direct measurement of cognitive load of participants in an appropriate design, which Mayer (2005) underlines as a crucial further research on Multimedia Learning.

There is plethora of studies on multimodal learning in which user behaviors were assessed. For example, Beymer, Orton and Russell (2007) conducted an eye-tracking study about how pictures influence online reading. They found that eye tracking patterns change according to the relatedness of the pictures to the text. Huang and Eades (2005) describe how people read graphs via eye tracking methodology. They argue that studies on graph reading behaviors are limited and their study is an attempt to provide systematic data about graph reading behavior. Yecan, Sumuer, Baran, and Cagiltay (2007) conducted a study in which participants’ eye movements were recorded while they followed video and text slides on a computer screen. They found that eye movements of the participants occur in a complementary way. In another study, Slykhuis, Wiebe and Annetta (2005) used eye-tracking method to observe students’ attention to PowerPoint photographs in science education settings. They used decorative and integrated photographs with text, accompanied with narration. Their results showed that participants fixated on the relevant photographs more compared to decorative ones. They also found that during narration, the difference between relevant and decorative pictures in terms of fixation is lost. They argue that the reason for this might be that participants spend more time when narration exists in slides.

Previous studies showed that eye tracking is a method which can be used to observe multimedia learning preferences (Huang & Eades 2005, Beymer, Orton, & Russell, 2007, Yecan, Sumuer, Baran, & Cagiltay, 2007, Slykhuis, Wiebe, & Anetta, 2005), and comprehension processes (Rayner et al., 2006) of participants. Although these studies include different multimedia channels into their design, individual differences, such as prior knowledge or expertise, are ignored. Since the heuristics of the multimedia design principles are affected by individual differences, inclusion of the prior knowledge to the design can be helpful to explain cognitive load changes during processing of multimodal information.

Ainsworth (2008) suggests that experiments related to multimedia learning should be renewed and varied by involving new methodologies. This is inevitable because traditional methods using black box experimentation has reached its boundaries and almost completed its life cycle. Whilst Ainsworth (2008) argues the end of the era, she praises the first generation multimedia studies for using the guidelines of experimental methods strictly and producing strong, statistically based results. Additionally, she mentions first generation experiment produced robust and replicable results, which conforms the guides and limitations of multimedia learning. As negative features for the first generation, she underlines following issues. (i) Using artificial populations with no prior knowledge, (ii) concentration only on science materials, (iii) timescale restrictions i.e. study once the given material and (iv) analyzing only learning output rather than learning process.

Tabbers, Paas, Lankford, Martens, and van Merriënboer (2008), in a chapter about studying eye movements in multimedia learning, indicate that none of the studies about CLT or CTML used eye tracking as a method to test their assumptions because of the dynamic nature of multimedia learning that makes analysis a hard job. They add that the availability of commercial programs for analyses make the eye movements studies possible. Using eye-tracking method in multimedia learning adds value by online measurement of the participant’s eye movements according to them. Although it does not cover the whole story of the interaction between text and pictures, a partial answer can be obtained by using this technology. The possible outcomes of these studies can be twofold. One is having accurate information about how pictures and text attract attention in design. Is the designer's effort to produce meaningful multimedia materials been recognized by the reader? The second is that eye tracking can yield information to compare different multimedia material materials. Thus, eye tracking can provide evidence for both theory and practice.

In this study, eye tracking and physiological measures, e.g., pupil dilation will be used as a complementary measure to the traditional measures multimedia learning such as retention or transfer tests, which will be carried out after studying the content. Furthermore, contrary to other studies in this study, prior knowledge of the participants will not be taken into account because the participants will create their own knowledge by repetitive study of the same content during experimentation. Since the content will be created by the researcher, the effect of prior knowledge about the domain will be minimized, if not completely eliminated. During the repetitive study of the same content, previous readings or listening of the content will build temporary experience about the domain. It is assumed that cognitive load of the participants will change as suggested by earlier studies and this will affect the use of multimodality.

2.4. Recent Articles on CTML, CLT and Eye-Tracking.

This section becomes a necessity because of the studies between 2008 and 2011. The framework of the current study was established in 2008. Therefore, eye-tracking tools have become widely available for many research labs. Increase in the use of eye-tracking technology resulted in several studies on the basis of CLT, CMLT, where this technique was used. Several special issues of periodicals currently exist about CLT, CMLT in addition to two books titled *Multimedia Learning* by Mayer (2009) and *Cognitive Load Theory* by Plass, Moreno and Brünken (2010).

In 2009, the Cognitive Load Theory Conference held at the Open University in the Netherlands lead to the publication of two special issues in which four articles published in "Educational Psychology Review (vol.22, 2010)" and 16 articles in "Computers in Human Behavior (vol 27, 2011). In addition articles about CLT, in 2010 6 articles published in "Learning and Instruction (vol 20, 2010)" about in which multimedia learning processes are analyzed via eye tracking. In Psychology of Motivation and Learning (vol 55, 2010) two papers published by Sweller about CLT and Mayer titled *Advances in the science of instruction* which were, unfortunately, not available. Some of these articles are cited elsewhere in this thesis.

There is another special issue about CLT in Applied Cognitive Psychology Journal in 2012 considered new directions and challenges. In introduction part, Ayres and Paas (2012) stated that new directions in CLT requires new methods to measure cognitive load used in last 20 years that is Likert type one question based on self-report. The other novelty they underlined was the use of eye tracking as an instrument to observe cognitive processes. They also mentioned that the use of animations, primary knowledge, and self-managed load as current topics. In the current study, embedding both static and animated images, study and restudy of the same content and system vs. self paced study encounter Ayres and Paas' (2012) considerations, respectively. The discussion about individual papers published in that special issue can be found in the next section.

To recap review of literature: recent developments showed that current research is covering up the agenda of CLT, CTML, and eye tracking and by focusing on further research topics. The next section explains the methodology selected to achieve this goal.

2.5. The Current Status and the Gap in the Literature

Several studies were conducted on multimedia learning in the last few decades. Despite the growing number of these studies, there is a paucity of competing theories in the area. Currently, the CTML dominates the literature in multimedia learning. This might result from its clear structure and the assumptions behind, in addition to its reference to CLT and WM. By the inclusion of cognitive load and working memory, CTML covers most of the literature regarding multimedia learning. In addition to the theoretical structure, CTML offers principles, which are easy-to-follow for instructional designers. Despite its limited number, designers could find something about their design considerations in CTML. So CTML In other words, CTML provides necessary but not sufficient conditions for successful multimedia experience. This means that in order to foster learning, multimedia presentations must fulfill the conditions set by CTML, but implementing those conditions into a multimedia presentation does not guarantee a successful design for learning. For example, the materials prepared for current study follows the principles of the CTML, but following these principles does not make the presentations as working examples of multimedia learning from the very beginning.

It would be a little bit speculative, but it seems that the principles of multimedia learning embrace only one the part of the necessary conditions for learning designs. The other part could not be derived from the studies of science but art. The artistic features of the learning integrated with the principles can lead to better results for multimedia learning. But, artistic features do not imply only color, font, and placement etc., rather the creative solutions that might not be discovered by scientists' goggles.

The current status of the multimedia learning can be seen in recent articles and books. In the previous chapter, the special issues published about multimedia learning and cognitive load theory mentioned. The special issue titled "New Directions and Challenges to Cognitive Load Theory" of Applied Cognitive Psychology Journal includes eight articles. The scope of those articles could be best tool to review current status of multimedia learning.

The special issue has eight articles. The first article focus on mental effort and the measurement of the load in which the findings reported in favor to repeated testing as measure of learning rather one test entire process ended. (van Gog, Kirschner, Kester & Paas, 2012). The second article is about narrated animation. The article by Skuballa, Schwonke & Renkl (2012) suggests that cueing external supports such as cueing and narrative pre-training have positive effects on WM capacity. The third article discusses the permanence of the spoken vs. written information. The paper suggests that if the narration is long, written form of information could be superior for learning than the narration version of the same information. (Singh, Marcus & Ayres, 2012). This finding is what is found at the very end of the study in the context of chemistry lessons. The next article reports from a different domain, radiology. They studied with students and expert radiologists on x-ray images. They found that the fixation metrics changed according to expertise on diffused and normal images. The authors discuss bottom-up processes and expertise (top-down) effects (Kok, de Bruin, Robben &. van Merriënboer, 2012). Bauhoff, Huff and Schwan (2012) tested spatial contiguity effect for finding differences between images. They found that the gaze pattern and cognitive load are two related constructs during visual search. The sixth article investigates whether training about general solving strategies can be transferred to a novel situation to solve a problem. They found that the training about general problem solving strategies could help students to solve problems in a different domain such geography (Youssef, Ayres & Sweller, 2012). The seventh article concentrated on managing cognitive load by learners themselves. It was argued that instruction about how to handle load can be utilized by learners in other settings (Roodenrys, Agostinho, Roodenrys & Chandler, 2012). In the final article, affective dimension of learning was taken into account. Hoogerheide and Paas (2012) studied the effects of the feelings on remembering. They suggest that what have been remembered can provide feedback to design more effective learning environments.

The articles mentioned above use different domains to study load and multimedia learning. However, another recent article by Johnson and Mayer (2012) showed that the classical materials of the CTML can still be profitably used; e.g., how car brakes work by integrating the testing environment into new measurement tools such as eye tracking.

CHAPTER 3

METHODOLOGY

In this study, factors affecting eye tracking measures and achievement in multimedia learning are investigated. Eye tracking technology provides an online tool to conduct experiments for boundary conditions of Cognitive Theory of Multimedia Learning (CTML) and Cognitive Load Theory as well as learning outcomes. In this study three experiments were conducted. The first experiment includes one multimedia content that is a chemistry lesson. In the second experiment, two other versions of chemistry lesson were tested with different group of participants. The third experiment, on the other hand, covers another content with two versions also. The detailed information about experiments are given later in the current chapter. The experiments in the study were designed to investigate following research questions.

1. First of all, the interrelationship between prior knowledge (familiarity) and advantages of multimedia were explored in this study. To achieve this goal, participants were exposed to the same multimedia content twice. It is assumed that studying strategies of learners have been modified in the condition of second study of the same lesson in which familiarity condition exists. This behavior change will be observed by recording eye movements in both studies of the lessons. The differences in eye movements between first and second study of lessons can be attributed to this factor, (Experiment 1).
2. The second concern of the study was finding out whether there are any relationship between eye tracking data and traditional learning outcomes of multimedia learning; such as recognition, recall, and transfer tests. The correlations between eye tracking measures and achievement scores of individual lessons will be calculated to investigate the relationship. The aim of this analysis is to investigate the relationship between eye movements and achievement scores. The existence or non-existence of correlations can provide data to build bridges between online and unobtrusive measures that are suggested by Mayer and traditional methods of measurement in multimedia learning (Experiment 1,2,3).
3. Pace of the presentation was another specific concern of the study. It is known that prior knowledge (familiarity) and pace interacts with each other in complex situations. To test this, another experiment was conducted. It is assumed that repetitive study of the same content with different pace options can yield informative results about the principles of multimedia learning (Experiment 3).
4. The last question of the study was about domains of multimedia learning. Multimedia learning studies especially in CTML focused on Mayer's classical cases. In this study, two new contents were used to investigate whether the principles of multimedia learning could be generalized to other domains or not. (Experiment 2)

To test these research questions three different experiments were conducted. The rationale behind designing experiments to test research questions and the steps followed to conduct experiments were given below.

3.1. Overall Research Design

In this experimental study, there are a series of small sized experiments to test research questions. Experiments are kept in simple because of researchers' tendency towards "minimally sufficient analysis" approach that was recommended by Wilkinson and APA Task Force to Statistical Inferences (1999). The same issue was brought to the agenda after a decade by Peterson (2009). He argues that selecting simpler methods to test research questions is a better approach rather than relying on too much complicated statistical programs and complex modeling. In this study, it is believed that keeping experimental designs simple and avoiding concurrent testing of variables produce more eloquent results, because eye tracking methodology produces huge amount of data that makes handling the data very complicated. Each experiment was conducted to address one specific question at a time. All experiments however were conducted to observe the entire process i.e. the changes in the learner's eye movements in a well-designed multimedia presentation.

In order to achieve a well-designed multimedia presentation, Mayer's suggestions (2009) were followed as much as possible. Mayer suggests three conditions, which makes multimedia presentation to work. The first condition is the existence of the words and pictures in presentation rather than only words or pictures. Words can be in the form of either narration or text. The second condition of multimedia learning is that the presentation of the materials should be simultaneous. The corresponding parts of the presentation should be on the screen at the same time. The third condition is avoiding extraneous materials, which are not directly related to the aims of the presentation. The unnecessary parts of the presentation should be eliminated according to last condition. These three conditions drawn by Mayer was applied to the testing materials of the experiments.

EXPERIMENT 1

In experiment 1, a chemistry lesson was chosen as the content of multimedia learning. The chemistry lesson designed as an example of Concise Narrative Animation (CNA) in which narration and video used as information channels. The chemistry lesson was studied twice by the same group of participants. Eye movements of participants were recorded in both study sessions, and an achievement test was applied after the second study to investigate the effect of the familiarity and the amount of learning. This version of the chemistry lesson was labeled as chemistry lesson (C).

EXPERIMENT 2

In Experiment 2, two other versions of chemistry lesson were used. In one version, subtitles were added to the chemistry lesson. The subtitles and narration were exactly same in terms of wording. This version of chemistry lesson was labeled as Chemistry with Subtitles Lesson (CWS) in the current study. CWS lesson includes a video, narration, and subtitles as information carrying channels. The last version of the chemistry lesson in Experiment 2 is the Chemistry with No Voice Lesson (CNV). In CNV lesson, the narration does not exist. The subtitles and video are the only information carrying channels in the CNV lesson. The CWS and CNV lessons were studied twice by each participant, and all participants were administered an achievement test after the second study. (Two different groups of participants were included in CNV and CWS lessons.

EXPERIMENT 3

In Experiment 3, the content of the lesson was chosen from the mechanism domain. The lesson consists of images and the text. In one version of the lesson, the presentation proceeds automatically, and in the other version of the lesson, an input from the user was required to proceed to the next slide. The input here is a keypress-either on keyboard or on mouse. These versions of mechanism lessons were labeled as System Paced Mechanism Lesson (M) and Self Paced Mechanism Lesson (MSP). Each participant studied one of the lessons twice, as in Experiment 2, but none of the participants studied M and MSP lesson together. The participants in M and MSP lessons were different.

In all experiments, the measures of eye movements were accepted as dependent variables. However, independent variables varied according to experiments. In Experiment 1, there are two independent variables. One is familiarity created by study and re-study treatment; and second is different fragments that lesson contains. One of the fragments includes list item names and pictures of materials used in chemistry experiments and the other procedures that are followed in chemistry experiments.

In Experiment 2, in addition to familiarity and fragments; narration and areas of interest were included as independent variables into the design. Areas of interest on screen and existence or nonexistence of narration in chemistry lessons is additional independent variables in Experiment 2.

In Experiment 3, familiarity is the common independent variable with Experiment 1 and 2. However, in Experiment 3, areas of interest (AOI) was defined as text and images region, instead of subtitles and video region in Experiment 2. There are two types of images in the mechanism lesson; animated and static. These, constituted another independent variable in Experiment 3. Therefore, the independent variables in Experiment 3 are familiarity, AOI, and image type. The list of these independent and dependent variables is given in Table 5.

Table 5: The List of Independent and Dependent Variables in Experiments.

	LESSON(S)	**INDEPENDENT VARIABLES**	**DEPENDENT VARIABLES**	**POSSIBLE ANALYSIS**
EXPERIMENT 1	Chemistry	Familiarity Fragments (Materials and Procedure)	Fixation Count Fixation Duration	Paired-Samples T-Test and Repeated Measure ANOVA
EXPERIMENT 2	Chemistry With Subtitles Chemistry With No Voice	Familiarity Existence of Narration Areas of Interest (Subtitles and Video)	Fixation Count Fixation Duration Visit Duration	Paired-Samples T-Test and Repeated Measure ANOVA
EXPERIMENT 3	System Paced Mechanism Self Paced Mechanism	Control (Self and System Paced) Areas of Interest (Text and Images) Type of Images (Animated and Static)	Fixation Count Fixation Duration Visit Duration	Paired-Samples T-Test and Repeated Measure ANOVA

3.2. Materials and Instruments

Repetitive study of the same content and measuring the achievement level of participants after multiple studies lie at the core of this study. This new methodology was proposed to control prior knowledge i.e. familiarity in the current study. In this method, it is expected that eye movements of the participants will yield differences among trials in terms of fixation and gaze patterns. Multimedia contents that were used in the experiments were developed by following

Mayer's (2009) suggestions. Mayer's definition for effective multimedia presentation on computers has several characteristics. First, presentation should include both words (preferably sound rather than text) and pictures. Second, words and pictures should be integrated, and be free from extraneous words, sounds, and pictures. Mayer (2009) called this type of presentation *concise narrated animation* (CNA). By the same token, the achievement test used in this study, includes not only recall and transfer questions but also recognition questions. The aim of including all types of questions in achievement test is to see whether learners show any preferences on these types of memories or not. The details about multimedia materials and achievement test are given in the following sections.

3.2.1 Multimedia Materials Used In Eye Tracking

In the current study, three chemistry and two mechanism lessons were developed to be used in three experiments. To develop these contents, chemistry and mechanism topics were selected from http://ocw.metu.edu.tr site. The METU OpenCourseWare is a free and open educational resource for faculty, students, and self-learners throughout the world. There are 118 lessons by 2012 on this site. Introductory chemistry lesson of OCW is the winner of People's Choice Award for Best Video Lectures given by education-portal in 2011. Since the OCW has gained acceptance from students and lecturers as a credible source, the lessons of current study were based on the materials on that site.

The reasons behind choosing these topics are their unfamiliarity to the participants and availability of Turkish versions. Moreover, both lessons have top visit statistics in the site. The multimedia contents used in the experiments were developed on the basis General Chemistry Laboratory Experiments (*Genel Kimya Laboratuvarı Deneyleri* in Turkish) and Mechanisms (*Mekanizma Tekniği* in Turkish) lessons on OCW. The development phases of the lessons are given below for all lessons separately.

3.2.1.1 Chemistry Lesson

The introduction to the chemistry experiment, which was published at the OCW, was about 9 minutes with background music. To use in the experiments, the original flash movie was converted to .avi format and embedded to an html page to fulfill technical requirements. In the original file, there were some unnecessary artistic segments and repetitions. In order to have a compact material and to reduce extraneous load, duration of the film was changed by re-editing. Total time was shortened to about three and a half minutes by omitting scenes without voice and standardizing the time of the presented materials. Background music was completely removed in post-editing. The short version of the chemistry lesson was vocalized by an adult female in neutral form for the purpose of this experiment.

After editing chemistry lesson with voiceover (narration), the subtitles were added to the film. The content of subtitles was exactly the same with the voice. This was the second version of the chemistry lesson. The number of words in subtitles is 258. The average length of subtitles per frame is 4.86 words. The mode of the words in subtitles is three and the longest subtitle contains thirteen words. Third version of the chemistry lesson was the no-voice version in which there were subtitles but no sound. A screenshot of the chemistry lesson with subtitles is given Figure 5. These lessons with different modalities were used in experiments 1 and 2. In Experiment 1, the chemistry lesson with narration was used as an example of CNA. In Experiment 2 Chemistry with Subtitles and Chemistry with No Voice versions were used as lessons since both have exactly the same visual presentation. The only difference between them is the lack of narration in Chemistry with No Voice lesson.

Figure 5: Screenshot Example of the Chemistry Lesson with Subtitles

3.2.1.2. Mechanism Lesson

The second topic studied in the study is the mechanisms lesson. The content of the lesson was originally prepared by Prof. Dr. Eres Soylemez from Mechanical Engineering department at METU. Unlike chemistry lesson, which was a video file, mechanism lesson was prepared as separate html pages by using SharePointDesigner©. There were seven pages total in the lesson. The first and the last pages were prolog and epilog pages. In those pages there were no information about the content. The introductory page contained information about how lesson proceeds, and last page contained information about how lesson could be ended. The rest of the pages had contained both figures and text about mechanism types and concepts. In lessons, text was placed at the left, and images were placed at the right side of the screen. The screen resolution was 1024x768 pixels. The area allocated to text was greater than the area allocated to images. Some of images were animated drawings in .gif format, while the rest were static drawings in .jpg format. There was no voice in mechanism lessons, the words were presented in visual (on-screen) form. Although this was not first choice for multimedia learning compared to the words in the narrative form, the design of the multimedia lesson was prevented the use of narration due to self paced version.

There were two versions of the lesson; the system paced and the self paced. In the system paced lesson the html pages were moving to the next page by themselves in a predetermined time interval. For determining appropriate time needed to study each page, preliminary tests were conducted before the experiments. In order to reach the optimum duration, 0.2 to 0.5 seconds per word were tested. The participants reported that 0.4 seconds per word creates a comfortable pace for reading pages. The duration of each page in the system paced mechanism lesson was calculated by multiplication of word count on page and the 0.4 seconds. Therefore, after this modification, the total presentation takes about 4 minutes to complete.

Self-paced study session included next page arrow at the right bottom of the page which does not exist in system pace version. When a participant decides to move to the next page, s/he simply clicks to the arrow with the mouse. There is no “back” option in the self paced session. The screenshot example of self paced mechanisms lesson was given at Figure 6. For example this page lasts 42 seconds in system pace version. The speed of the presentation was determined by participants in self paced version. The variance in speed resulted from individual differences prevented the use of narration in mechanism lesson.

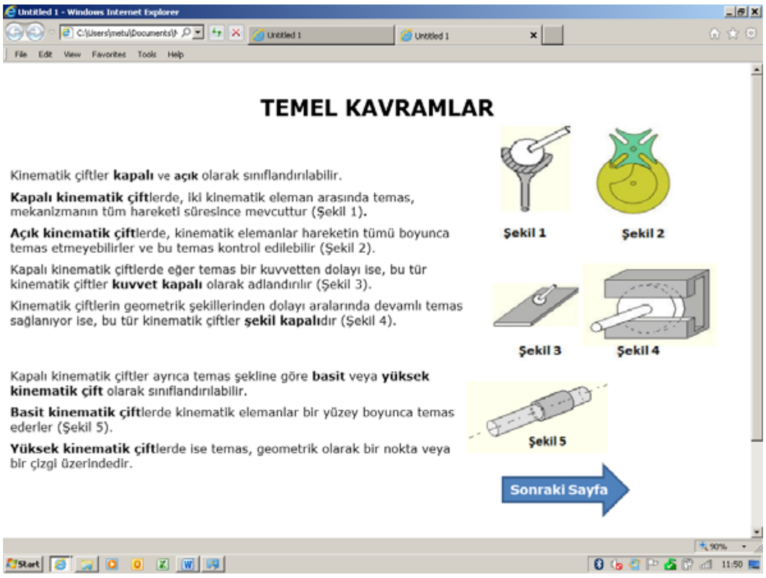


Figure 6: Screenshot Example of the Self Paced Mechanism Lesson

During the phase at the development of multimedia materials, expert views were taken for chemistry and mechanism lessons. In addition to the content, placement of the images and text were consulted by an expert on multimedia from Cognitive Science Department at Informatics Institute. Before the experiments, pilot studies were conducted to observe whether materials were compatible with eye tracking tools and the participants could follow the lessons without any problem. Pilot studies showed that; the speed, sound, colors, and fonts are quite suitable and comfortable for participants.

3.2.2. Achievement Test

The section above explains the multimedia lessons used in the study. In this section recall and recognition tests used after eye tracking recording was explained. The informed consent and questions were given in Appendix A and Appendix B respectively. The test contains of 36 questions. Nineteen of the questions were about mechanism lesson and 17 of them were about chemistry lesson. The test consisted different type of questions to measure recognition, recall or transfer. Recall and recognition questions were considered as retention questions by Mayer. They were based on remembering. The test mechanism part contained 6 recognition, 8 recall, and 5 transfer questions. Additionally, chemistry part contained 13 recall and 4 recognition questions. There are no transfer questions in chemistry part because of the definitive character of the presented material in the lesson.

In the mechanism lesson, the recall questions were fill-in-the-blanks questions. Participants had to remember the words and concepts given in the lesson. One recognition question was the multiple choice and the others were matching questions. One of the transfer questions was multiple choice, and other 4 questions were matching questions. In chemistry lesson 13 recall questions were fill-in-the-blanks questions like in the mechanism lesson. Recognition questions were true or false questions in chemistry lesson.

The scoring for multiple choice, true or false and matching questions was 1 for true answers and 0 (zero) for wrong or empty ones. Fill-in-the-blanks questions were scored 1 for true answers and as, 0 for wrong or empty ones. If the question was recalled correctly, but not exactly the way as it was presented in the lesson, they were scored 0.5 points. The total and sub-scores of these tests will be reported at the results chapter.

The test questions were prepared by the researcher. The statements in questions in terms of linguistic fluency and comprehensibility were controlled by an external reviewer. In addition to the external reviewer, the comments made by the participants of the pilot studies were taken into account before using applying tests after experiments. Reliability scores and some item statistics were given in results and conclusions section.

3.2.2. Instruments

All experiments were conducted in a controlled environment. Technology Enhanced Learning Research and Applications Lab at Middle East Technical University (METU) and Clinical Psychology Application & Research Lab at Atılım University were the two places where all data were collected. The tests were conducted individually. A Tobii 1750® eye tracker connected to a PC was used to record eye movements. The paper-based recall and recognition tests were given to all participants following each experiment. Before starting multimedia lessons, participants filled an informed consent (See Appendix A) and they were asked whether they have difficulty in vision or any obstacle to join experiment. None of the participants had reported such an excuse, which prevents them to join the experiments. After completing consent, a calibration was made by using TobiiStudio® program. Studying multimedia lessons took about 20 minutes to complete.

The data collected by the eye tracker program and tests items were scored by the researcher. The parameters recorded by eye tracker device are given in Appendix C. The metrics of eye movements calculated based on parameters are given in Appendix D. Since not all of the metrics are related to the aims of the study, some of these metrics were used in the current study. The metrics used in current study are listed in Table 6.

Table 6: Eye Tracking Metrics included in Current Study

Eye Tracking Metrics
Fixation Count
Fixation Duration
Visit Count
Visit Duration
Pupil Dilation
Percentage Fixation

3.3. Participants

All subjects are university students or personnel from METU and Atılım University. Participants joined to study voluntarily without any reward. Participants were from Computer Education and Instructional Technology (METU), Psychology (Atılım U.), and Mathematics (Atılım U.) departments.

A total of 55 subjects participated in experiments. However, two of them excluded from subsequent analyses due to different reasons. One of METU participant's eye tracking data were not eligible for the analysis due to insufficient number of fixations (i.e. less than 70%) and the other participant was excluded from subsequent analyses, because she was from a different university to omit any possible future objection. Gender, university, age, department, and grade information were taken as demographics for the participants. Twenty one participants joined Experiment 1 from METU and Atılım U. in which only C lesson was studied. Thirty two participants joined Experiment 2 from METU and Atılım U. in which chemistry with subtitles and chemistry with no voice lessons were studied. There are a total 43 participants in Experiment 3 from METU and Atılım U. in which both mechanism lessons were studied. The distribution of the participants is tabulated in Table 7 according to gender, university and experiments joined. The possible analyses about those cells is given in data analysis section.

Table 7: Participant Characteristics across the Experiments and the Types of Lessons

	Experiment 1	***Experiment 2***		***Experiment 3***	
	Chemistry	**Chemistyr No Voice**	**Chemistry With Subtitles**	**System Paced Mechanism**	**Self Paced Mechanism**
METU	**10**	**10**	**10**	**10**	**10**
Female	5	5	5	5	5
Male	5	5	5	5	5
AU	**11**		**12**	**11**	**12**
Female	6		7	6	7
Male	5		5	5	5
Total	**21**	**32**		**43**	

The mean age of the participants was 25.2 (*SD*=4.6). The mean age for Atılım U. and METU are 23.3 (*SD*=2.6) and 26.7 (SD=5.2) respectively. The mean age of the participants from METU is slightly higher than that of Atılım University. The reason for this difference is a result of the higher number of graduate students at METU. Atılım University students were consisting of only undergraduates. The mean ages of gender are almost identical 25.0 (SD=4.4) for females and 25.4 (SD=4.9) for males. The mean ages by gender within the university are higher for males from Atılım U. and higher for females from METU. But this difference is lost when all participants are taken into account.

3.4. Procedure

Materials and Instruments section explains the structures of chemistry and mechanism lessons in addition to achievement test used in the experiments. In this section, the settings and the sequences of the experiments were stated. The two lessons with five different versions used in three experiments were tabulated in Table 8. However, the data in experiments were not collected in serial order. They were applied concurrently. Table 8 shows the distribution of lessons to experiments. The three version of chemistry lesson is used for Experiment 1 and 2, and mechanism lessons used of Experiment 3.

Table 8: The List of Lessons Used in the Experiments

Lesson	***Format***	***Voice/Text***	***Pace***	***Name Given***	***# of Experiment***
Chemistry	Video	Voice	System	Chemistry (C)	Experiment 1
Chemistry	Video	Voice+Text	System	Chemistry with Subtitles (CWS)	Experiment 2
Chemistry	Video	Text	System	Chemistry with No Voice (CNV)	Experiment 2
Mechanism	HTML	Text	System	Mechanism System Paced (M)	Experiment 3
Mechanism	HTML	Text	Self	Mechanism Self Paced (MSP)	Experiment 3

The timeline for experiments is given in Figure 7 in which the position of experiments are depicted. The empty boxes with black outline point Experiment 1, orange outlines point Experiment 2, and green outlines point Experiment 3. The distribution of content, participants and the timing of achievement test are given also in Figure 7. The rationale behind this methodology is maximizing the amount of data collected from participants. Since there are no hypothetical connection between experiments, using same group of participants in experiments creates no problem. In addition to mixing the experiments, the order of lessons were also counterbalanced to exclude any recency or primacy effect in experiments. The procedure for individual experiments is explained below.

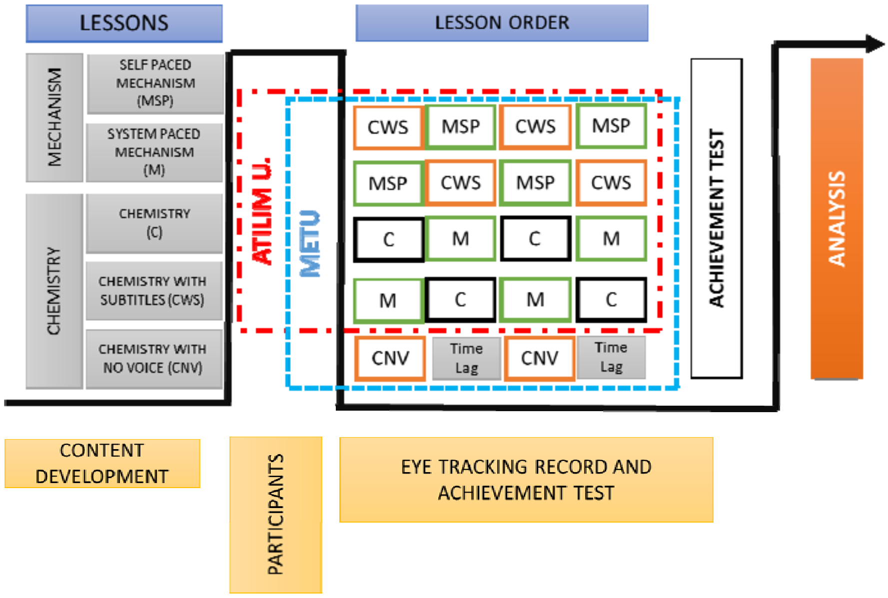


Figure 7: The timeline of the Experiments.

3.4.1. Procedure of Experiment 1

There is only one chemistry lesson with narration (C) in Experiment 1 that was studied twice. The participants in Experiment 1 are from both Atılım U. and METU. The participants of Experiment 1 are also the participants of Experiment 3. The order of the lessons is switched to avoid maturation effect. After studying chemistry lesson twice, participants took achievement tests just after the eye recording session. There are three measures in Experiment 1, eye tracking records of the first and the second study of individuals and an achievement test score for the chemistry lesson.

3.4.2. Procedure of Experiment 2

In Experiment 2, there were two types of chemistry lesson that are chemistry with subtitles (CWS) and chemistry with no voice (CNV) lessons. The participants of those lessons were from Atılım U. and METU also. The groups of participants in Experiment 1 and 2 are different. Participants of Experiment 1 did not join Experiment 2 since both content is the same except existence of subtitles in Experiment 2. However, the participants of Experiment 2 are also the participants of Experiment 3 as in Experiment 1. But unlike Experiment 1, the participants of Experiment 2 studied self paced mechanism lesson instead of system paced mechanism lesson. The CWS lesson is studied from participants from both university, but CNV lesson is conducted only in METU, because of technical reasons. Since there was no third type of mechanism lesson, the participants of CNV lesson studied only one lesson with time breaks. In the breaks, there were distractors whose functions are to empty short-term memory.

3.4.3. Procedure of Experiment 3

In Experiment 3, the content is different from Experiment 1, and Experiment 2 in which two version of mechanism lesson were used. Two version of mechanism lesson applied twice to the two different group of participants are system paced mechanism lesson (M) and self paced mechanism lesson. Two group of participants from Atılım U. and METU studied twice same lesson in both groups, after an achievement test applied those groups. The participants of Experiment 3 also studied one of the chemistry lesson in changing order to increase the amount of data collected since Experiment 3 was totally different from Experiment 1 and Experiment 2 due to its content. No interaction expected in both types of experiments. The methods of analyzing data are given next section.

3.5. Data Analysis

The research questions listed in Section 1.3 point the purpose of the current study. These questions emphasize the boundary conditions of CTML. The multimedia literature suggests that further research should be done in these areas. These questions encompass most of the CTML's missing connection between principles. These problems are generally converged on prior knowledge, pace, and domain invariability. Although there are many emerging new studies about this topic, perseverance on conventional methodologies and closedness of research groups have carrying out the danger of failing to notice the need of new visions for theory.

Fortunately, this is not the case among the theoreticians in multimedia learning. Recent studies underlie the need for new methodologies. Eye tracking supported by other methods can be one of the usable tools to study multimedia learning (Mayer, 2010). In addition to eye tracking Moreno and Park (2010) suggest that more authentic settings are required to study multimedia learning. Using eye tracking supported by studying same content more than once might satisfy suggestions of leading theoreticians in this area.

Current studies' novelties inherently force the design to be more complicated, keeping analysis simpler the variables analyzed separately, even the risk of increasing type one error is preferred for analyses. Studying the same content more than once is the first novelty in this study. Measuring multimedia factors by using eye tracking is another novelty. Using other domains other than classical ones for exploring multimedia principles is the last novelty. To find out the factors effecting eye tracking measures and achievement in multimedia learning three experiments were designed in the current study. The aims of the study are to find out how studying the same content more than once in different presentations interacts with CTML's basic assumptions.

Some of the inherited independent variables in the design such as university, gender and age will be analyzed by repeated measure t-test before analyzing hypothesis, because no significant differences are assumed between the groups. If results will be as expected, these variables will be eliminated from further analyzes to keep results easy-to-follow. In addition to variance resulted from demographics of participants, there are designedly purposefully included variables such as, familiarity, content, and control. The gender, age, and university variables will not be discussed for each experiments, however the multimedia design issues will be discussed in detail for all experiments in the light of data obtained by eye tracking and achievement test. The methods applied for analyses are given below for each experiment separately.

In addition to eye tracking data, an achievement test was applied in the study. The analysis conducted for the test is given at end of the current section.

3.5.1. Data Analysis for Experiment 1

In the Experiment 1 chemistry with narration lesson was used. The lesson studied twice with same group of people. Two set of eye movement data was obtained for each study. The differences in terms of eye movement data will be reported in results chapter.

Since the same group of participants studied lesson twice, paired sample t-test analysis was applied for chemistry with narration lesson. The independent variables in this experiment are familiarity, different fragments of lesson in addition to gender and university. One way of exploring the differences between the first and the second study in terms of eye tracking measures is the application of two paired sample t-tests in this section. The other possible analysis in this section is 2 (Familiarity) X 2 (Fragments) repeated measure ANOVA analysis to explore the possible interaction effect.

The dependent variables in Experiment 1 are the eye tracking measures listed in Table 6. But only fixation duration and fixation count will be reported in this experiment. The only Areas of Interest in this experiment is the whole screen, as a result the visit (gaze) statistics were not valid for this experiment.

3.5.2. Data Analysis for Experiment 2

In Experiment 2, all variables of Experiment 1 are used. Moreover, Experiment 2 consists of two different versions of chemistry lessons studied twice with different groups. Consequently, there will be between group comparisons in addition to within (repeated) group comparisons in Experiment 2. The between group comparisons will include comparisons between chemistry with subtitles and chemistry with no voice lessons in terms of eye tracking metrics. Visit durations as dependent variable will come to picture in Experiment 2, because of existence of different areas of interests on screen namely, subtitles area and video area.

In Experiment 2, the between group comparisons will be made by using independent group t-test with unequal sample size because of the number of participants in chemistry with subtitles and chemistry with no voice lessons are not equal.

3.5.3. Data Analysis for Experiment 3

In Experiment 3, the content of the lesson is different from Experiment 1 and Experiment 2. The content is replaced with mechanism instead of chemistry in Experiment 3. The groups of participants joined to Experiment 3 is different for system paced mechanism and self paced mechanism lessons although both presentation is same except for pace or control given to the participants. The data analysis procedure for Experiment 3 will consist of between group comparisons for independent t-test for equal groups since both groups have same number of participants.

In addition to t-tests there will be a repeated measure ANOVA for Experiment 3, because of existence different areas of interest regions, namely text area and images area. The images area also divided into two as static and dynamic images. So the analysis will include both within and between group comparisons for Experiment 3. Within comparisons can based on static and dynamic image differences for the same lessons and between group comparisons can based on image type control. So, the comparisons can be made by 2 (AOI) X 2 (Image Type) X 2 (Control).

3.5.4. Data Analysis for Achievement Test

The structure of achievement test is given in previous section. All participants took achievement test, after studying each of the lessons. The descriptive statistics about the test will be given in results section. These will include means, standard deviations, and tendency values in addition to reliability scores for gender and age for chemistry and mechanism tests separately.

The correlations between achievement test and eye tracking measures will be also given for different lessons to discuss whether eye tracking measures can be connected with traditional methods in results section. The possible implications of correlations will be discussed in discussion and implications chapter.

3.6. Limitations of the Study

The current study has some common limitations as previous research about multimedia learning. The first common limitation is that the nature of the multimedia materials. They were developed specifically to be used in this study. The second limitation is the subjects' affiliations; they were either university undergraduates or graduates. These limitations are valid most of the multimedia studies like the present one.

The third limitation is not specific to neither multimedia learning theories nor to the current study. Current study is a fine example of actual practice of social science (i.e. neither theory driven nor data driven). It keeps both deductive and inductive reasoning in its design. In other words, the research questions do not include directional hypotheses. They enclose the current status of the theories and suggest new methods to study those theories by testing boundary conditions. The study is not only designed as an inductive study to test hypotheses based on theories, but also it follows a data driven deductive methodology to find out the relation between theory and data. CTML has a theoretical model, but CLT has never been a learning theory (Moreno and Park, 2010). CLT's primary consideration is to decode the relations of cognitive systems rather than suggesting methods to achieve better learning. Unlike CLT, CTML is a full multimedia theory, which suggests heuristics for better learning. The limitation in this case is that not having a directional hypothesis to be confirmed or not.

On the other hand, some of the limitations resulted from design in this study are not common with others. One of the limitations of this type is the nonexistence of follow up study. Although the level of prior knowledge change considerably by time, the time lag used in the study is not more than half an hour. Although this time is longer than STM's retention time that is about 30 seconds, increasing or changing time between study and test could produce more real world-like and valid results.

The other limitation in this study is the lack of systematic reliability and validity studies in the pilot study. There were no formal reliability and validity analyses as a pilot phase in the study. In the development phase of recall and recognition test with lessons, few expert reviews were taken in addition to eye tracking trials but these studies had not been statistically verified. The pilot studies based on individual reports of experts, and participants of pilot studies.

The sampling of the study can be other limitation of current study. The number of participants is at the border of acceptable numbers. The lack of randomization selection of participants and assigning them into cells in experiments weakens the generalization power of study. The results should be considered as peculiar to current study in discussion and implications.

The last limitation of the current study is the lack of affective domain in the design. Motivation of the participants controlled by neither researcher nor by the materials. Future research where motivation was also taken into consideration is required. The inclusion of affective domain could enable using qualitative measures in study, which is unavailable currently. The study based on only quantitative measures, however, using self-reports or online think-aloud protocols during eye tracking might be studied to observe participants' behaviors.

CHAPTER 4

RESULTS AND DISCUSSION

In this section, an overview of results will be given at the beginning. This outline will include the descriptive statistics for achievement test and multimedia lessons. After giving the descriptive statistics for each lesson, the experiments will be reported by the help of inductive statistics. At the end of the chapter, the overall comparisons will be made between achievement test and experiments to sum up.

4.1. An Overview of Results

The outline has two main parts that are achievement test results and eye tracking data results. The achievement test conducted just after completing the study session of multimedia content consisted of recognition, recall and transfer questions about chemistry and mechanism lessons. The test materials were described in the previous chapter. The next section contains the analyses about achievement test.

4.2. General Results of Achievement Test

There are 36 questions in recall and recognition test. Nineteen of them are about mechanism lesson and 17 are about chemistry lesson. Forty three participants answered both chemistry and mechanism questions. Ten participants who studied only Chemistry without Voice (CNV) lesson answered only chemistry questions. The scoring for recall questions was 1 point for each correct answer, and 0.5 point for incomplete answers. Incomplete answers are the correct answers but they are not recalled as they taken place in lessons. Recognition questions were assessed as true (1 point) or false (0 point). Five transfer questions contained one multiple choice question and 4 matching questions scored either true or false that is 1 or 0.

The mean score of the chemistry test is 8.23 (*SD*=2.44) out of 17 and 6.45 (SD=3.84) for mechanism test out of 19. The maximum score or chemistry lesson was 13 achieved by two of the participants, and the maximum score is 15 in mechanism lesson for one participant. The distribution of the scores among participants for chemistry and mechanism lessons are given in the Figure 8 and Figure 9, respectively. The reason for giving these distributions are to control their normality, since the smaller values than 1 for central tendency measures can be accepted as normality indicators for distributions.

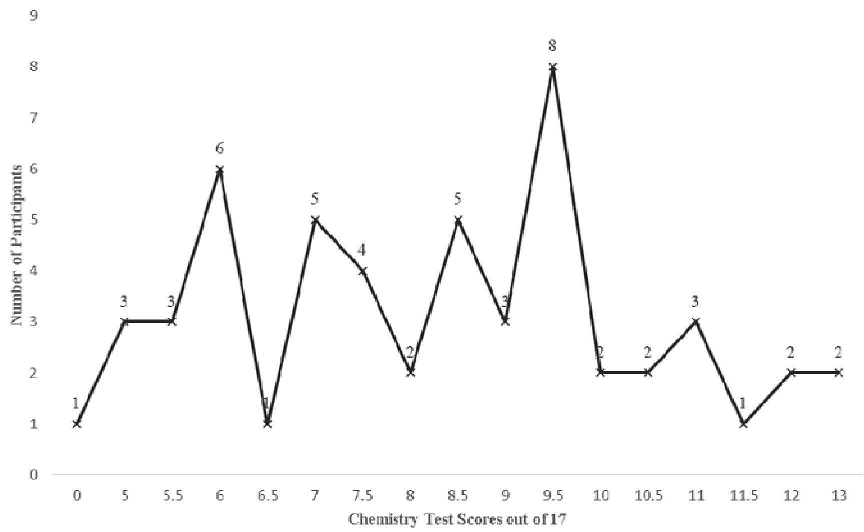


Figure 8: The Distribution of Recall and Recognition Test Scores – Chemistry (N=53).

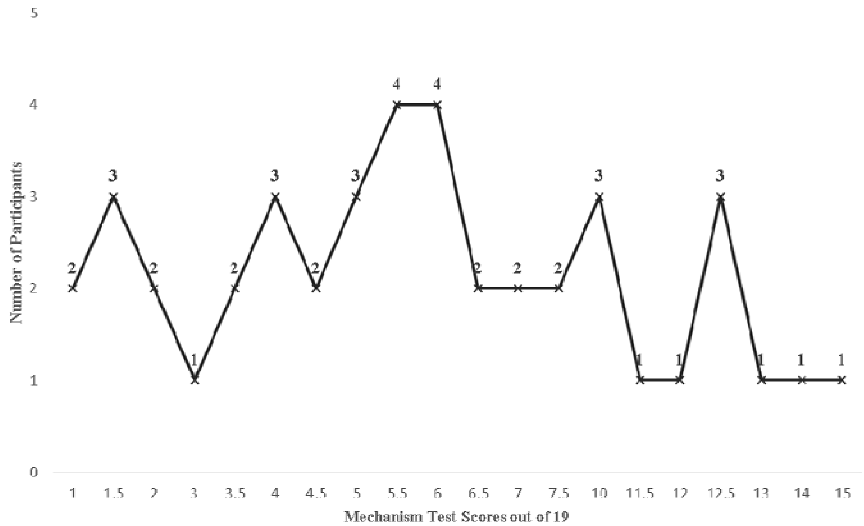


Figure 9: The Distribution of Recall and Recognition Test Scores – Mechanism (N=43).

A quick look at the histograms as an easy measure of skewness and kurtosis to visualize the distribution of participants' scores does not indicate extreme results to worry on. Skewness of the distribution for chemistry and mechanism scores are .31 and .62 respectively. Skewness values imply that chemistry scores were a little more symmetrical when compared to mechanism scores. Chemistry and mechanism tests were unlike in kurtosis values. The distribution of the chemistry test is more peaked that mechanism test which is flatter than a normal distribution. The kurtosis value was 1.31 for chemistry test, and -.52 for mechanism test. The skewness values smaller than 1, is an indication assessment for normality assumptions in achievement tests.

Although the success rate (mean scores) for the test seem to be low if it had been a classroom achievement test; these scores are acceptable for the aims of the current study. Since the test scores have not been placed to the focal point in overall design unlike the eye tracking measures, these scores showed that participants learn something at the level of recognition, recall, or transfer level. The following sections contain more detailed descriptive analyses about scores by taking into account variables such as university, gender and age.

4.2.1. Test Scores by University

In the previous section, descriptive statistics are given for all participants. In this section, the scores are analyzed according to participants from different universities. Participants were students of two different universities in the study. One is METU and the other is Atılım University. The METU students were from Computer Education and Instructional Technology and Atılım University students were from Psychology department mainly.

The mean scores for mechanism and chemistry tests are given in the Table 9. Table shows that the scores of METU are higher than Atılım University scores; $t(30) = 5.9$, $p=.000$. The mean scores of mechanism test for METU students (M=9.28; SD=3.46) higher than Atılım University students (M=4.00; SD=2.06). This gap for chemistry test is not as high as mechanism scores but still significant, $t(51) = 4.4$, $p=.000$. The mean scores for chemistry test are 6.78 (SD=2.10) for Atılım University and 9.33 (SD=2.05) for participants from METU. There are 23 participants from Atılım University for both test and 20 participants in mechanism lesson and 30 participants in chemistry lesson from METU. Although that was not present in demographics questions, when participants asked whether they took chemistry lesson or not in university level, their responses was "no".

Table 9: Mechanism and Chemistry Test Scores by University

University	**Mechanism Test**		**Chemistry Test**	
	Mean	**SD**	**Mean**	**SD**
Atılım U.	4.00	2.06	6.78	2.10
METU	9.28	3.46	9.33	2.05

4.2.2. Test Scores by Gender

In the previous section, the test scores have been analyzed university by university. In this section, scores will be grouped according to gender. There are total 28 female, 25 male participants from two universities. The observed scores of females and males almost identical in both tests (Table 10). Same pattern can be observed when scores are reported by gender and university.

Table 11 showed that the mean test scores of either chemistry or mechanism test are similar and the differences are insignificant; $t_{mech.}(41) = .23$, $p=.82$ and $t_{chem.}(51) = .52$, $p=.60$. Since gender has no visible effect on results as expected, no further calculation has been made for gender about test scores.

Table 10: Mechanism and Chemistry Test Scores by Gender.

University	**Mechanism Test**		**Chemistry Test**	
	Mean	**SD**	**Mean**	**SD**
Female	6.33	3.87	8.39	2.40
Male	6.60	3.89	8.04	2.47

Table 11: Mechanism and Chemistry Test Scores by University and Gender.

University **Gender**	**Mechanism Test**		**Chemistry Test**		**n**
	Mean	**SD**	**Mean**	**SD**	
AU	**4.00**	**2.06**	**6.78**	**2.10**	**23.00**
Female	3.92	2.15	7.35	1.75	13
Male	4.10	2.04	6.05	2.37	10
METU	**9.28**	**3.46**	**9.33**	**2.05**	**30.00**
Female	9.45	3.37	9.30	2.57	15
Male	9.10	3.72	9.37	1.45	15

4.2.3. Reliability and Item Statistics of Achievement Test

As mentioned earlier, items were prepared by researcher and consisted of several type of questions. These are true/false, fill in the blanks, multiple choice and matching questions. The basic descriptive statistics of these items such as mean, standard deviation and percentage of true and half true responses (item difficulty) are presented here. Further statistics such as item discrimination or validity issues will not be considered in this section.

Overall results for chemistry and mechanisms tests are given in the previous section. Since these tests consisted of recall, recognition and transfer questions, these three different sets of questions have analyzed independently.

The reliability score for mechanism test for 19 questions is .87 for full scale and .93 for split half reliability. The reliability score for chemistry test for 17 questions, on the other hand, .66 for both full scale and split half reliability.

The overall difficulty for the test for all participants is 40.78. The mean of item difficulty scores are p_{chem}=48.39 for chemistry questions and p_{mec}=33.97 for mechanism questions. The p value for mechanism test is lower than chemistry test in total. This difference can be resulted from the content of the lesson or the nature of the questions. The differences between lessons mentioned in methods chapter. The number of recall questions in test is 21 (13 in chemistry test, 8 in mechanism test) and number of recognition questions is 10 (4 in chemistry test and 6 in mechanism test. The difficulty score for transfer questions in mechanism test is obtained as p_{trans}=29.77. The mean difficulty score for recall questions is p_{recall}=38.41. The p value for recognition test is calculated higher than recall and transfer questions that is $p_{recognition}$=51.25 as expected. The p scores for recall and recognition test is tabulated in Table 12 and Table 13. The results show that the difficulty of the items did not depend on the type of questions. The recall Qs of chemistry and mechanism parts are almost identical, whereas recognition questions of chemistry part are quite easier when compared to questions of mechanism part.

Table 12: Item Difficulty Scores for Recall, Recognition, and Transfer Questions by Test Type.

Test/ Question Type	Mean of p value	Number of Qs
Chemistry	**48.39**	**17**
recall	39.33	13
recognition	77.83	4
Mechanism	**33.97**	**19**
recall	36.92	8
recognition	33.53	6
transfer	29.77	5
Average	**40.78**	**36**

Table 13: Item Difficulty Scores for Recall, Recognition, and Transfer Questions by Question. Type

Question Type/ Test	Mean of p value	Number of Qs
Recall	**38.41**	**21**
chemistry	39.33	13
mechanism	36.92	8
Recognition	**51.25**	**10**
chemistry	77.83	4
mechanism	33.53	6
Transfer	**29.77**	**5**
mechanism	29.77	5
Average	**40.78**	**36**

The *p* scores for individual questions are given Figure 10. The hardest question was a recall question that is graduated cylinder (*mezur* in Turkish) answered less than one tenth of participants (p_{mezur}=9.43). The easiest questions were two recognition questions in chemistry test (p_{evet}=98.11). In these questions participants were asked to remember whether the study material contained a still scene or not. Participants performed superior in this recall task, but if they were asked to recall a scene that is not contained in the chemistry lesson, their responses were unstable. One has *p* value 24.53 the other has 90.57. The harder one had mislead most of the participants, which is resembled one of scenes in test.

The transfer questions had place in the test because of learning is described as transfer in multimedia learning. However, these two questions' difficulty obtained as 29.77, which is at the bottom of the difficulty distribution among mechanism test. This finding is in line with the multimedia literature in which transfer is assumed as state that can be achieved after fulfilment of recall and recognition in learning. However, to be more decisive, discrimination values should be calculated, but these statistics are out of the scope in in this study. So, these part of the study remains unfinished and requires further studies.

The item statistics of the tests showed that learning can be achieved through recognition, recall, and transfer questions. Relaying only transfer questions could not be only possibility to measure advantages of multimedia principles. Although transfer is the most preferred outcome for learning, recall and recognition tests can be used as supplementary conditions for multimedia learning.

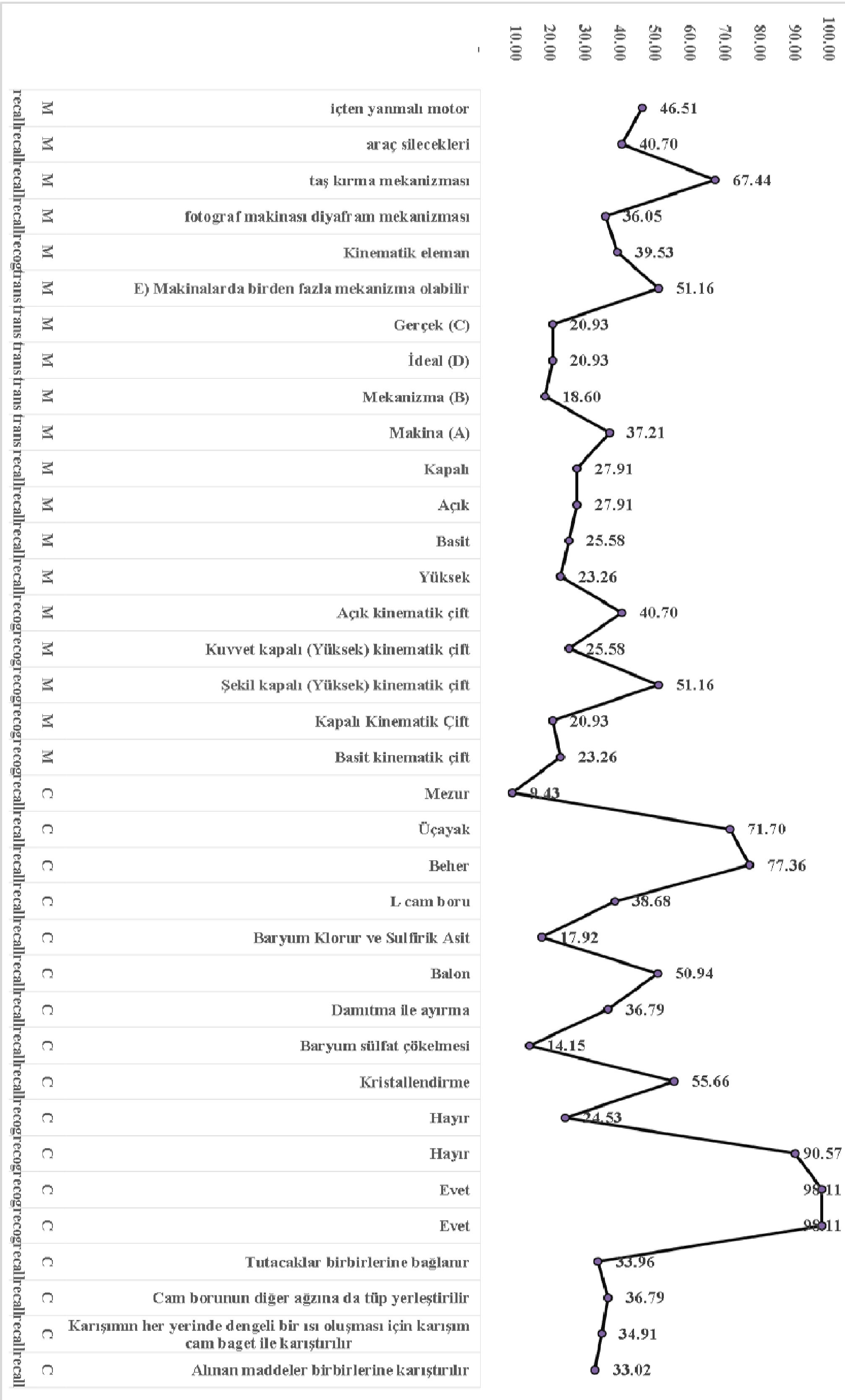


Figure 10: Item Difficulties of All Questions in Achievement Test.

4.3. Correlations between Achievement Scores and Eye Tracking Measures

In this section the correlations between achievement test scores and selected eye tracking measures are given. The fixation metrics are chosen for correlation matrix because, the whole screen taken into account. When total screen is taken into account, the correlations about visit metrics became meaningless. Table 14 shows that there are some correlations between test scores and eye tracking metrics. Mechanism lesson test has correlation with total fixation duration in first reading, $r(19)=-.36$, $p<.05$. Self paced mechanism lesson has correlations in first reading with fixation duration, $r(20)=-.43$, $p<.05$, and both readings with total fixation duration, $r(20)=-.60$, $p<.01$; $r(20)=-.36$, $p<.05$. Chemistry lesson has correlation in second reading with total fixation duration, $r(19)=-.40$, $p<.05$. Chemistry with no voice lesson has correlation in first reading with total fixation duration, $r(20)=-.66$, $p<.01$. Chemistry with subtitles lesson has significant correlations for all fixation metrics in both readings that are FC, FD, and TFD; $r(8)=-.36$, $p<.05$, $r(8)=-.43$, $p<.05$, $r(8)=-.62$, $p<.01$, $r(8)=.53$, $p<.01$, $r(8)=.43$, $p<.05$, and $r(8)=.37$, $p<.05$ respectively. Since, the existence or non-existence of correlations between test scores and eye tracking measure might be used as a connection between online and traditional measures without any causality. As a result, the correlation measures seem to be worth study in further studies. The possible uses of this connection will be revisited in discussion and conclusion chapter again.

Table 14: Correlations between Achievement Scores and Eye Tracking Measures.

Lessons	Fixation Metric						
Record	Fixation Count (FC)		Fixation Duration (FD)		Total Fixation Duration (TFD)		*n*
	First	Second	First	Second	First	Second	
Mechanism	.32	.27	.04	.17	-.36*	.32	21
Self Paced Mechanism	.62**	.24	.28	-.43*	-.60**	-.36*	22
Chemistry	.05	.02	-.03	.09	.34	.40*	21
Chemistry with No Voice	.11	-.33	.07	.48	-.66**	-.28	22
Chemistry With Subtitles	-.36*	-.43*	.62**	.53**	.43*	.37*	10

**p*<.05

***p*<.01

4.3. Results of Experiment 1

The Experiment 1 conducted to investigate the effect of the familiarity on a concise narrative animation that is a chemistry lesson with narration in background in that case. The description of lesson is given in methodology chapter. The design of Experiment 1 consist of several intrinsic and extrinsic factors (For a recent review see Grissom & Kim, 2012). The intrinsic factors in design are gender, age, and university of participants. The extrinsic factors are fragments and familiarity. Fragments variable has two level, one is experiment part in which the procedures are explained and the other is materials part in which the materials used in experiments are listed. The familiarity variable refers to treatment created by first and second study of lesson.

The results of the Experiment 1 analyzed in two steps. First step consist of independent t-test analyses for intrinsic factors, which are not central to the study. The second step include 2x2 factorial design for extrinsic factors that are aimed to investigate the effects of content of multimedia presentation (fragments) and the prior knowledge (familiarity).

The group statistics of independent sample t-test analysis for gender are given in Table 15. Table 16 presents independent t-test Results of Eye Tracking Metrics for Gender in Chemistry Lesson. Six independent-samples t-tests indicated that scores were not significantly different for gender in terms of eye tracking measures recorded in first and second study. The appropriate eye tracking results in Experiment 1 are fixation count, fixation duration, and total fixation duration.

Since the Levene's test for equality of variances indicates unequal variances for the mean of fixation durations in first reading, the corrected degrees of freedom is reported in that row. Six independent-samples t-tests indicated that scores were not significantly different for gender in terms of eye tracking measures recorded in first and second study. The appropriate eye tracking results in Experiment 1 are fixation count, fixation duration, and total fixation duration. Since the Levene's test for equality of variances indicates unequal variances for the mean of fixation durations in first reading, the corrected degrees of freedom is reported in that row.

Table 15: Group Statistics of Eye Tracking Metrics for Gender in Chemistry (C) Lesson

Group Statistics					
Gender		N	Mean	Std. Deviation	Std. Error Mean
Mean of Fixation Count of First Reading on Chemistry Lesson	Female	11	442.09	43.80	13.21
	Male	10	438.10	97.44	30.81
Mean of Fixation Count of Second Reading on Chemistry Lesson	Female	11	430.64	66.78	20.13
	Male	10	433.40	71.21	22.52
Mean of Fixation Duration of First Reading on Chemistry Lesson	Female	11	0.45	0.04	0.01
	Male	10	0.46	0.13	0.04
Mean of Fixation Duration of Second Reading on Chemistry Lesson	Female	11	0.46	0.06	0.02
	Male	10	0.45	0.10	0.03
Mean of Total Fixation Count of First Reading on Chemistry Lesson	Female	11	195.59	7.59	2.29
	Male	10	191.24	9.53	3.01
Mean of Total Fixation Count of First Reading on Chemistry Lesson	Female	11	192.94	8.03	2.42
	Male	10	187.70	10.32	3.26

Table 16: Independent t-test Results of Eye Tracking Metrics for Gender in Chemistry (C) Lesson

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference: Lower	Upper
Mean of Fixation Count of First Reading on Chemistry Lesson	4.3	.052	.123	19	.903	3.99	32.42	-63.87	71.86
Mean of Fixation Count of Second Reading on Chemistry Lesson	.001	.980	-.092	19	.928	-2.76	30.11	-65.79	60.26
Mean of Fixation Duration of First Reading on Chemistry Lesson	9.3	.007	-.357	10.6	.728	-0.01	0.04	-0.11	0.08
Mean of Fixation Duration of Second Reading on Chemistry Lesson	.91	.352	.271	19	.789	0.01	0.03	-0.06-	0.08
Mean of Total Fixation Count of First Reading on Chemistry Lesson	.716	.408	1.163	19	.259	4.35	3.74	-3.48	12.19
Mean of Total Fixation Count of First Reading on Chemistry Lesson	.783	.387	1.305	19	.208	5.24	4.01	-3.16	13.64

As a result of independent sample t-test values, gender as an intrinsic independent variable ignored in the rest of analysis in Experiment 1. Table 17 and Table 18 show the results for group statistics and independent t-test results of eye tracking metrics for university in chemistry lesson like in gender. Results showed that all none of the differences for universities are significant in terms of eye tracking metrics. As in the gender, the analyses for university as a factor excluded from following factorial designs.

Table 17: Group Statistics of Eye Tracking Metrics for University in Chemistry (C) Lesson

Group Statistics					
University		N	Mean	Std. Deviation	Std. Error Mean
Mean of Fixation Count of First Reading on Chemistry Lesson	AU	11	438.18	87.19	26.29
	METU	10	442.40	56.37	17.83
Mean of Fixation Count of Second Reading on Chemistry Lesson	AU	11	435.18	72.86	21.97
	METU	10	428.40	64.07	20.26
Mean of Fixation Duration of First Reading on Chemistry Lesson	AU	11	0.46	0.11	0.03
	METU	10	0.45	0.06	0.02
Mean of Fixation Duration of Second Reading on Chemistry Lesson	AU	11	0.44	0.08	0.02
	METU	10	0.46	0.08	0.02
Mean of Total Fixation Count of First Reading on Chemistry Lesson	AU	11	191.64	10.33	3.11
	METU	10	195.59	6.19	1.96
Mean of Total Fixation Count of First Reading on Chemistry Lesson	AU	11	187.10	10.90	3.29
	METU	10	194.12	5.79	1.83

Table 18: Independent t-test Results of Eye Tracking Metrics for University in Chemistry Lesson

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Mean of Fixation Count of First Reading on Chemistry Lesson	0.70	0.41	-0.13	19.00	0.90	-4.22	32.42	-72.08	63.64
Mean of Fixation Count of Second Reading on Chemistry Lesson	0.02	0.90	0.23	19.00	0.82	6.78	30.08	-56.17	69.73
Mean of Fixation Duration of First Reading on Chemistry Lesson	2.06	0.17	0.20	19.00	0.84	0.01	0.04	-0.08	0.09
Mean of Fixation Duration of Second Reading on Chemistry Lesson	0.07	0.79	-0.64	19.00	0.53	-0.02	0.03	-0.09	0.05
Mean of Total Fixation Count of First Reading on Chemistry Lesson	2.86	0.11	-1.05	19.00	0.31	-3.95	3.77	-11.84	3.93
Mean of Total Fixation Count of First Reading on Chemistry Lesson	3.29	0.09	-1.81	19.00	0.09	-7.02	3.87	-15.11	1.08

In second steps, the other independent variables in Experiment 1 are fragments and familiarity that are extrinsic factors by design will be investigated. To test the effects of these variables on eye tracking metrics a 2x2 repeated measure factorial design analysis applied since both variables have two levels. The independent measures in are fixation metrics of participants. Three different repeated measure analyses will be reported in this section for the means of fixation count, fixation duration and total fixation duration.

4.3.1. Results for Fixation Count per Second in Experiment 1

Before giving the results of the experiment, some points should be underlined. First, the durations of experiment and materials fragments are different which can be affect the number of fixations and total fixation duration. To omit the possible effects of the different durations for materials and experiment fragments, the mean values of fixation count and total fixation duration metrics per second values calculated by dividing length of each fragments. The second point is the sphericity assumption that is to control the equal variances in within groups for repeated measures of ANOVA. The sphericity is controlled by Mauchly's Test of Sphericity in repeated measures factorial designs. Since, Mauchly's Test of Sphericity control the covariances for different treatments in repeated measure, it is required minimum three levels for one of the dependent variables at least. Although, availability of correction methods for the conditions in which sphericity assumptions have not been met; the two level with two factorial designs does not need a test to control sphericity because of non-existence of the covariance resulted from a third variable (Grissom & Kim, 2012).

To calculate the F values of factorial designs, there are several methods. These are Wilk's lambda, Pillai's trace, Hotelling-Lawley trace and Roy's largest root in which variance in dependent variable calculated differently. The results of last three measure are converges when the number of participants increased. The differences in calculations resulted from using discriminant analyses or the use of eigenvalues between them. Tabachnick and Fidell (1996) reported that when an effect has only two levels effects those values turn into be identical. Additionally, if there is more than two level, the F values can be slightly different but, significances computed identical also. In this study Wilk's lambda will be reported because of two reasons, one is it demonstrates the amount of variance in the dependent variable produced by the independent variable and the second is it became the most frequently reported measure in repeated measure analyses, since its interpretation is the easiest one among others.

Table 20 shows within subjects design results for mean of fixation count per second. The two main effects in design are fragments and familiarity. There is a significant main effect of fragment in Experiment 1, $F(1,20)=46.92$, p=000. Participants' mean of fixation counts per second differ significantly on experiments and materials areas. But the main effect of familiarity is not significant, $F(1,20)=.167$, p=.687, that is studying same material second time does not change the mean of the fixation count per second. The interaction between fragment and familiarity is also insignificant also, $F(1.08)=1.08$, p=.311. The partial eta squares the effect size of fragment is large (partial η^2=1), but familiarity (partial η^2 =008) and interaction (partial η^2=.05) are small. Similarly, observed power is large for fragments, but quite small for familiarity and interaction effects.

Table 19: Descriptive Statistics for Means of Fixation Count per Second in Experiment 1

Descriptive Statistics				
Fragments	Familiarity	Mean	Std. Deviation	N
Experiment	First Study	1.95	0.42	21
	Second Study	1.96	0.37	21
Materials	First Study	2.43	0.46	21
	Second Study	2.35	0.45	21

Table 20: Repeated Measures Design Results of Mean of Fixation Count per Second in Experiment 1

Tests of Within-Subjects Contrasts							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Fragments	3.941	1	3.941	46.922	.000	.701	1.000
Error(Fragments)	1.680	20	.084				
Familiarity	.025	1	.025	.167	.687	.008	.068
Error(Familiarity)	2.951	20	.148				
Fragments * Familiarity	.051	1	.051	1.079	.311	.051	.168
Error(Fragments*Familiarity)	.938	20	.047				

a. Computed using alpha = .05

Before going to the discussion of the results, another justification is needed here that is the strength of the association or effect size and power, since it is required many editors or journals from articles to be published. For example, Journal of Educational and Psychological Measurement and at least 22 other journals as of 2012 require the reporting of effect size measures (Grissom & Kim, 2012). The effect size measures the strength of the association between dependent and independent variable. But, two decades ago, Cohen (1992) has complained about the availability of power reports in articles although there were no controversy about the methodologies and importance of power since its suggestion in 60's. The occurrence of effect size and power analysis increased after 90's but the way of using them by researchers still have problems especially in factorial designs (Pierce, Block, & Aguinis, 2004). They discuss the measures of strength of associations reported for in analysis of variance. They include, eta-squared, omega squared, and epsilon squared. They reported that although partial eta-squared measure is the most frequent one to report strength of association, it is resulted from the availability of this measure in computer programs such as SPSS (IBM SPSS, 2011). In addition to widely used strength measures, there are also other suggestions to compute power in literature specifically for repeated measures, which are not get attention from researchers (D'Amico, Neilands, & Zambarano, 2001). Bakeman (2005) suggests the use of general eta squared measure to report effect size instead of partial eta squared which is available in SPSS for repeated measures designs, in cases when between comparisons needed for the dependent variables. But, in current study there will be no multiple DV's in analyses, in each analyses only one DV i.e. eye tracking measure in current study will be used. So this difference will be lost its meaning for this study. Consequently there is no problem in using partial eta square measure for effect size, and observed power for estimating power in provided by SPSS for current study methodically, but the problem can be arise in interpretation as explained below.

The problem about use and reporting about effect size and power is generally based on the priori or posteriori use of them. The power analysis can be used to estimate the number of the participant joined into study before conducting experiments, but it is used generally to discuss the size of sample posteriori. Researchers should be aware of the relation between significance levels, size of the samples and their relation to effect sizes. Aiming a high level of effect size or power may not be feasible always. The researcher should use the general preferences to achieve an optimum level of significance and sample size level. For example Tabachnick and Fidell (1996) suggest at least 5 subjects for each cell in repeated measures designs minimum, but they revised their suggestion to 10 for each cell to ensure robustness in 2007 (Tabachnick & Fidell, 2007).

Murphy and Myors (2004) explains relationship between number of the participants and the generability of the results. They stated that in some studies the use of repeated measure design could be advantageous if not necessity such as sleep or vision research. Repeated measures give the opportunity to collect huge physiological data for a reasonable size of sample without requirement of between measures. Additionally the data collected in repeated measures has more strength than the same number of observations collected by between group designs due to decreased error term if dependent variables correlated. For example, a design for four treatment and 57 subject led 228 observations in repeated measure. The number of observations in between group design reached 356 for same power=.80. But, Murphy and Myors added (2004) that strengths of the repeated measures can be a potential weakness if they are not implemented appropriately into study, so the number of the participants and the observations from them should be comparable to reach interpretable results if not only significant.

As the light of discussion made above, the number of participants of current study satisfy the sample size, which is greater than 10 for groups. Although the number of measures obtained by eye tracking devices is quite high, only one of them treated as DV, mean of fixation count per second in the case above in analysis. So the partial eta squared and observed power values provided by SPSS used for only significant results in the reports, but these are not used as tools to criticize the sample size of the study or the significant levels of the analyses to find any justification or excuse, rather they are used to provide cues for current study to find a place in multimedia research in general. Because, the effect sizes of the multimedia researches in the literature used as the key data to establish principles in Cognitive Theory of Multimedia Learning by Mayer (2010)

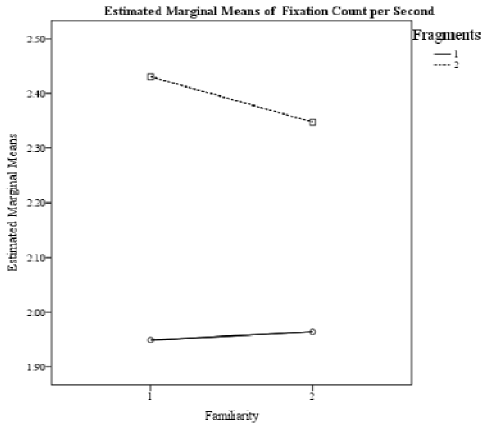


Figure 11: Means of Fixation Count per Second in Chemistry Lesson Familiarity by Fragments in Experiment 1

Figure 11 depicts the observed values for means of fixation counts for familiarity (1 for first study, 2 for second study) and fragments (solid line for experiment, dashed line for materials). It is shown that the mean of fixation counts in for both study is higher for materials part when compared to experiment part. The mean of count increased in second study for experiment part and decreased in materials part. This shows an interaction but this interaction does not significant when compared to main effect of fragments that is underlies the difference between mean of fixation counts in materials part and experiment part.

The results showed that the fixation count is effected from the content of material. If the content has list the number of counts increased, when compared to a fragment in which a process explained. Up to this point only fixation counts discussed, other eye tracking metrics of Experiment 1 will be reported in following sections. That are fixation duration and total fixation duration.

4.3.2. Results for Fixation Duration in Experiment 1

The descriptive statistics for the mean of the fixation durations in Experiment 1 is given in Table 21. In this analysis per second transformation was not applied because of mean of fixation durations have free from the length of the fragments. The results showed that the mean of duration are larger in experiment fragment (*M*=.5222, *SD*=.1496 for first study and *M*=.4993, *SD*=.0996 for second study) when compared to materials part (*M*=.4165, *SD*=.1038 for first study and *M*=.4240, *SD*=.1227 for second study).

Table 22 showed that the main effect of fragments is significant, $F(1,20)=20.385$, $p=.000$, $\eta=.505$ for mean of fixation duration per second as in fixation count. Familiarity also does not produce a significant main effect, $F(1,20)=.096$, $p=.760$ for fixation duration like in fixation count. There is no interaction observed for fixation duration in Experiment 1 ($F(1,20)=1.288$, $p=.253$).

Table 21: Descriptive Statistics for Means of Fixation Duration Experiment 1

Descriptive Statistics				
Fragments	Familiarity	Mean	Std. Deviation	N
Experiment	First Study	.5222	.1496	21
	Second Study	.4993	.0996	21
Materials	First Study	.4165	.1038	21
	Second Study	.4240	.1227	21

Table 22: Repeated Measures Design Results of Mean of Fixation Duration in Experiment 1

Tests of Within-Subjects Contrasts							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Fragments	.172	1	.172	20.385	.000	.505	.990
Error(Fragments)	.169	20	.008				
Familiarity	.001	1	.001	.096	.760	.005	.060
Error(Familiarity)	.262	20	.013				
Fragments * Familiarity	.005	1	.005	1.388	.253	.065	.202
Error(Fragments*Familiarity)	.070	20	.004				

a. Computed using alpha = .05

Figure 12 showed that the mean of fixation durations are slightly higher in materials part when compared to experiment part. And this pattern is observed also in second reading, whereas the gap decreased. The next part investigate another eye fixation metric that is total fixation duration. For total fixation duration the values will be divided into durations of fragments unlike fixation duration metric, since total duration is affected from the durations of fragments those durations are different for experiment and materials part directly.

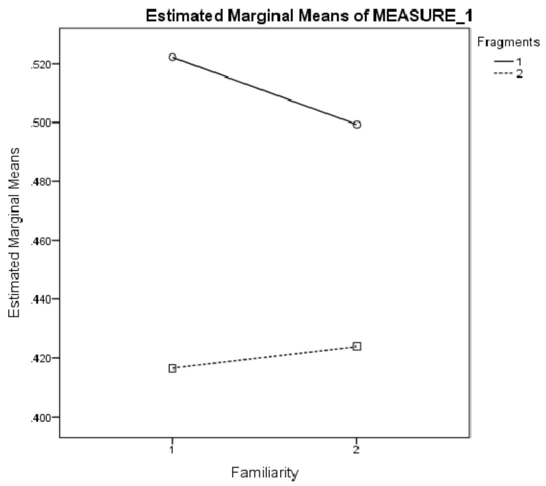


Figure 12: Means of Fixation Duration per Second in Chemistry Lesson Familiarity by Fragments in Experiment 1

4.3.3. Results for Total Fixation Duration per Second in Experiment 1

The descriptive statistics for the mean of the total fixation durations per second in Experiment 1 are given in Table 23. The results showed that the mean of duration are greater in first study both in experiment part (*M*=.9618, *SD*=.0321) and materials part (*M*=.9692, *SD*=.0456) when compared to second study (*M*=.9467, *SD*=.0455 for experiment part and *M*=.9454, *SD*=.0420 for materials part.

Table 24showed that the main effect of fragments is not significant, *F*(1,20)=.162, p=.691 for mean of total fixation duration per second unlike in fixation count and duration. Interestingly, Familiarity produces a significant main effect, *F(*1,20)=16.6, p=.001,η=.454 for total fixation duration unlike previous ones. There is no interaction observed for total fixation duration in Experiment 1 (*F*(1,20)=1.216, p=.283).

Table 23: Descriptive Statistics for Means of Total Fixation Duration per Second in Experiment 1

Descriptive Statistics				
Fragments	Familiarity	Mean	Std. Deviation	N
Experiment	First Study	.9618	.0321	21
	Second Study	.9467	.0455	21
Materials	First Study	.9692	.0456	21
	Second Study	.9454	.0420	21

Table 24: Repeated Measures Design Results of Mean of Total Fixation Duration per Second in Experiment 1

Tests of Within-Subjects Contrasts							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Fragments	.000	1	.000	.162	.691	.008	.067
Error(Fragments)	.024	20	.001				
Familiarity	.008	1	.008	16.600	.001	.454	.972
Error(Familiarity)	.010	20	.000				
Fragments * Familiarity	.000	1	.000	1.216	.283	.057	.183
Error(Fragments*Familiarity)	.007	20	.000				

a. Computed using alpha = .05

Figure 13 showed that the mean of total fixation durations are quite higher in first study of chemistry lesson when compared to second study both fragments. This pattern is not compatible with the patterns observed in Figure 11 and Figure 12 for fixation count and duration respectively. Although the total fixation duration is a product of fixation count and duration, the times spend outside of the lesson or saccades in which no recording made change the main effect familiarity from fragments.

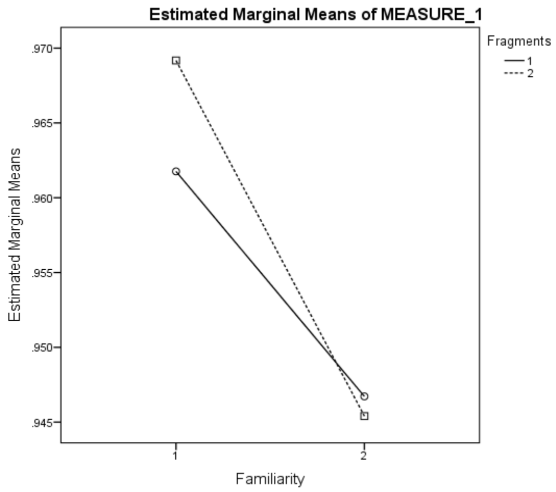


Figure 13: Means of Total Fixation Duration per Second in Chemistry Lesson Familiarity by Fragments in Experiment 1

To sum up, the main effects of familiarity and fragments investigated in Experiment 1. There was only one lesson in Experiment 1, which is chemistry lesson with narration. Participants study the chemistry lesson constituted experiment and materials parts (fragments) twice (familiarity effect). The fixation count and total fixation durations divided the durations of fragments, but fixation duration used as they are. The results showed that fragments has effect on fixation duration and count of participants; however total fixation duration affected from familiarity. No interaction observed for all eye tracking metrics in Experiment 1.

4.4. Results of Experiment 2

Experiment 2 includes two different version of chemistry lesson. These are chemistry with subtitles (CWS) and chemistry with no voice (CNV) lessons. The only difference between CWS and CNV lessons is the narration at background. There are two new intrinsic variable in Experiment 2 in addition to familiarity and fragments variables. These are existence of narration and areas of interest. To observe the effect of narration CWS and CNV lessons studied by different groups. And areas of interest in chemistry lesson contains video and subtitles areas. Therefore, the design of the Experiment 2 is different from Experiment 1 which contains only within group comparisons. Experiment 2, on the other hand will include between group comparisons of lessons which makes the design a mixed one.

In addition to fixation count, fixation duration, and total fixation duration in Experiment 1, visit count, visit duration, and total visit duration metrics in can be used in Experiment 2, due to inclusion of the subtitles and video to the lessons as AOI. But instead of using all six eye tracking metrics, only fixation count, fixation duration and total visit durations will be reported. Total visit duration is quite alike total fixation count, both include the time spend on a specific area, but their calculations are different. The visit count and visit metrics will not be reported, since there are only two areas of interest. Those metrics are meaningful if three are three or more AOI defined on screen.

Before conducting the experiments, the outputs of Tobii Studio© transformed to reach more reliable results for fragments and AOI. The fixation counts weighted according to the duration of the fragments and area proportions of the AOI. The process of how fixation counts weighted according to fragments explained in previous sections. Since the areas of subtitles and video are not equal, the mean of fixation counts weighted according to screen proportions of AOI. Subtitles covers 23% of screen and video covers 77% of the screen. The fixation counts on subtitles and video areas multiplied by .77 and .23 respectively to neutralize the sizes of AOI's in following analyses.

The last point for Experiment 2 is the number of participants. There are 31 participants in Experiment 2 from Atılım U. and METU in Experiment 2, but there are no participant to CNV lesson from Atılım U. because of technical reasons. For that reason, to omit the variance that might be resulted from university, 12 participants from Atılım U. omitted from repeated measures design. The analyses conducted only with participants from METU.

Table 25 shows the structure of the experimet2 in which there are three within and one between subject factors. Each subject factor has two levels. The name and number of the levels in addition to dependent variables reported in following three sections are given in Table 25. Dependent variables reported in separate titles to increase the readability of the results. At the end of the section for Experiment 2, as short summary will be given for all subtests.

Table 25: Within and Between Subject Factors of Experiment 2

Familiarity/Fragments/AOI			**Dependent Variable**
Within Subject Factors			
First Study (1)	Experiment(1)	Subtitles(1)	
		Video(2)	
	Materials(2)	Subtitles(1)	Fixation Count
		Video(2)	Fixation Duration
Second Study(2)	Experiment (1)	Subtitles(1)	Total Visit Duration
		Video(2)	
	Materials(2)	Subtitles(1)	
		Video(2)	
Between Subject Factors			
Lesson	CNV		
	CWS		

4.4.1. Results for Fixation Count per Second in Experiment 2

The descriptive statistics for the mean of the fixation counts in Experiment 2 are given in Table 26. The results showed that the transformed means for fixation counts on video area are greater when compared to subtitles area first. Nevertheless, to reach any conclusions, the ANOVA tables of within and between subject factors for repeated measures are given in

Table 27. The main effect of the familiarity is not significant, $F(1,17)=1.593$, p=.224, but main effects of fragments and AOI are significant, $F(1,17)=8.716$, p=.009, partial η=.339 and $F(1,17)=9482.936$, p=.000, partial η=.861 respectively for mean of fixation count. The between subject factor of lesson is also significant as main effect, $F(1,17)=20.106$, p=.000, partial η=.861 for fixation count. Four interactions are significant out of eleven possibilities. These interactions are fragments*lesson interaction, $F(1,17)=34.603$, p=.000, partial η=.671; AOI*lesson interaction $F(1,17)=33.536$, p=.000, partial η=.664; fragments*AOI interaction, $F(1,17)=4.906$, p=.041, partial η=.224; and fragments*AOI*lesson interaction, $F(1,17)=24.697$, p=.000, partial η=.592.

Table 26: Descriptive Statistics for Means of Fixation Counts in Experiment 2

Familiarity/Fragment/AOI/Lesson				**Mean**	**Std.Deviation**	**N**
First Study (1)	Experiment (1)	Subtitle (1)	CNV	30.06	5.54	10
			CWS	4.13	2.80	9
			Total	17.78	13.99	19
		Video (2)	CNV	27.81	6.22	10
			CWS	36.35	3.70	9
			Total	31.86	6.68	19
	Materials (2)	Subtitle (1)	CNV	19.10	3.70	10
			CWS	13.44	3.05	9
			Total	16.42	4.40	19
		Video (2)	CNV	35.30	7.65	10
			CWS	36.67	4.56	9
			Total	35.95	6.24	19
Second Study (2)	Experiment (1)	Subtitle (1)	CNV	28.09	9.08	10
			CWS	7.78	5.41	9
			Total	18.47	12.76	19
		Video (2)	CNV	26.69	6.64	10
			CWS	32.78	4.81	9
			Total	29.57	6.49	19
	Materials (2)	Subtitle (1)	CNV	18.40	3.84	10
			CWS	14.22	6.20	9
			Total	16.42	5.39	19
		Video (2)	CNV	33.90	8.44	10
			CWS	34.67	8.15	9
			Total	34.26	8.08	19

Table 27: Repeated Measures Design Results for Mean of Fixation Counts in Experiment 2

Tests of Within-Subjects Contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Familiarity	23.849	1	23.849	1.593	.224	.086	.222
Familiarity * Lesson	9.669	1	9.669	.646	.433	.037	.118
Fragments	85.293	1	85.293	8.716	.009	.339	.795
Fragments * Lesson	338.629	1	338.629	34.603	.000	.671	1.000
AOI	9842.936	1	9842.936	105.058	.000	.861	1.000
AOI * Lesson	3141.956	1	3141.956	33.536	.000	.664	1.000
Familiarity * Fragments	.053	1	.053	.007	.934	.000	.051
Familiarity * Fragments * Lesson	3.093	1	3.093	.417	.527	.024	.094
Familiarity * AOI	57.520	1	57.520	2.777	.114	.140	.349
Familiarity * AOI * Lesson	60.925	1	60.925	2.942	.104	.148	.367
Fragments * AOI	281.410	1	281.410	4.906	.041	.224	.551
Fragments * AOI * Lesson	1416.532	1	1416.532	24.697	.000	.592	.997
Familiarity * Fragments * AOI	4.982	1	4.982	.138	.715	.008	.064
Familiarity * Fragments * AOI * Lesson	21.232	1	21.232	.586	.454	.033	.112

a. Computed using alpha = .05

Table 27: Repeated Measures Design Results for Mean of Fixation Counts in Experiment 2 (Cont.)

Tests of Between-Subjects Effects							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Intercept	94451.859	1	94451.859	2075.103	.000	.992	1.000
Lesson	915.179	1	915.179	20.106	.000	.542	.988
Error	773.784	17	45.517				

a. Computed using alpha = .05

The figures of significant interactions are given in following tables. Figure 14 showed that the mean of fixation counts in experiment part (1) and materials part (2) quite different for CNV and CWS lessons. Fixation count is smaller for experiment part in CWS lesson than materials part, but this phenomena is opposite for CNV lesson. The number of counts per second are higher in experiment part when there was no narration at background. The difference between in fixation count between lessons can be attributable to the existence of narration in lessons. That is if narration exist at background learners fixated less for both fragments, but the decrease observed in when a process explained in content. If narration removed from environment, the number of fixation increased to follow subtitles. This variance can be observable at Figure 15, in which the number of fixations in subtitles area (1) is minimum for CWS lesson when compared to CNV lesson. But, the number of fixations does not differ for video area (2) for both lesson between lessons.

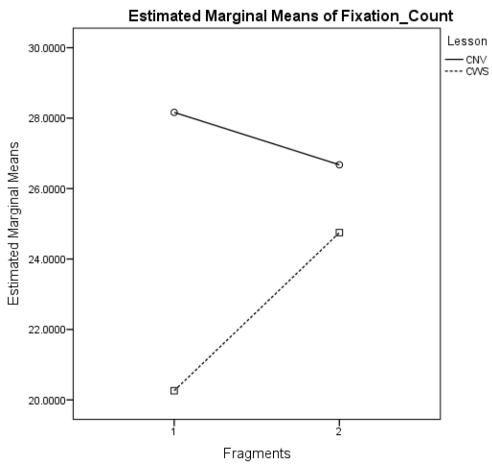


Figure 14: Plot of Interaction Fragments by Lesson in Experiment 2

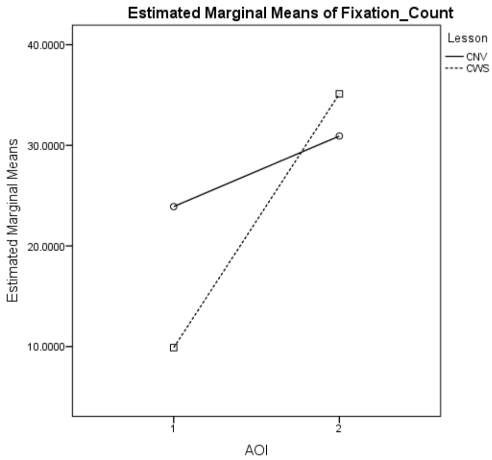


Figure 15: Plot of Interaction AOI by Lesson in Experiment 2

Figure 16 depicts an interaction between fragments and AOI also. The number of fixations on video area is higher for video area than subtitles area. That means, learner prefer looking at video area rather than reading subtitles. Moreover, the difference increased in materials part that might be resulted from the lengths of subtitles that are shorter than the experiment part. Figure 17, on the other hand, show a three way interaction by fragment, AOI and lessons. The difference between CNV and CWS lessons resulted from narration, which is direct effect on the part on subtitles at experiment part. The fixation counts on materials part does not seem to be affected by the presence of narration as in experiment part. So it can be said that, if the content has process rather than list, learners' attention on word affected most by the narration at background.

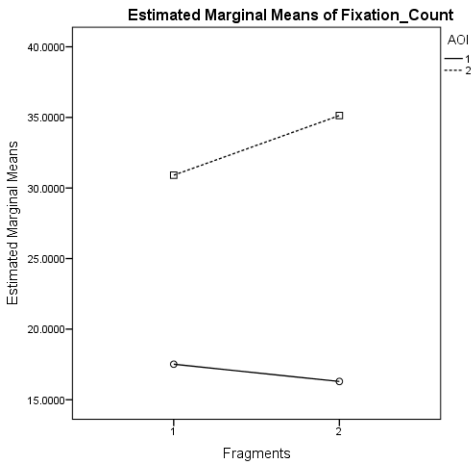


Figure 16: Plot of Interaction Fragments by AOI in Experiment 2

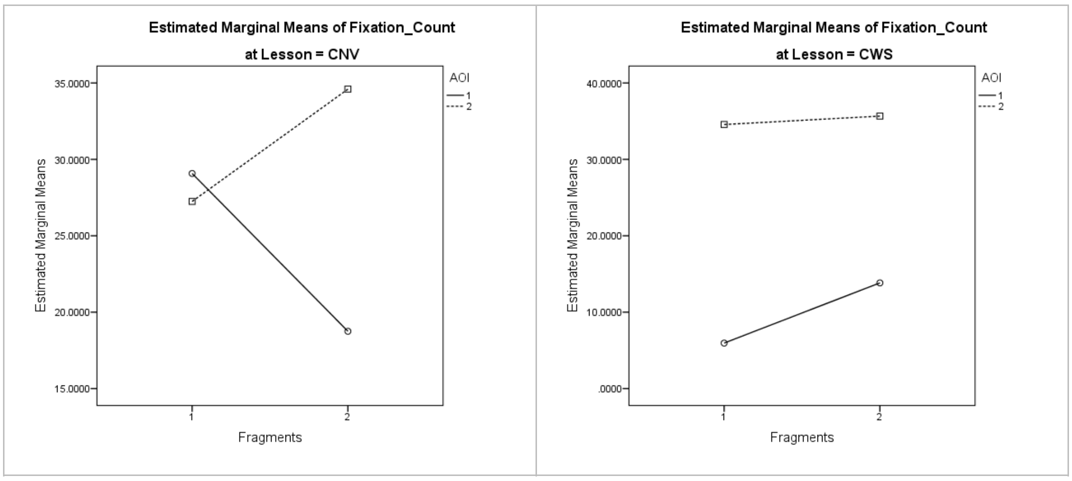


Figure 17: Separate Plots of Interaction Fragments by Lesson and AOI in Experiment 2

4.4.2. Results for Fixation Duration in Experiment 2

The descriptive statistics for the mean of the fixation counts in Experiment 2 is given in Table 28 The results showed that the transformed means for counts on materials fragment are greater when compared to experiments fragment at first sight, the ANOVA tables of within and between subject factors for repeated measures are given in Table 29 to reach any conclusion. The main effect of the familiarity is not significant, $F(1,17)=.787$, $p=.387$, but main effects of fragments, and AOI are significant, $F(1,17)=1186.750$, $p=.000$, partial $\eta=.986$ and $F(1,17)=23.514$, $p=.000$, partial $\eta=.580$ respectively for mean of fixation duration as fixation count in previous section. But none of the interactions is not significant for fixation duration in experiment to. The between subject factor of lesson is not significant as main effect also, $F(1,17)=1.780$, $p=.200$ for fixation duration. Since there is no interaction effect except fragments and lesson main effects, there will be no plots in this section.

Table 28: Descriptive Statistics for Means of Fixation Durations in Experiment 2

Familiarity/Fragment/AOI/Lesson				**Mean**	**Std.Deviation**	**N**
First Study (1)	Experiment (1)	Subtitle (1)	CNV	0.074	0.005	10
			CWS	0.081	0.018	9
			Total	0.077	0.013	19
		Video (2)	CNV	0.140	0.020	10
			CWS	0.158	0.017	9
			Total	0.148	0.020	19
	Materials (2)	Subtitle (1)	CNV	0.323	0.056	10
			CWS	0.323	0.089	9
			Total	0.323	0.071	19
		Video (2)	CNV	0.398	0.059	10
			CWS	0.412	0.076	9
			Total	0.405	0.066	19
Second Study (2)	Experiment (1)	Subtitle (1)	CNV	0.077	0.006	10
			CWS	0.093	0.049	9
			Total	0.084	0.034	19
		Video (2)	CNV	0.143	0.035	10
			CWS	0.165	0.028	9
			Total	0.154	0.033	19
	Materials (2)	Subtitle (1)	CNV	0.329	0.109	10
			CWS	0.332	0.069	9
			Total	0.330	0.090	19
		Video (2)	CNV	0.387	0.084	10
			CWS	0.428	0.089	9
			Total	0.406	0.086	19

Table 29: Repeated Measures Design Results for Mean of Fixation Durations in Experiment 2

Tests of Within-Subjects Contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Familiarity	.001	1	.001	.787	.387	.044	.133
Familiarity * Lesson	.001	1	.001	.710	.411	.040	.125
Fragments	2.372	1	2.372	1186.750	.000	.986	1.000
Fragments * Lesson	1.195E-05	1	1.195E-05	.006	.939	.000	.051
AOI	.213	1	.213	23.514	.000	.580	.995
AOI * Lesson	.003	1	.003	.297	.593	.017	.081
Familiarity * Fragments	2.002E-05	1	2.002E-05	.012	.913	.001	.051
Familiarity * Fragments * Lesson	.000	1	.000	.085	.774	.005	.059
Familiarity * AOI	.000	1	.000	.059	.811	.003	.056
Familiarity * AOI * Lesson	.000	1	.000	.102	.753	.006	.060
Fragments * AOI	.001	1	.001	.152	.701	.009	.066
Fragments * AOI * Lesson	.001	1	.001	.144	.709	.008	.065
Familiarity * Fragments * AOI	1.844E-05	1	1.844E-05	.007	.932	.000	.051
Familiarity * Fragments * AOI * Lesson	.001	1	.001	.230	.637	.013	.074

a. Computed using alpha = .05

Table 34: Repeated Measures Design Results for Mean of Fixation Durations in Experiment 2 (Cont.)

Tests of Between-Subjects Effects							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Intercept	8.834	1	8.834	1835.462	.000	.991	1.000
Lesson	.009	1	.009	1.780	.200	.095	.242
Error	.082	17	.005				

a. Computed using alpha = .05

4.4.3. Results for Total Visit Duration in Experiment 2

In this section, a new metric will be used as dependent variables for analyses. Total visit duration chosen as eye tracking metric to investigate the variance for all the time if spend on areas of screen taken into account. But the total visit duration is spend on have not been used in analyses as they are. They proportioned by considering sizes of AOI, and weighted in terms of durations for fragments. As a result of these transformations, descriptives should not be accounted as seconds rather they are constructed values to reach more trusted analyses. The descriptive statistics for the total visit durations in Experiment 2 is given in Table 30. The repeated measures ANOVA tables are given in Table 31. The main effects of the familiarity and AOI are not significant, $F(1,17)=.582$, p=.456, and $F(1,17)=3.636$, p=.074 but main effect of fragments is significant, $F(1,17)=562.601$, p=000, partial η=.971 for transformed values of total visit duration. The between subject factor of lesson is significant as main effect, $F(1,17)=12.676$, p=.002, partial η=.427. Two interactions AOI*lesson, $F(1,17)=12.676$, p=.002, partial η=.427; and fragments*AOI $F(1,17)=5.089$, p=.038, partial η=.230 interactions are significant.

Table 30: Descriptive Statistics for Means of Fixation Counts in Experiment 2

Familiarity/Fragment/AOI/Lesson				Mean	Std.Deviation	N
First Study (1)	Experiment (1)	Subtitle (1)	CNV	0.088	0.019	10
			CWS	0.014	0.010	9
			Total	0.053	0.041	19
		Video (2)	CNV	0.045	0.006	10
			CWS	0.068	0.003	9
			Total	0.056	0.012	19
	Materials (2)	Subtitle (1)	CNV	0.239	0.064	10
			CWS	0.175	0.070	9
			Total	0.208	0.073	19
		Video (2)	CNV	0.159	0.019	10
			CWS	0.178	0.021	9
			Total	0.168	0.022	19
Second Study (2)	Experiment (1)	Subtitle (1)	CNV	0.088	0.031	10
			CWS	0.025	0.017	9
			Total	0.058	0.040	19
		Video (2)	CNV	0.046	0.009	10
			CWS	0.064	0.005	9
			Total	0.054	0.012	19
	Materials (2)	Subtitle (1)	CNV	0.242	0.093	10
			CWS	0.193	0.108	9
			Total	0.219	0.101	19
		Video (2)	CNV	0.158	0.028	10
			CWS	0.172	0.032	9
			Total	0.165	0.030	19

Table 31: Repeated Measures Design Results for Total Visit Duration in Experiment 2

Tests of Within-Subjects Contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Familiarity	.000	1	.000	.582	.456	.033	.111
Familiarity * Lesson	.000	1	.000	.359	.557	.021	.087
Fragments	.686	1	.686	562.601	.000	.971	1.000
Fragments * Lesson	.000	1	.000	.141	.712	.008	.065
AOI	.018	1	.018	3.636	.074	.176	.436
AOI * Lesson	.063	1	.063	12.676	.002	.427	.918
Familiarity * Fragments	3.159E-05	1	3.159E-05	.054	.819	.003	.056
Familiarity * Fragments * Lesson	1.205E-06	1	1.205E-06	.002	.964	.000	.050
Familiarity * AOI	.001	1	.001	.582	.456	.033	.111
Familiarity * AOI * Lesson	.001	1	.001	.359	.557	.021	.087
Fragments * AOI	.021	1	.021	5.089	.038	.230	.567
Fragments * AOI * Lesson	.001	1	.001	.141	.712	.008	.065
Familiarity * Fragments * AOI	.000	1	.000	.054	.819	.003	.056
Familiarity * Fragments * AOI * Lesson	4.132E-06	1	4.132E-06	.002	.964	.000	.050

a. Computed using alpha = .05

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Intercept	2.260	1	2.260	1561.860	.000	.989	1.000
Lesson	.018	1	.018	12.676	.002	.427	.918
Error	.025	17	.001				

a. Computed using alpha = .05

The figures of significant interactions are given in following tables. Figure 18showed that the subtitles in CNV lesson visited most, but video areas visited least. For CWS lesson on the other hand, subtitles area gathers less attention when compared to video area. The reason for this discrepancy might be narration as in fixation count. Figure 19showed that the in the experiment part proportion of total visit duration almost same for video and subtitles area, but there is difference for same value in materials fragment. The proportion of total visit duration is greater for subtitles area in materials fragment that might be resulted from the lengths of the subtitles.

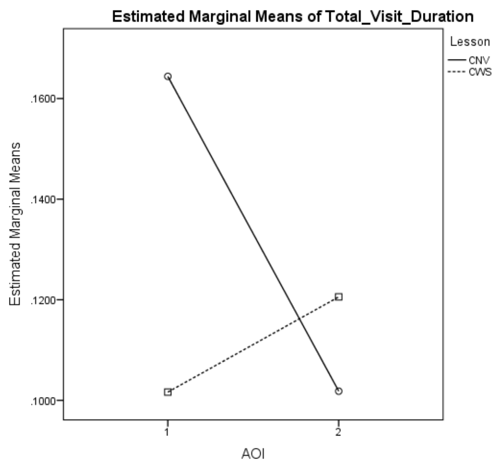


Figure 18: Plot of Interaction AOI by Lesson in Experiment 2

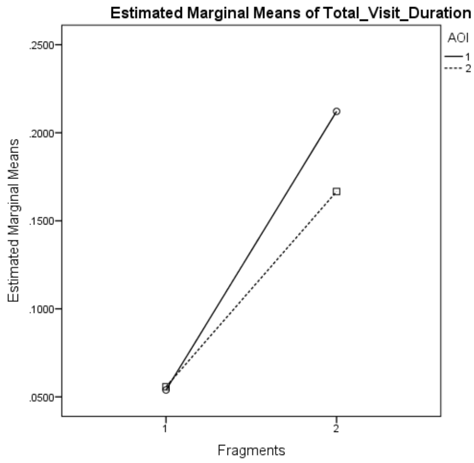


Figure 19: Plot of Interaction AOI by Fragments in Experiment 2

4.5. Results of Experiment 3

The design of Experiment 3 is identical to Experiment 2 except two changes. One is the lessons changed into mechanism from chemistry. The differences between lessons that creates between subjects variance in repeated measures is the pace instead of narration as in Experiment 2. Moreover, the second change is the fragments converted into type of images in experiment two. The within content variation is created by materials and experiment part in chemistry lesson, converted into animated and static images on pages. The text and images constitute AOI in mechanism lesson as in chemistry lesson but in different proportions. No video and narration is used in mechanism lesson. The familiarity created by first and second study stayed same in Experiment 3. The variables in Experiment 3 to test system paced mechanism (M) and self paced mechanism (MSP) lesson are given in Table 32.

Table 32: Within and Between Subject Factors of Experiment 3

Familiarity/Types of Images/AOI			**Dependent Variable**
Within Subject Factors			
First Study (1)	Animated (1)	Images (1)	Fixation Count Fixation Duration Total Visit Duration
		Text (2)	
	Static (2)	Images (1)	
		Text (2)	
Second Study(2)	Animated (1)	Images (1)	
		Text (2)	
	Static (2)	Images (1)	
		Text (2)	
Between Subject Factors			
Lesson	M MSP		

Transformations in eye tracking measures for the fixation count and total visit duration is made before analyses in Experiment 3. The area of images covers 37% of the screen and text covers 63% of the screen as AOI. The durations for animated and static images are different also. Although the durations of animated and static images constant for all participants, these values differ in MSP lesson for all participants. Since, the durations of MSP lesson changed for all participants, the individual durations used as durations in MSP lesson. The mean of fixation count and total visit durations weighted for AOI and divided the durations of types of images to achieve standard scores.

4.5.1. Results for Fixation Count per Minute in Experiment 3

The descriptive statistics for the mean of the fixation counts in Experiment 3 is given in Table 33. The results showed that the means for fixation counts on text area are greater than images although they are weighed for AOI. Table 34 presents the repeated measures ANOVA results. All main effects in within factor design are significant. These are familiarity, $F(1,41)=35.465$, p=.000, partial η=.464, type of image, $F(1,41)=30.502$, p=.000, partial η=.427, and AOI $F(1,41)=338.172$, p=.000, partial η=.892. However, between subject factor of lesson is not significant as main effect, $F(1,41)=2.689$, p=.109 for fixation count. There are five significant interactions in model. These interactions are AOI*lesson interaction, $F(1,41)=4.246$, p=.046, partial η=.094; familiarity*TypeOfImage interaction $F(1,41)=45.490$, p=.000, partial η=.526; familiarity*AOI interaction, $F(1,41)=49.312$, p=.000, partial η=.546; TypeOfImage*AOI interaction, $F(1,41)=76.662$, p=.000, partial η=.652, and Familiarity*TypeOfImage*AOI interaction, $F(1,41)=42.239$, p=.000, partial η=.507.

Table 33: Descriptive Statistics for Means of Fixation Counts in Experiment 3

Familiarity/Type of Image/AOI/Lesson				Mean	Std.Deviation	N
First Study (1)	Animated (1)	Image (1)	M	7.38	3.66	21
			MSP	11.50	7.22	22
			Total	9.49	6.06	43
		Text (2)	M	109.94	10.32	21
			MSP	141.70	80.21	22
			Total	126.19	59.38	43
	Static (2)	Image (1)	M	9.64	3.80	21
			MSP	10.13	6.95	22
			Total	9.89	5.57	43
		Text (2)	M	62.31	7.70	21
			MSP	79.59	43.99	22
			Total	71.15	32.75	43
Second Study (2)	Animated (1)	Image (1)	M	8.91	4.63	21
			MSP	6.98	4.10	22
			Total	7.92	4.42	43
		Text (2)	M	60.16	10.92	21
			MSP	69.16	35.91	22
			Total	64.76	26.87	43
	Static (2)	Image (1)	M	14.83	7.02	21
			MSP	11.45	8.43	22
			Total	13.10	7.87	43
		Text (2)	M	61.52	9.57	21
			MSP	66.69	33.81	22
			Total	64.16	24.94	43

Table 34: Repeated Measures Design Results for Mean of Fixation Counts in Experiment 3

Tests of Within-Subjects Contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Familiarity	23578.595	1	23578.595	35.465	.000	.464	1.000
Familiarity * lesson	2693.430	1	2693.430	4.051	.051	.090	.502
TypeOfImage	13313.738	1	13313.738	30.502	.000	.427	1.000
TypeOfImage * lesson	734.495	1	734.495	1.683	.202	.039	.245
AOI	436678.484	1	436678.484	338.172	.000	.892	1.000
AOI * lesson	5482.331	1	5482.331	4.246	.046	.094	.521
Familiarity * TypeOfImage	18741.310	1	18741.310	45.490	.000	.526	1.000
Familiarity * TypeOfImage * lesson	220.638	1	220.638	.536	.468	.013	.110
Familiarity * AOI	26143.868	1	26143.868	49.312	.000	.546	1.000
Familiarity * AOI * lesson	837.214	1	837.214	1.579	.216	.037	.233
TypeOfImage * AOI	20027.740	1	20027.740	76.662	.000	.652	1.000
TypeOfImage * AOI * lesson	235.151	1	235.151	.900	.348	.021	.153
Familiarity * TypeOfImage * AOI	13197.707	1	13197.707	42.239	.000	.507	1.000
Familiarity * TypeOfImage * AOI * lesson	95.914	1	95.914	.307	.583	.007	.084

a. Computed using alpha = .05

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Intercept	719404.054	1	719404.054	368.460	.000	.900	1.000
Lesson	5250.980	1	5250.980	2.689	.109	.062	.360
Error	80050.845	41	1952.460				

a. Computed using alpha = .05

The figures of significant interactions are given in following tables. Figure 20 showed that mean fixation counts on images (1) does not change according to lesson, but it changes on text (2) that is higher for MSP lesson. Figure 21 showed that there are differences on animated and static images at first and second reading. Although animated images gather attention at first reading, they completely loose attention at second reading. The relation between AOI and familiarity is given in Figure 22. The fixation count on text is smaller than the count on images in both reading and stay stable in both readings, but there is a decrease for fixation count on images in second reading. The same pattern observed for type of images also as in AOI (Figure 23). The last interaction in Experiment 3 is a triple one. The interaction between familiarity, type of image, and AOI is depicted by two separate plots in Figure 24. The mean of fixation counts on images does not change according to type but the fixation counts of text changed according to the images net to them. If animated images accompany to the text, the number of fixation count getting higher in first reading .But in second reading the fixation count in text does not changed whether animated or static images placed next to it.

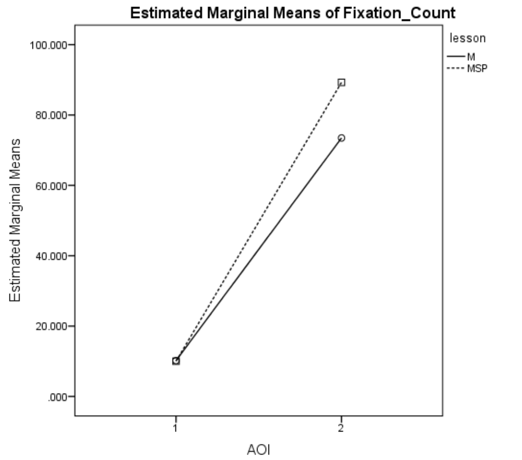


Figure 20: Plot of Interaction AOI by Lesson in Experiment 3

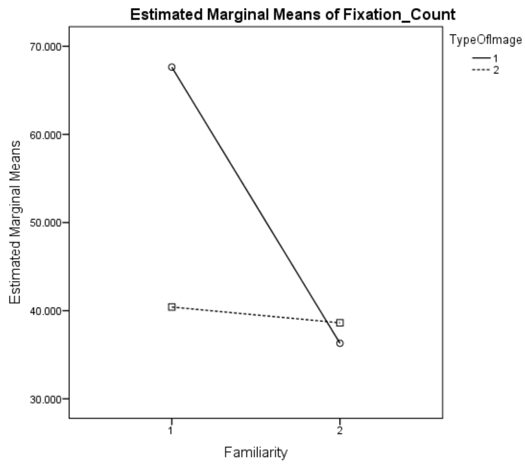


Figure 21: Plot of Interaction Familiarity by TypeOfImage in Experiment 3

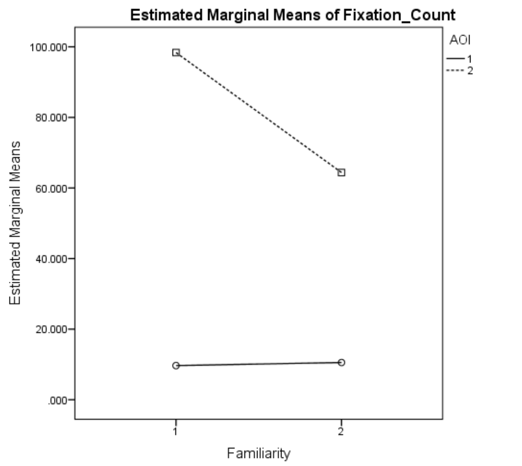


Figure 22: Plot of Interaction Familiarity by AOI in Experiment 3

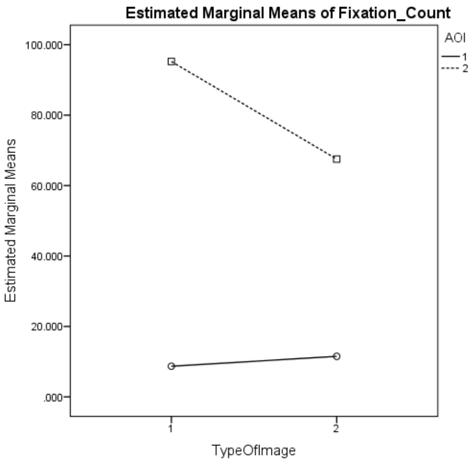


Figure 23: Plot of Interaction TypeOfImage by AOI in Experiment 3

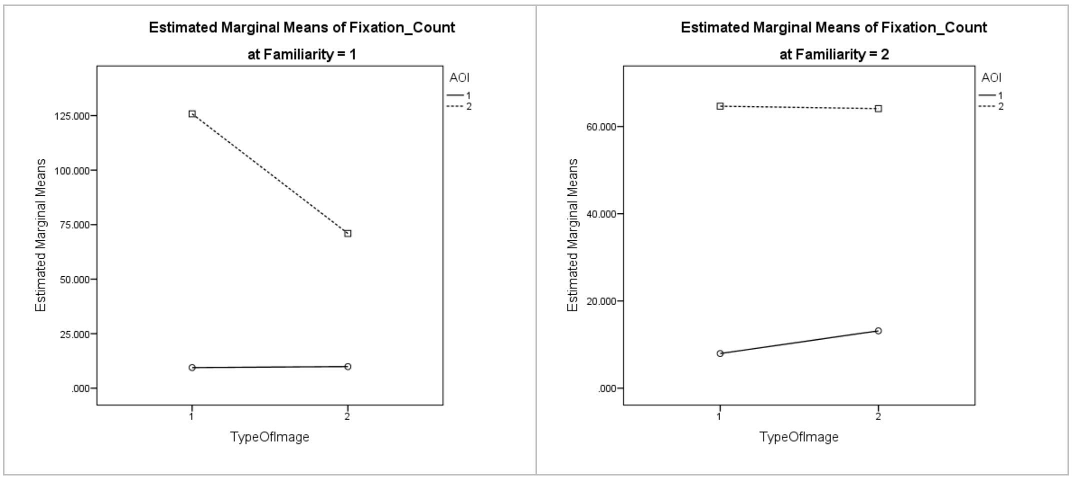


Figure 24: Separate Plots of Interaction Familiarity by TypeOfImage and AOI in Experiment 3

In this section, the fixation count used as dependent variable in Experiment 3. It is found that, fixation count is affected from familiarity, type of images, and AOI as main effects. In addition to main effect, many interactions discussed in this section.

4.5.2. Results for Fixation Duration in Experiment 3

The descriptive statistics for the mean of the fixation durations in Experiment 3 were given in Table 35. The results showed that the means for fixation durations on images are greater than text. The significances investigated in ANOVA table (Table 36).Familiarity as within subject effect is not significant, $F(1,41)=.733$, p=.397. Type of image, $F(1,41)=.108.571$, p=.000, partial $\eta=.726$; AOI, $F(1,41)=95.153$, p=.000, partial $\eta=.699$; and lesson $F(1,41)=13.939$, p=.001, partial $\eta=.254$ as between subject factor produce significant results. There are seven significant interactions in the model. Only familiarity*TypeOfImage, familiarity*AOI, and familiarity*TypeOfImage*AOI interactions are not significant for fixation duration. The significant interactions are familiarity*lesson$F(1,41)=12.236$, p=.001, partial $\eta=.230$, TypeOfImage*lesson, $F(1,41)=28.009$, p=.000, partial $\eta=.406$, AOI*lesson, $F(1,41)=21.186$, p=.000, partial $\eta=.341$, familiarity*TypeOfImage*lesson, $F(1,41)=5.828$, p=.020, partial $\eta=.124$, familiarity*AOI*lesson, $F(1,41)=9.623$, p=.003, partial $\eta=.190$, TypeOfImage*AOI, $F(1,41)=99.838$, p=.000, partial $\eta=.709$, TypeOfImage*AOI*lesson,$F(1,41)=28.259$, p=.000, partial $\eta=.408$, familiarity*TypeOfImage*AOI*lesson, $F(1,41)=5.429$, p=.025, partial $\eta=.117$. The plots for two-way interactions are given below.

Table 35: Descriptive Statistics for Means of Fixation Durations in Experiment 3

Familiarity/Type of Image/AOI/Lesson				Mean	Std.Deviation	N
First Study (1)	Animated (1)	Image (1)	M	.6931	.2590	21
			MSP	.5207	.1479	22
			Total	.6049	.2247	43
		Text (2)	M	.2811	.0400	21
			MSP	.2984	.0410	22
			Total	.2900	.0410	43
	Static (2)	Image (1)	M	.3431	.0975	21
			MSP	.3355	.0588	22
			Total	.3392	.0792	43
		Text (2)	M	.2805	.0336	21
			MSP	.2921	.0331	22
			Total	.2864	.0335	43
Second Study (2)	Animated (1)	Image (1)	M	.8736	.4323	21
			MSP	.4027	.1872	22
			Total	.6327	.4040	43
		Text (2)	M	.2836	.0228	21
			MSP	.2879	.0361	22
			Total	.2858	.0301	43
	Static (2)	Image (1)	M	.3898	.0915	21
			MSP	.3246	.0663	22
			Total	.3564	.0853	43
		Text (2)	M	.2890	.0330	21
			MSP	.2869	.0363	22
			Total	.2880	.0343	43

Table 36: Repeated Measures Design Results for Mean of Fixation Durations in Experiment 3

Tests of Within-Subjects Contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Familiarity	.012	1	.012	.733	.397	.018	.133
Familiarity * lesson	.197	1	.197	12.236	.001	.230	.927
TypeOfImage	1.624	1	1.624	108.571	.000	.726	1.000
TypeOfImage * lesson	.419	1	.419	28.009	.000	.406	.999
AOI	3.368	1	3.368	95.153	.000	.699	1.000
AOI * lesson	.750	1	.750	21.186	.000	.341	.994
Familiarity * TypeOfImage	.000	1	.000	.024	.879	.001	.053
Familiarity * TypeOfImage * lesson	.077	1	.077	5.828	.020	.124	.654
Familiarity * AOI	.014	1	.014	.943	.337	.022	.158
Familiarity * AOI * lesson	.146	1	.146	9.623	.003	.190	.857
TypeOfImage * AOI	1.610	1	1.610	99.838	.000	.709	1.000
TypeOfImage * AOI * lesson	.456	1	.456	28.259	.000	.408	.999
Familiarity * TypeOfImage * AOI	.002	1	.002	.135	.715	.003	.065
Familiarity * TypeOfImage * AOI * lesson	.078	1	.078	5.429	.025	.117	.624

a. Computed using alpha = .05

Table 37: Repeated Measures Design Results for Mean of Fixation Durations in Experiment 3 (Cont.)

Tests of Between-Subjects Effects							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Intercept	51.337	1	51.337	1134.924	.000	.965	1.000
Lesson	.631	1	.631	13.939	.001	.254	.954
Error	1.855	41	.045				

a. Computed using alpha = .05

Since there are many interactions for fixation duration, only two-way interactions reported. The first interaction is between familiarity and lesson. That is the mean of fixation durations and lesson in that case control is interrelated with each other. The fixation counts in first reading is closer for lessons than second reading. The fixation duration increased in M lesson in second reading but same value decreases in second one if learner has control over pace (Figure 25). Figure 26 showed that fixation durations are varied according to lesson (MSP vs M) and type of images (animated vs. static). The difference of fixation duration is small for static images for different lessons, but the gap increase if the images are animated rather than static.

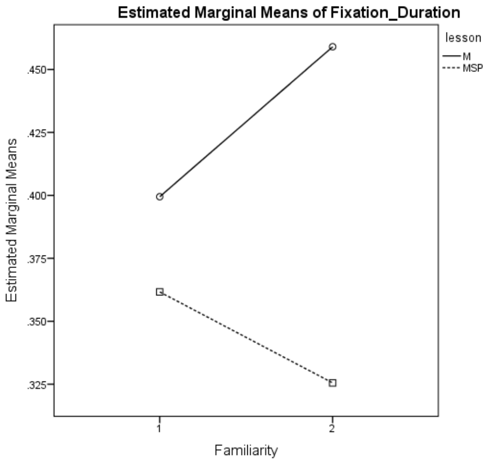


Figure 25: Plot of Interaction Familiarity by Lesson in Experiment 3

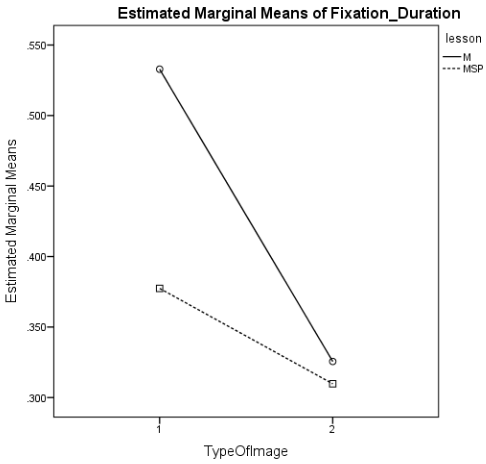


Figure 26: Plot of Interaction Lesson by Type of Image in Experiment 3.

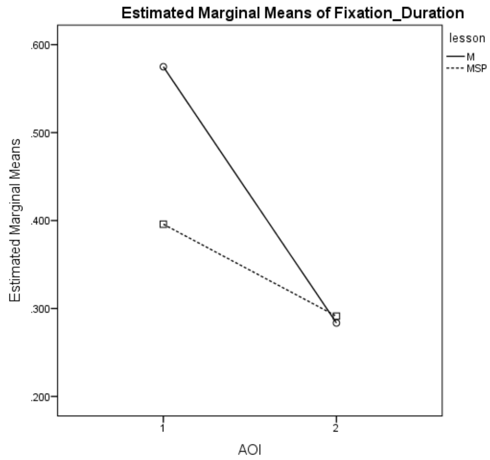


Figure 27: Plot of Interaction Lesson by AOI in Experiment 3.

Figure 27 points that difference between M and MSP lesson for fixation duration is lost when fixations occur on text. But on images the difference between MSP and M lesson can easily observable, in which MSP lesson has higher duration averages than M lesson. That might be interpreted as reading behaviors does not affected by having control on pace. Learner does not change their reading habit even they have the chance to do. The same pattern can be also observable when the text is presented by animated or static image. Although the fixation durations changed according to type of image, text stayed steady in both condition (Figure 28).

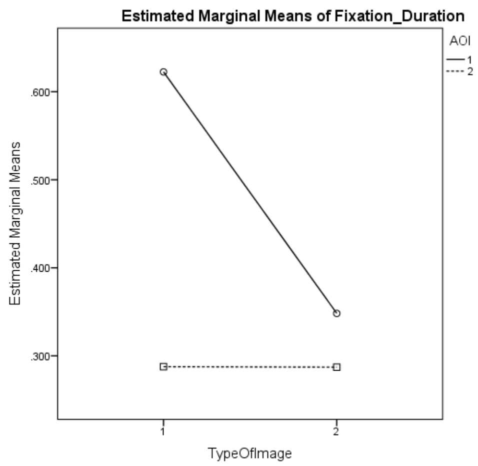


Figure 28: Plot of Interaction TypeOfImage by AOI in Experiment 3

In this section, the fixation duration used as dependent variable. It is found that, fixation count is affected from familiarity, type of images, and AOI as main effects. In addition to main effect, many interactions discussed in this section. The next section investigated total visit duration as dependent variable.

4.5.3. Results for Total Visit Duration in Experiment 3

The reasons for choosing total visit duration as eye tracking metric for comparing lessons are explained in Experiment 2 part. To remind that the values used in this section for total visit durations are not raw scores that are time spend on a specific area, rather they are corrected values by considering durations of different parts in lessons and the sizes of areas of interests. The descriptive statistics for the total visit durations in Experiment 3 is given in Table 38.Table 39 presents the repeated measures ANOVA statistics. The significant main effects are type of image, $F(1,41)=79.464$, p=.000, partial η=.660, and AOI, $F(1,41)=563.065$, p=.000, partial η=.932. The lesson is also a significant factor in variance, $F(1,41)=6.458$, p=.000, partial η=.136. There are four significant interaction for total visit duration. The first significant interaction is between TypeOfImage*lesson, $F(1,41)=4.466$, p=.000, partial η=.098; the second one is the interaction between AOI*lesson$F(1,41)=15.182$, p=.000, partial η=.270; third is familiarity*AOI*lesson interaction, $F(1,41)=7.442$, p=.009, partial η=.154, the last one is TypeOfImage*AOI*lesson $F(1,41)=64.619$, p=.000, partial η=.612 interaction. As in other sections, the plots and interpretation of the those interactions is given below.

Table 38: Descriptive Statistics for Total Visit Durations in Experiment 3

Familiarity/Type of Image/AOI/Lesson				Mean	Std.Deviation	N
First Study (1)	Animated (1)	Image (1)	M	0.145	0.081	21
			MSP	0.158	0.156	22
			Total	0.152	0.124	43
		Text (2)	M	0.907	0.081	21
			MSP	1.008	0.310	22
			Total	0.959	0.232	43
	Static (2)	Image (1)	M	0.240	0.092	21
			MSP	0.260	0.179	22
			Total	0.250	0.142	43
		Text (2)	M	1.328	0.152	21
			MSP	1.746	0.664	22
			Total	1.542	0.525	43
Second Study (2)	Animated (1)	Image (1)	M	0.222	0.156	21
			MSP	0.097	0.084	22
			Total	0.158	0.138	43
		Text (2)	M	0.827	0.152	21
			MSP	1.072	0.354	22
			Total	0.952	0.298	43
	Static (2)	Image (1)	M	0.250	0.124	21
			MSP	0.205	0.110	22
			Total	0.227	0.118	43
		Text (2)	M	1.318	0.195	21
			MSP	1.801	0.733	22
			Total	1.565	0.589	43

Table 39: Repeated Measures Design Results for Total Visit Durations in Experiment 3

Tests of Within-Subjects Contrasts							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Familiarity	1.511E-08	1	1.511E-08	.000	.999	.000	.050
Familiarity * lesson	5.781E-05	1	5.781E-05	.003	.957	.000	.050
TypeOfImage	9.875	1	9.875	79.464	.000	.660	1.000
TypeOfImage * lesson	.555	1	.555	4.466	.041	.098	.541
AOI	95.477	1	95.477	563.065	.000	.932	1.000
AOI * lesson	2.574	1	2.574	15.182	.000	.270	.967
Familiarity * TypeOfImage	3.927E-08	1	3.927E-08	.000	.999	.000	.050
Familiarity * TypeOfImage * lesson	5.934E-05	1	5.934E-05	.003	.957	.000	.050
Familiarity * AOI	.005	1	.005	.147	.704	.004	.066
Familiarity * AOI * lesson	.229	1	.229	7.442	.009	.154	.759
TypeOfImage * AOI	5.627	1	5.627	64.619	.000	.612	1.000
TypeOfImage * AOI * lesson	.296	1	.296	3.398	.073	.077	.437
Familiarity * TypeOfImage * AOI	.020	1	.020	1.237	.272	.029	.192
Familiarity * TypeOfImage * AOI * lesson	.031	1	.031	1.902	.175	.044	.270

a. Computed using alpha = .05

Tests of Between-Subjects Effects							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Intercept	180.199	1	180.199	703.963	.000	.945	1.000
Lesson	1.653	1	1.653	6.458	.015	.136	.699
Error	10.495	41	.256				

a. Computed using alpha = .05

The interaction between type of image and lesson is given in Figure 29 for total visit duration. The time spend on static images are higher for both M and MSP lesson, in which MSP lesson has higher visit durations than M lesson. Figure 30 points the difference between lessons in terms of AOI. The total visit duration on images is smaller than text, additionally both lesson has not differentiated in total visit duration on images, but the total visit durations changed on text for M and MSP lesson. MSP lesson higher proportion in terms of time spend on it compared to M lesson. This difference can be resulted from either increased spend time on text, or decreased spend time on images, since learners does not attend images on MSP as reported before. The interaction of type of image and areas of interest is given in Figure 31. The text has higher proportions in terms of time spend on it when compared to images. Moreover, total visit durations increased in the case of static images. Although the familiarity has not main effect, it interacts with AOI (Figure 32). There is a difference between first and second study in terms of text and images. The relative time spend on images increased in second study, but relative time spend on text decreased in second study, however the proportion of the time spend on text is quite higher than images in both conditions.

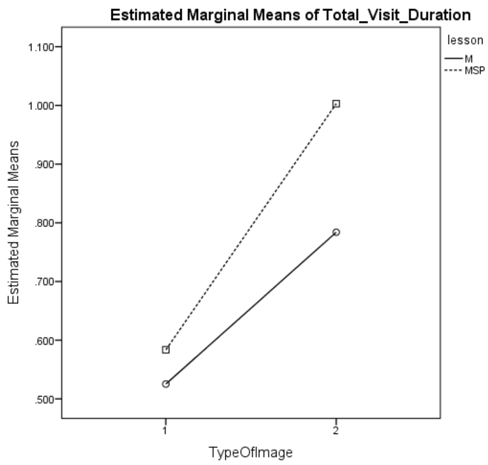


Figure 29: Plot of Interaction TypeOfImage by Lesson in Experiment 3

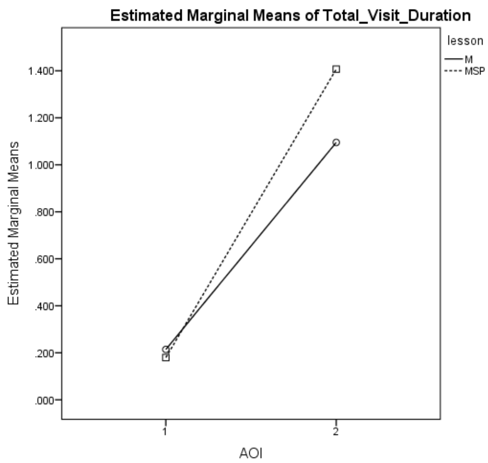


Figure 30: Plot of Interaction AOI by Lesson in Experiment 3

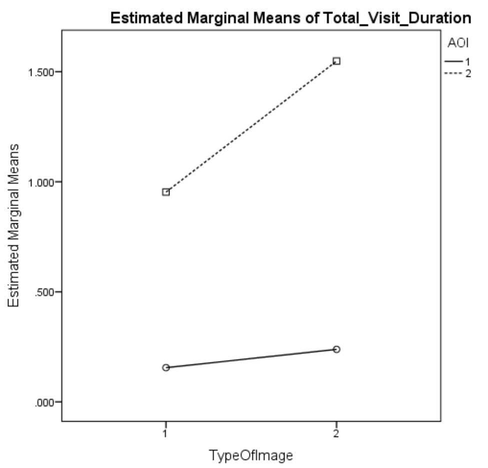


Figure 31: Plot of Interaction TypeOfImage by AOI in Experiment 3

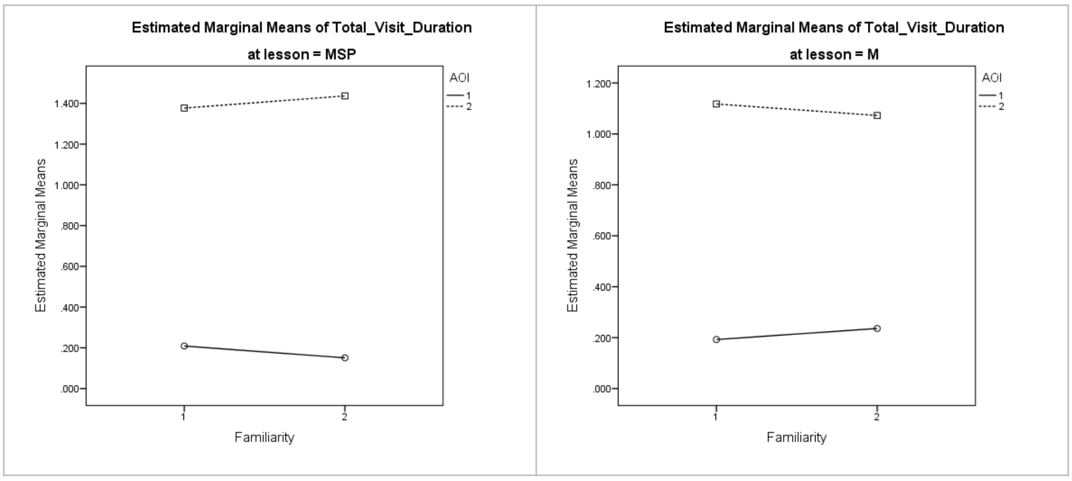


Figure 32: Separate Plots of Interaction Familiarity, and AOI by Lesson in Experiment 3

The total visit duration was the last dependent variable investigated in Experiment 3, even in the study. The results of three experiments will be summarized with some discussions in the next section.

4.6. Summary for Results and Discussion

The general structure of the study stands on the results of the achievement tests and measures of the eye tracking. At the very beginning of the chapter, the results of the paper pencil based achievement tests are given. These results showed that the redundancy could function as a supporting phenomenon if adequate conditions created, intentionally or unintentionally. In our case, for example, although the chemistry with narration version of lesson is the best example of the multimedia learning, participants show greatest performance in chemistry with subtitles lesson, in which narration or subtitles act as a redundant information. In self paced mechanism lesson, availability of control on speed does not increase the achievement in post test, probably, because of the optimum speed of system paced lesson. At the end of the achievement test result, correlations between achievement test and eye tracking measures are reported, as a suggestion to create a bridge between well established measurement methods of multimedia learning and online measures emerging in literature.

Afterward three experiments explained in detail. The *t* and *F* statistics were given with the strength of association values. Experiment 1 consists of one lesson. The intrinsic and extrinsic factors are analyzed differently. The intrinsic factors such as gender and university excluded from further studies, because of not having a place at the core of the study. The extrinsic factors i.e. familiarity, fragments analyzed by repeated measures ANOVA in which three eye tracking metric used as dependent variable. These are fixation count, fixation duration, and total fixation duration. In Experiment 1, the difference between materials (list of materials used experiments) and experiment (explanation of process in chemistry experiments) found to be significant for fixation count and fixation duration dependent variables. Total fixation duration produced significant results for familiarity in Experiment 1. In Experiment 2 fragments, areas of interests' variables are significant as within subject design factors, in addition to lesson that is significant as between design factor. Fragments and AOI are only significant as main factors for fixation duration metric in experiment to. Last dependent variable tested in Experiment 2 is the total visit duration, which produces significant differences for fragments and lesson. In Experiment 3, two versions of mechanism lesson is studied by different group of participants twice. The same dependent variables used in Experiment 3 as in Experiment 2. Fixation count values are significantly different in first and second study (familiarity), on text and images (AOI), and animated and static images (TypeOfImage). For fixation duration, on the other hand, TypeOfImage, AOI, and lesson are significant main factor. Last metric observed in Experiment 3 is the total visit duration, in which same factors are significant as in fixation duration.

The possible reasons or the outcomes of the findings reported in this chapter will be reconsidered in the next chapter again. The next chapter will include the design issues that might be affected current results. Consequently, the study will be concluded by connecting the results to the framework of multimedia theory with implications.

CHAPTER 5

CONCLUSIONS AND IMPLICATIONS

Current study joins multimedia literature with a somewhat newer instrument, eye- tracker. In this study, not only did a new instrument come into the picture, but also a new design approach was advanced to investigate the factors affecting the multimedia principles. In other words, in this study, what happened in the case of same multimedia content studied more than once for same learner with a time break was investigated. Eye tracking instruments were used to record eye movements of learners during in both study sessions. To observe the factors affecting eye tracking measures and achievement in multimedia learning three experiments were designed. The conclusions about experiments and implications of those experiments can be found in the following sections coupled with the principles of multimedia. In other words, how selected principles that are embedded in the contents are used in experiment interacted with multiple exposures to learning material and how they affected the achievement tests applied following the experimental sessions.

At the beginning of the chapter the results will be discussed in terms of the dependent variables tested in the experiments, afterwards the implication of those results are connected to the multimedia literature in the light of principles of Cognitive Load Theory and Cognitive Theory of Multimedia Learning. This connection will be established by the help of the research questions asked at the end first chapter. Those questions and the outputs of the experiments integrated to achieve conclusions about study. During discussing research outcomes, unexpected results of the study were discussed in line with their possible reasons. After stating the conclusions, the study summed up with implications in terms or theory, practice, and future implications.

5.1. Revisiting Research Questions by the Light of Results

The research questions of this study were given in section 1.3. The reasoning behind those questions were explained in Chapter 1 and Chapter 2 in detail. Each research question listed below is linked to the experiments of the study to provide guidance for interpreting results. Those linked will be discussed later on current chapter.

1. It is known that, multimedia heuristics has been varied according to learner's prior knowledge. The advantages of the multimedia content have been lost when the learners' knowledge about the domain increases. In this study, it will be observed that, what changes in subject's studying strategies when learners have familiarity about the material, which is created by studying same content repeatedly? It is assumed that the online and unobtrusive measures of eye tracking could reveal data about "how the responses of the participants will change in the case of multiple readings of same content".

   To test the question stated above, it is hypothesized that the familiarity created by multiple study of same multimedia content will not produce significant differences in terms of eye tracking results.

This question is tested throughout the study. All experiments in the study include whether paired sample comparisons of means or repeated measure analysis of variance. Any difference in within group comparisons can be attributable to the factor created by repeated study of same content. The findings of three experiment given in previous chapter will be integrated here to achieve conclusions.

2. Does the eye tracking data yield compatible results with the traditional measurement methods of multimedia learning such as recognition, recall, and transfer tests in the presence of familiarity about content?

   The question above hypothesized as follows "there is no relation between paper-pencil based achievement test applied after studying multimedia content and eye tracking measures recorded during multiple studies of same content"

This question points the link between eye tracking measures and results of eye tracking. Eye tracking is the measure for where the learner looks on screen during reading. It is not a direct measure of learning by looking or watching. Eye tracking should be considered just as an online tool to observe the eye movements of the learners during study. Recording eye movements can be provide data about utilization of the multimedia content by the learner, but tell nothing about the amount of learning in advance. So investigating the strength of the association between eye tracking metrics and amount of learning measured by achievement tests is a necessary condition for working eye movements as a tool to study multimedia learning.

3. Since the pace of the presentation has effect on existence multimedia principles, how do responses of the learners change when they re-study of the same content in a pace controlled environment? It is assumed that re-study of the same content and pace of the study will interact by learner's perceived complexity of material, increased familiarity with revisits and cognitive load resulted from ability to control speed. Eye tracking will provide online data about "what might be going of the learners mind" during self-study.

   The hypothesis about control is that "having control over multimedia lesson does not produce significant differences in multiple study of same content"

Eye tracking is not a direct tool for observing the learners' mind, but investigating variances of eye movements in different conditions can provide cues to find out if there is a relationship between eye movements and the processing of the multimedia content. Since, pace is one of the boundary conditions in complex situations; presence of the pace in a relatively complex presentation could have effect on movements. This question explores the relationship between eye movements and the having control on the speed of presentation of multimedia materials to be learned. Experiment 3 was designed to examine this relationship.

4. Does an example of effective well-designed multimedia material proposed by Mayer work within different science domains such as chemistry or mechanism? It is known that multimedia advantage appears most "*when the material is complex and presented at a rapid pace for the learner"* (p.275) in science domains. In this study it will be investigated that what happens when a parallel design is used with topic other science topics."

   As a last hypothesis in the current study, it is assumed that new domains will be produce differences terms of working principles in multimedia learning.

The worked examples of multimedia leaning do not show broad variance. The classical examples of multimedia learning are how breaks work or how lightning strike etc. However, Mayer suggest that a diversification in topic is needed for multimedia learning. Experiment 1 and 2 were designed to test well established principles of multimedia learning in a different domain that is chemistry. In addition to Experiment 1 and 2, Experiment 3 also can be accounted as a domain different from classical examples.

The four questions stated above constitute the framework of the experiments reported in this study. The factors that might affect eye tracking or achievement tests results are given below. The theoretical connections to multimedia learning of those factors and results of the experiments are integrated in following sections.

5.1.1. Familiarity as Measure in Multimedia Learning (Experiment 1, 2, and 3).

Familiarity is not construct, which is directly referred, in multimedia content. The closest concepts to familiarity in multimedia theory are the prior knowledge. Prior knowledge is one of the topics that multimedia learning has many considerations on it. Mayer (2009) conceptualizes prior knowledge as a portion of Long Term Memory (LTM) and connects prior knowledge with a not to verbal or pictorial model for integration process in Cognitive Theory of Multimedia Learning (CTML) (See Figure 2:). Prior knowledge considered as a whole in Mayers' studies and have not been discussed for its subcomponents. Prior knowledge treated as a unique concept, and used close to commonsense meaning. He does not have deliberate special interest of the structure or the functions of previous knowledge. Although he added prior knowledge as a component into his theory, he left discussing details of connection between prior knowledge and CTML to further studies as mentioned several times in his book (Mayer, 2009).

Mayer connects prior knowledge concept with coherence, pre-training, and personalization principles. Prior knowledge is described in his writings as low- high knowledge learners or less-more experienced learners. Mayers' conceptualization of prior knowledge came into picture throughout principles as "expertise reversal effect" (Kalyuga, 2005; 2007) most of the time. CTML's theoretical consideration on prior knowledge seem to be derived from "reversal effect"s practical point of view but Cognitive Load Theory (CLT) has wider point of view for prior knowledge because of its' inherited boundary on load since prior knowledge has more related to load compared to CTML's modular structure in which prior knowledge is placed at the end of whole process.

Traditionally prior knowledge and load was measured by asking simply "do you have any idea about … topic" or "how much effort you spent to complete this task". And participants fill a Likert type question scaled none to all. (Ayres & Paas, 2012; Ozcelik, Arslan-Ari, & Cagiltay, 2010). Although these self report measures are simple to administer for researchers, van Gog, Kirschner, Kester, & Paas, F. (2012) argue that those subjective measures are subject to change test to test. So, instead of using subjective self reports to achieve the levels of prior knowledge and cognitive load, using alternative methods worth to be try. But, in the case of prior knowledge we do not have a nominal dichotomy such as gender or standardized scale such as IQ that might be accepted by many, and might be put all objections aside. On the other hand, if self reports used as a tool to measure participants' prior knowledge level before study, in the same manner post tests can be used as an option to measure learning. Since this preposition was not acceptable, researchers should be particular about post tests' certain qualities such as reliability and validity. In current study, the achievement test was developed by considering those limitations. The test reliability and their difficulties were reported in results section.

In experimental design, there are several methods to eliminate the possible effects of confounding variables that are randomization, matching, holding extraneous variable constant, building extraneous variable into design, counterbalancing, and ANCOVA (Johnson & Cristensen, 2004). Familiarity about content is created by treatment itself in this study. The differences between first and second study observed by eye tracking. Consequently, familiarity became a factor in the study, and eye tracking measures became dependent variables by design. The differences in eye tracking metrics between first and second study will indicate the effects of familiarity on strategies to study a multimedia learning. In this study familiarity used a term for prior knowledge because of "Creating knowledge by treatment itself" would have been a bold statement, and can be falsifiable easily by a different point of view, for example expertise-novelty continuum used heavily in multimedia studies specially in CLT (see Van Merriënboer & Sweller, 2005for review). Familiarity is used to underline the connection between the concerns of the study and structure of multimedia learning rather than quantifying knowledge that participants might have at the beginning but eliminated by design. But one should be cautious about to make any generalization for familiarity created by multiple studies. In this study, the inner characteristics of the familiarity had not been assessed in detail. For example, what learners remember about first lesson were not observed in design. Their attention might be whether on content or design in the first study. The change on eye behaviors can be resulted of either explicit or implicit learning. On the other hand, the dichotomy of implicit and explicit learning can useful in discussing the differences between first and second study. Participants said nothing about how should study lesson. They freely decide where should be attended or what should be remembered. Since in both conditions they decided in their strategies of study by themselves, the difference between first and second study is probably an unconscious one. In first study they have acquire unconscious implicit knowledge in addition to explicit knowledge about content. The point is whether they are changed their strategies of study in second reading according to explicit knowledge or implicit knowledge they have acquired in first study. Since they try to increase their amount of explicit knowledge in both studies, the difference can be accountable to the implicit learning, which is defined as becoming sensitive to regularities in environment unconsciously (Cleeremans & Dienes, 2008). Since the learners do not changed their strategies consciously, the difference resulted from the familiarity can be a part implicit learning.

Why prior knowledge has been operationalized by first and second study and it is named as familiarity has been discussed so far. In the previous chapter, results of experiments for familiarity for eye tracking metrics. These metrics are fixation duration, fixation count, total fixation duration, in Experiment 1; fixation duration, fixation count, and total visit duration in Experiment 2 and Experiment 3.

The significant results regarding familiarity as a within-subject factor observed for total fixation duration in Experiment 1. In Experiment 2, which investigated the presence of the narration, there is no significant effect resulted from familiarity. But in Experiment 3 familiarity affected some eye tracking metrics in addition to some interactions. Mean of fixation count is affected from familiarity as a main factor that is there is difference in the numbers of fixations in first study and second study in mechanism lessons. Familiarity interacts with lesson, types of images, and areas of interests also. In addition to a two-way interaction, there are interactions among familiarity, type of images, and areas of interest in Experiment 3 in terms of fixation count. Although there is main effect of familiarity on fixation duration, there are other significant interactions with familiarity. These are familiarity*lesson, familiarity*type of image*lesson, familiarity*areas of interest*lesson, and familiarity*type of image*areas of interest*lesson. For total visit duration only familiarity*areas of interest*lesson interaction is significant.

What tell us those main effects and interactions? Familiarity is one of the factors that might affect the eye movements. Those eye movements can be affected from the content of multimedia also. These are text vs. images, animated vs. static images, and pace of the lesson (Experiment 3). The fixation duration is affected from familiarity in chemistry lesson in which there is no subtitles (Experiment 1). However, the familiarity does not affected if there are subtitles in presentation. The effect of familiarity is lost in Experiment 2, whether narration exists or not. This shows that there are design factors effecting eye tracking measures, but these measures are sensitive to the presence of the text and narration as medium of information channel. If text is used as an information source, familiarity is not observed as a main effect, although there is still considerable difference between chemistry with subtitles and chemistry with no voice lessons in terms of fixation count.

Since eye tracking metrics affected from length of the lesson, except fixation duration, only mean of fixation duration metric reported here to visualize the differences between lessons (Figure 33). The highest mean of fixation duration observed in chemistry lesson in which a narration exists in background. Since there is no distractor in screen, Participants' attention can be remained on video only during study. CWS lesson's means for fixation duration is smaller than C lesson. This difference resulted from the presence of subtitles probably. But, difference between first and second study for the CWS lesson is greater than C lesson. The smallest mean values of fixation duration for chemistry lessons were obtained in CNV lesson.

C lesson was an example of Mayers' CNA (Concise Narrated Animation) in its context there were no extraneous words or pictures. Mayes's definition of CNA has the highest means for fixation duration values for five lessons. So, it can be said that, longer fixation duration values produced when Mayer's CNA definition was fulfilled.

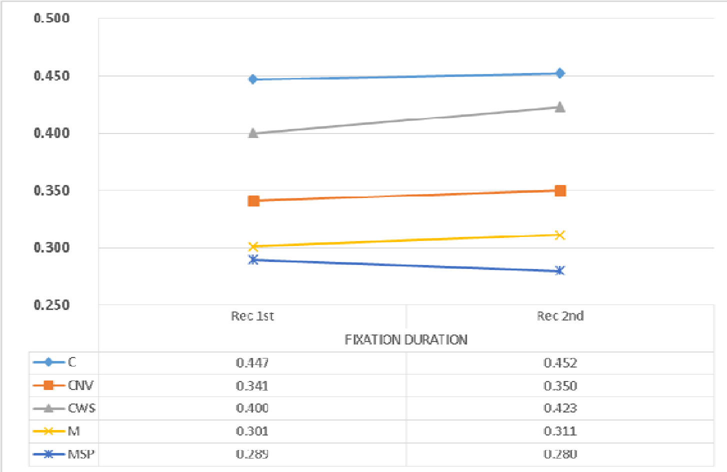


Figure 33: Mean of Fixation Duration for All Lessons

For the two version of mechanism lessons fixation values showed for first and second study. For system paced mechanism lesson mean of fixation value is smaller than second, self paced mechanism lesson, same value is greater. Participants' mean of fixation values are getting smaller when they have information about context if they have control on it like in MSP lesson. On the other hand, mean of fixation durations get bigger if they have no control on presentation like in M lesson.

In this point, this question comes to mind achievement test scores have been changed according to type of presentation the answer of this question can be averages of five lessons' achievement test scores. Since there are no participants from Atılım U in the CNV lesson condition, the comparisons were made only among METU students. The achievement test scores observed in chemistry lessons and mechanism lessons are given in Table 40. The greatest score for chemistry group is for CNV lesson. The second one is CWS lesson, and the score of C lesson is the third one. This showed that learners showed greatest performance in CNV lesson in which no narration exists. The redundancy in chemistry lesson had not been helped to learning in chemistry. But, text creates better performance when compared to narration, contrary to the redundancy principle of CTML. Redundancy principle states that, "People learn better from animation and narration than from animation, narration, and on on-screen text", but this is not case in current study.

Table 40: Achievement Test Scores According to Lesson

Lesson	N	Achievement Test Score
Chemistry with Narration (C)	10	9.00
Chemistry with Subtitles (CWS)	10	9.25
Chemistry with No Voice (CNV)	10	9.75
System Paced Mechanism (M)	10	10
Self Paced Mechanism (MSP)	10	8.55

In mechanism lesson system paced version has higher values than self paced counterpart (Table 40). The MSP lesson has higher duration than M lesson in the first study, but this difference is lost in the second study, their mean durations to complete lesson almost identical. So, what can be the reason of the differences between M and MSP scores, if the pace was not a factor? The remarkable decrease in second studies of MSP lesson can be a sign of motivation to study content. Participants has tendency to finish lesson as soon as possible if they have control on pace. If they had no control on speed, they might have been enforced to study content even though they are unmotivated or reluctant to study. The detailed discussion considering CTML' principles and components of CLT are given in later parts in this section.

Previous paragraphs were about the second research question of the study. That was about a comparison between achievement test scores of the current study. The achievement test showed that, the scores of recall and recognition test could be considered as an alternative approach to the transfer questions for original studies of Mayer's multimedia learning experiments. Mayer thought that the ultimate aim of the learning is the transferability of knowledge to the new situations. Although transfer of knowledge is a desired ability, it requires having knowledge and comprehension, which are more basic levels of cognitive domain. In classroom settings, unfortunately higher levels of cognitive domain, application, analysis, synthesis, and evaluation could not be achieved easily. The further discussion and suggestions about this topic can be found in the last section (Section 5.4).

5.1.2. Pace and Speed of Presentation (Experiment 3)

The speed of presentation is one of the characteristics that makes multimedia learning complicated. The speed of presentation or the ability to control pace of presentation affects the perceived complexity of the design. This complexity seem to be dependent on the learner's previous knowledge. If learner has prior knowledge about content, the principles of multimedia design seem to be lost the advantages over non-multimedia counterparts. To solve the equation between complexities, pace, speed and prior knowledge for multimedia designs many studies have been conducted in both CTML and CLT. In this study, factors effecting eye tracking measures were investigated in Experiment 3. In these lessons, there were text areas, static images areas and animated images areas. The difference between those areas assumed to be resulted from pace and prior knowledge. The differences between first and second study refers to prior knowledge about content, and the differences between self and system paced mechanism lessons refers to sense of control and speed of presentation variables.

In Experiment 3, type of images, and areas of interests main effects are significant for fixation duration. Moreover, fixation counts familiarity is significant also. But, the between subject main effect of lesson (that is the overall difference between system paced mechanism vs. self paced mechanism lesson) is not significant for fixation count, but significant for fixation duration. But the interactions of lesson with type of image, areas of interest and familiarity are significant. So although pace does not a factor in lessons, but it has effects on design characteristics of lessons.

To visualize the differences between type of images and text for the first and second study are given in the Figure 34 for fixation duration. Figure 34 showed that mean of fixation durations on static images for first reading were almost identical. But, there is a gap between in second reading. The duration in MSP lesson decrease slightly, and the duration in M lesson increased slightly. For animated images, the mean of fixation durations were higher than static image both first and second reading. Although the durations for second reading increased on animated images in M lesson, the durations in MSP lesson decreased for second reading. The mean of fixation durations on static images in M lesson and animated images in MSP lesson for second reading became almost identical.

Figure 34 is a summary for hypothesis about speed, pace, and control, in fact. If participants have control on pace, their attention on animated images seem to be lost. Having information about animation and its contribution (none in this case) can be reason for loosing attention. The mean of fixation durations on animated images became identical like in the case of no control in the second reading. In complex materials, having no control over content maximizes multimedia advantage. In our experiments achievement test scores for M lesson were higher than MSP lesson. This finding supports Mayer's studies. Participants have no control over pace in M lesson get higher scores than participants of MSP lesson.

The time given in M lesson is close to the duration observed in second study of MSP lesson. This shows that the time limit in M lesson is not high; they could easily study content without feeling any restraint. The possible effects of longer or shorter durations of M lesson are discussed in below in this section. It is expected that the change in time would change the cognitive load of the participants, which is one of the key factors on redundancy and temporal contiguity.

On the other hand, M and MSP lessons do not consist of only images. Text accompany images in those pages. The verbal information was presented to the learner in printed form unlike Mayer's CNA's. In Concise Narrative Animations the verbal information is carried by sound to the learner. But, in M and MSP lessons no sound channel have been implemented due to its design. If audio channel exist in the M lesson, the MSP lessons' self paced characteristics would be effected by the speed of narration. By ignoring narration in M and MSP lesson, learners get the full control on pace and speed in lesson.

Figure 34: The Mean of Fixation Durations Differences between Type of Images, and text for First and Second Study

The Figure 34 presents also the mean of fixation durations on text area in M and MSP lessons. The format of the text was same in both M and MSP lesson in every page. The images varied either static or animated. Figure 34 showed that mean of fixation durations change the neighboring images and study order. But, the difference between mean of fixation durations are not greater than .020 milliseconds. Shortest fixation durations observed during the first study in M lesson. In the second study, the mean fixation durations increased on text of static pages especially. The second study of MSP lesson has smaller mean fixation duration although they have greater values than M lesson for first study. This difference can be a result of different strategies of learners in the presence of control over pace. Their strategy of learning from the content seems to be dependent on several factors.

The first factor might be the familiarity of the content. In the first study, learners have no idea about the content and the overall design of the lesson. During the first study, they could not predict what is next and how important the content presented in current page when compared to lesson as a whole. Lack of familiarity in the first study reversed to familiarity and prior knowledge in the second study. They exactly know what they have been studied in lesson and order. Their strategy to learn the content is guided by prior knowledge about content. The differences between first and second study was discussed in the light of Cognitive Theory of Multimedia Learning and Cognitive Load Theory later in this chapter.

The second factor that might affect the eye tracking measures could be the way of placing materials into mechanism lesson. Although the analyses were focused on only static and animated images, not all images either animated or static were identical in their class. For example, some of the static images referred by text while others were not and the sizes of images different in lesson. The differences between image characteristics such as referred or not, sizes, its function in text, and their relation to text can be effective on eye tracking metrics, but in analyses in that level makes the things more complicated. The analysis in that level requires more controlled experimental designs, for example participants' characteristic, images comparability etc. For being sound such an analysis, the design should be revised for the number of variables included to the experiment.

As a last word about pace of the presentation, the existence of the "continue" button in the lesson has effects on the learner's behavior towards content and its' process shown by data revealed eye tracking metrics. On the other hand, only fixation duration has given as eye tracking metric above, because of lack of correspondence between durations and number of images on pages.

5.1.3. The Domain Variability of Multimedia Learning (Experiment 2)

Another research question to be investigated in this study is the domain of materials used in multimedia learning. Mayer (2009) says that still more research is needed to pinpoint the nature of multimedia learning. But, he does not point any specific domain to study on. As most of the other researchers, in this study new materials developed to test specific assumptions of research questions. These new materials contribute to the ecological validity (Ainsworth, 2008) of the multimedia studies although they still in science domain. The chemistry and mechanism lessons' content have different levels of technical terminology. Mechanism lesson is quite technical when compared to chemistry lesson. Chemistry lesson is a very introductory level for higher education. It has only descriptions of three experiments and names of materials used in those experiments. However, mechanism lesson has many concepts related to each other. In mechanism lessons, concepts tried to became concrete by every day examples.

In this section, research question is exploring the multimedia advantages of lesson developed for this study, if exists. The multimedia advantage is maximized in complex and rapid presentation in Mayer's CTML. The closest lesson to this definition in the current research was system paced mechanism lesson. The speed of lesson in self paced mechanism lesson was determined by learner. But, system paced mechanism does not fit the CNA (Concise Narrative Animation) definition of Mayer like chemistry lesson. The speed of chemistry lessons could not be varied due to narration since; narration restricts the change in speed. This is a drawback of the study. There was no best session to observe the maximization of multimedia advantage unfortunately.

Although current study does not contain treatment to observe maximization of multimedia advantage directly, however, it still covers indirect measures to observe the advantage. Since the level of complexities in the materials was not plain, the fluctuations in complexities in within the lessons could provide information about existence of multimedia advantage.

To achieve this goal, following segmentation can be suggested. All chemistry lesson videos can be divided into two separated fragments. These fragments in chemistry lessons have different scenes with different information load. For example, speed of the fragments in which the names of materials presented used in experiments flow faster than the scenes in which the activities in experiments defined. So, a within comparison for chemistry lesson could provide clues for complexity in addition to its interactions with multimedia advantages. The name of materials is not related to fragments, as they were presented for about two seconds and their names were given in written or narrative form during the lesson. These fragments looked like simple short term memory span-task experiments. Learners should remember a list of materials. But, the other fragments of chemistry in which how experiments conducted contain interdependent serial activities which have require use of inactive components of memory unlike serial span-task (Shipstead, Redick, & Engle, 2010)

The main effects and interactions of fragments with other design factor such as areas of interest can be accounted as the measures domain variability in chemistry lessons. The results given in Chapter 4 showed that fragments in chemistry lesson showed significant differences for fixation count and fixation duration in Experiment 1. In Experiment 2, in addition to fixation count and duration, the difference is significant for total visit duration also. This showed that, there are within differences in chemistry lessons in terms of eye tracking metrics. That means eye movements changed according to the context of lesson. Eye movements affected fragments, and those fragments show interactions with other design factor such as areas of interest. The interaction of fragments with familiarity cannot be accounted as a design factor, rather interpretable as the interaction of content variability and implicit prior knowledge created by repeated study.

The figure in chapter 4 presented two way interactions. To depict whole output obtained from different experiments for fragments, mean of fixation durations on is given in Figure 35. The mean of fixation durations showed greater variance on experiment areas than on materials areas. The smallest values of fixation duration means observed in CWS lesson during experiment fragment. The longest durations for mean of fixation duration observed on experiment fragments of chemistry lesson in which no subtitles exist. It shows that in chemistry lesson, learners could have been concentrated only on the activities in video. The mean of fixation durations increased in second study of CWS lesson on materials, but same values decreased in CNV lesson.

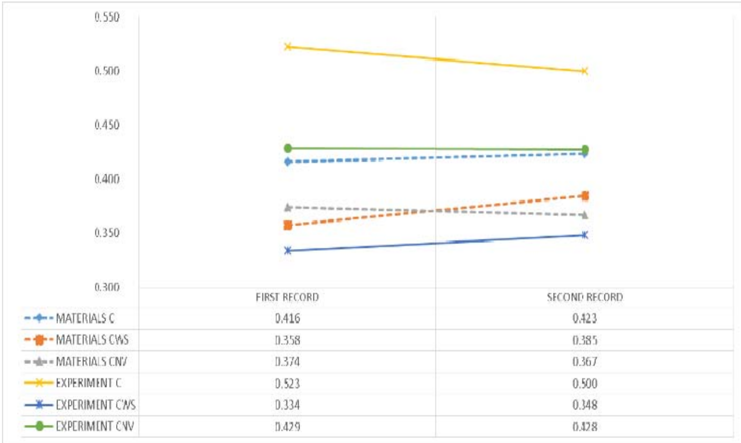


Figure 35: Comparison of Fixation Duration Means on Materials and Experiment Fragments in Chemistry Lessons

The values in Figure 35 pointed that the fixation durations differed, if the content of the lesson changed. This difference could be result of either content or familiarity of learner created by experiment itself. Since, only difference between chemistry lessons are the presentation mode of the verbal information, the differences between lessons in the same study can be interpreted as effects of mode of presentation. In Chemistry lesson words are transferred to learner by ears. But, CWS lesson uses both only eyes and ears for word unlike CNV lesson in which only eyes are used for words. The connection of these differences into CTML and CLT will be made in the following section. The next sections discussed the theoretical framework and the possible uses of study in practice followed by drawbacks and suggestions for further studies.

5.2. Theoretician "How this Study Made Contribution to CTML and CLT?"

In this study two main theory of multimedia learning were considered. One was Mayer's Cognitive Theory of Multimedia Learning and the other was Sweller's Cognitive Load Theory. Although both theories use some common roots as Baddaley's Working Memory, their scopes about multimedia learning are different. CLT concerns about reducing cognitive load to maximize learning, and CTML concerns the design heuristics to maximize learning. CLT uses generally mathematical models and schema construction to test its hypothesis, but CTML uses science experiments as the source of multimedia principles (Reed, 2006).

This study is designed to contribute prior knowledge as familiarity and some design issues of CTML and CLT by using new methodology in general. The components of CTML are given in Figure 36 to refresh the concepts that are given at the very beginning of the study again. In this figure, three different routes for spoken/printed words and pictures are given in one diagram unlike in Figure 2: in which three different diagrams used originally by Mayer.

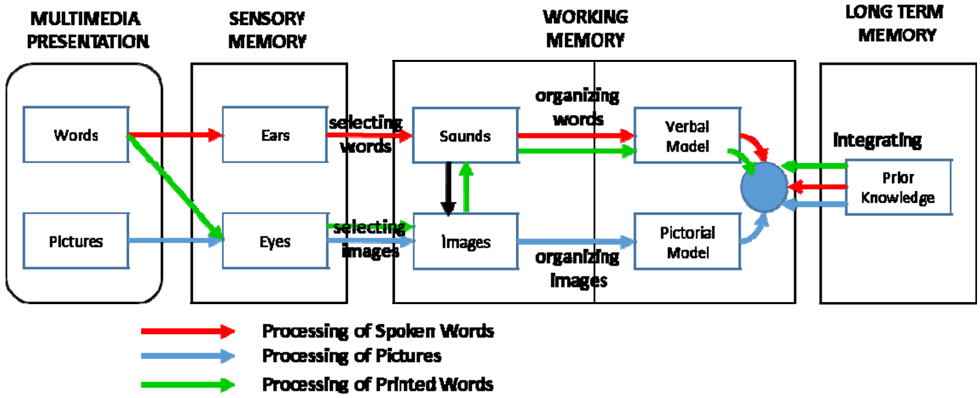


Figure 36: Cognitive Theory of Multimedia Learning, Mayer (2009) (A new presentation)

The focus point of current study in diagram is the circle between WM and LTM. CTML considers prior knowledge as unique structure that fits all. But, LTM is not unique construct that interacts with working memory in one way. The relation between WM and LTM is more complicated than CTML assumed. LTM has indefinite capacity theoretically and many subcomponents i.e. declarative, semantic, implicit, procedural, explicit etc. Since prior knowledge is measured by Likert type question(s) in multimedia learning, this method might be a little bit imprecise what learners have in their mind before study, or how they integrate prior knowledge with WM model. In other words, LTM box in model should be detailed to define relationship between multimedia principles and prior knowledge more accurately by considering prior knowledge has many levels and components in definition.

The arrows in CTML model are one way, their direction point bottom up processing. The only top down process in model takes place in integration part, but it is known that, bottom up and top down processes have always interact in all steps. Those interactions were not observable in the model. This study showed that eye tracking measures are affected from implicit previous knowledge of learners. If learners has familiarity about the content, their eye movements and attention changed in terms of areas of interest, fragments and types of images consequently. The findings of this study can be placed somewhere experts and novices, but closer to novices. Since familiarity has effect on the design and processing issues, observation of the components of the familiarity is needed in further studies. Additionally in those studies declarative and implicit dichotomy should be underlined for prior knowledge.

The schemata or their generation (from CLT) is not observable directly in novice-expert continuum. Building schemata is not none or all process taken in place at one time. In CTML before achieving the level of transfer there should be many hidden cognitive levels and steps that learners should overcome. Recently retention and transfer tests have used to measure the outcome of learning even if eye tracking methodology had been integrated to study. (Johnson & Mayer, 2012). Retention is defined as remembering ability of presented material, and transfer is defined as understanding of material. In retention test (Mayer, 2009) any type of questions to measure what learners remember from presented material can be used, but in transfer questions only essay type questions were used to see whether learner apply the information given in presentation to the new situation or not. Although, Mayer said that the aim of multimedia learning is to promote transfer as well as retention, these two types of learning are not observable in the model. In the current study, by studying same material more than once creates prior knowledge as a counterpart of the retention tests of Mayer. Since the target of the retention test is to measure the ability of remembering, reproducing, or recognizing presented material, the first study could provide learners' knowledge at that level, however achieving to the level of transfer was questionable by presenting short materials only twice which is novel for learners.

Cognitive Load Theory, on the other hand, view LTM as a source of schemata that affect load the most, however the processes of schemata formation was unclear for CLT (Moreno, Park, 2010) Additionally, the CLT has no visual model accepted as CTML. The models for CLT were given in Figure 3 and Figure 4, but these figures are not informative to discuss the relationship between CLT and current study. Figure 4 contains evaluation of CLT from first to third version. First stage contains schemata as an only source for load. In the first version of CLT, controlled and automatic memory processes were thought to be only sources for load which are qualitatively different. The second version of CLT include intrinsic and extraneous load as factors, which occupy WM capacity. In second version, the loads thought to be not effected by instructional design. But in third version germane load came to picture in addition to intrinsic, and extraneous load types. Moreover, those loads thought to be affected by instructional design. Intrinsic load refers to the content to be learned. The higher intrinsic load means the harder materials to be learned by nature. Extraneous load refers to the design factors, which are not directly related to content but load created by design. The schemata concept of older version come to picture in third version of CLT as germane load which refers to the creation of automatization or change in quality in process. Figure 37 contains an analogy for four (three types of load and WM capacity) construct of the CLT to achieve learning. In this diagram, the volume of the cone refers to working memory capacity; the diameter of the bottom refers to germane load. The green marbles refers to intrinsic load of the content. And the 12 point star refers to extraneous load. The instructional design can change the diameter of the cone bottom, and the existence of the star. The increase in diameter make easier flow of green marbles to achieve learning. Reducing extraneous load make the star smaller, smoother or even disappear which also positively affect the flow of targeted content. Extraneous load and germane load can be changed by design, but the characteristics of the intrinsic load seem to be unaffected by design at for now. To follow same analogy, increasing flow seem to be dependent on the sizes or the shapes of marbles. If the content can be divided into smaller parts, or the shape conserved as spherical, the flow can be affected positively.

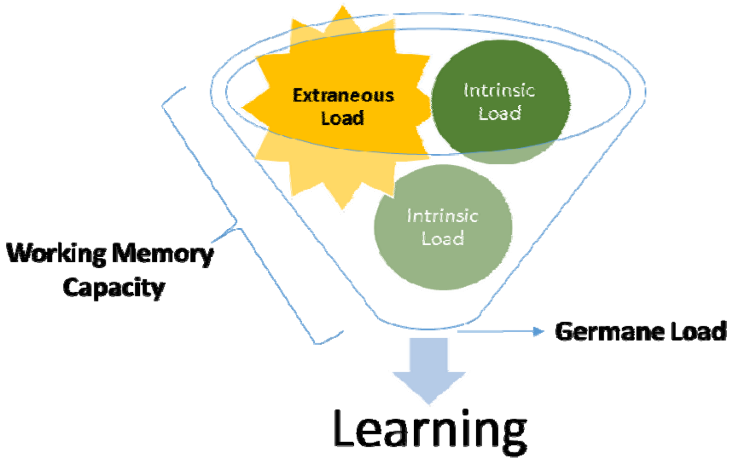


Figure 37: A New Diagram for Cognitive Load Theory.

Current study supports in this analogy by providing examples of loads. The intrinsic loads are the complexities of chemistry and mechanism lessons. The load of mechanism lesson is higher in Experiment 2, when compared chemistry lessons used in Experiment 1 and 2. The extraneous load are implemented in the designs of lessons. For example, availability of control or narration are two factors that might be affected extraneous load. The effect of familiarity in on processing can be suggested as an example of germane load, since by implementing familiarity can be accounted as the first step for constructing a schemata. All these factors in the study have effects on WM capacity, and as a result of change in capacity and processing, the outcome in learning can be changed in different lessons. Nevertheless, this analogy is a suggestion that the findings should be treated as a set of results paving the way for future studies with experimentally controlled learning environments.

Another way to foster the learning could be capacity of WM by training. There were many studies to use WM more efficiently, but closer look to these studies made their validity questionable (Shipstead, Redick, & Engle, 2010). Therefore, as a method to foster learning, WM capacity as an individual construct is not an agreeable method of learning yet.

The contribution of the current study to the models of CTML and CLT does not only come from the repetitive nature in experiments but also testing different modalities, and different load levels. The processing routes of CTML are tested with different modalities presented in lessons of chemistry. Chemistry with narration lesson follows the route of processing of spoken words and pictures (red and blue line in Figure 36), whereas Chemistry with no voice lesson follows green and blue lines, in which eyes are used for both processing for spoken words and pictures. The last version of chemistry with subtitles lesson use all colors to process information. But, in CWS lesson learner has the opportunity to use either routes for processing of words. Contrary to expectations, the amount of learning in CNV lesson was greater than C and CWS lessons in study. This finding is compatible with the redundancy principle, but incompatible with modality principle, which assumes that learning is better from animation and narration, than from animation and on-screen text.

The comparison made in Figure 35 provides information to compare those processes by fixation duration metrics. The differences between fixation durations indicate that the fixation metrics are affected from the modalities of presentation. Highest mean of fixation durations obtained in C lesson shows that learners watch the presentation, and listen to the narration. But, in other version of chemistry lesson, the existence of subtitles are shortened the fixation durations because of reading behavior produce fixation duration about 150 milliseconds. In that point, the comparison between CNV and CWS can yield information about processing of materials. If the existence of the subtitles generates shorter the fixation durations, it would be expected that CNV should produce shorter fixation durations if the narration helps processing of information, that is learners have opportunity to watch content if narration present in environment. Contrary to expectations, CWS lesson produces shortest fixation durations in experiment part in both studies. However, if the material parts are taken in to account, the fixation durations of CNV crossed with CWS lesson. That means the presence of narration interacts with familiarity of content. The expected order of shortest fixation durations in CNV were achieved in the second study. The reason for producing shortest fixation durations in CWS lesson can be the redundancy created by narration. If the narration and on-screen test available in the environment, learners can be distracted or disoriented, and their concentration on content might be disappear. An alternative explanation can be for shortest fixation durations in CWS lesson can be as follows. In the presence of narration, learners fixation durations get shorter due to input from ears triggers the conversion images to sound in WM, that might be shorten the fixation durations on printed words consequently. This concept is known as priming in psychology. But, if segmentation is ignored the means of fixation durations turn upside down. In that case CWS gets higher scores when compared to CNV (See Figure 33). The implication of the incongruity is made in next section.

If the statement above was true, this finding would be compatible with the CLT's assumption that narration can prime the automated processes and result in increased memory load by processing or printed words. The function of load in this study should be observed in system and self paced mechanism lesson. It is assumed that the availability of self pace decreases the load by design, and increase the learning. But the scores of recall and retention test higher for system paced mechanism lesson. This could be resulted from the comfortable pace of system for learners. This might be create an optimum level of load to encourage learning. Having control on the speed does not foster learning itself, even, it might be act in opposite way as seen in self paced mechanism lesson. Participants spent more time on study and have control on pace, but their scores were not higher than system paced counterpart in term of recall and recognition test scores. This showed that some level of cognitive load could be a necessary but not sufficient condition to better learning. So, the current study made partial support to the theoretical structure of the CLT, by providing an online eye tracking data from multiple readings of same content. Learners changed behaviors in different designs if they have prior knowledge or familiarity about what they have been study. Their strategies about redundancy or spatial contiguity were affected by prior knowledge. The prior knowledge could reduce extraneous load and increase the germane load. Both has positive affect on learning outcome. But the effect of prior knowledge intrinsic load was not explored by current study. Suggestions for reducing intrinsic load were made later in this section.

5.3. Practitioner "What Makes Useful This Study for Me?"

What were the lessons a designer could extract from this study? The first outcome of the current study verifies that prior knowledge is not a static nature, which learners have. It is known that, the expertise which requires higher order cognitive processes have interaction with multimedia principles but, a lower version of expertise let me call this familiarity can be interact with multimedia principles also. This familiarity resembles in CTML's pre-training principle, which is used for managing essential processing (or germane load).

To explain the possible impacts on practical use of current study, the principles of CTML and applications of CLT on design should be kept in mind. The principles of CTML grouped into three; these are principles to reduce extraneous load, principles to manage essential processing, and fostering generative processing. These three group are comparable with CLT's extraneous load, intrinsic load, and germane load. However, which theory borrows the concepts from other was unclear. The principles of multimedia learning are repeated here with the treatments in the current study. The findings also included to the Table 41.

Table 41: Multimedia Principles Tested in Study and Results Obtained

Intention of Principle	Multimedia Design Principle	Treatment in Current Study	Obtained Results
-reduce extraneous process	-coherence -signaling -redundancy -spatial contiguity -temporal contiguity	-Redundancy tested in experiment 1 and 2	-The achievement in lessons CNV>CWS>C -The redundancy effect observed in eye tracking metrics.
-managing essential processing	-segmenting -pre-training -modality	-Pre-training tested by familiarity in exp. 1, 2, and 3. -Modality tested with experiment 2. -Segmenting tested with experiment 3.	-The main effect of familiarity measured by eye tracking metrics. -The modality effect observed in CWS and CNV lessons as between groups subjects main effect. -The segmenting effect observed in M and MSP lessons as between groups subjects main effect
-fostering generative processing	-multimedia -personalization -voice -imaging		

The principles of reducing extraneous load are coherence, signaling, redundancy, spatial contiguity, and temporal contiguity. In this study to only redundancy principle covered by chemistry lesson. The other principles of reducing extraneous load were not included. Redundancy principle assume people learn better from narration and graph than graphics, narration and on-screen text. In chemistry lesson there was only narration and graphs. In chemistry with subtitles lesson there was narration, graphs, and on-screen text. Lastly in chemistry with no voice lesson there was graphs and on-screen text. According to redundancy principle of CTML the highest scores were expected from Chemistry lesson, but in this study the highest scores were obtained from CNS lesson, in which narration and on-screen text violate modality principle. The reason for this output could be internal characteristics of the lesson or the treatment of design that is studying twice that is novel for multimedia learning setting. An instructional designer should aware of the learners' familiarity to the content if s/he uses familiar concepts or materials for learners.

The comparison of C and CNV lesson could yield a testing environment for the modality principle in managing essential processing. Modality principle assume that that narration and graphs are better from on-screen text and animation. In C lesson the both verbal (red line in Figure 36) and pictorial modes (blue line in Figure 36) used. In CNV lesson instead of verbal, pictorial (green line in Figure 36) mode is used. It is expected that, using same components of WM should create load and reduce the scores consequently. But the scores of recall and recognition test after CNV is higher than C lesson contrary to expectation. Re-reading or the internal lesson characteristic could be reason for this difference, again. Before going into segmenting and pre-training principle, checking eye tracking results of those chemistry lessons could be provide information about what happens during learning for redundancy and modality principle. But, to investigate redundancy principle the C and CNW lesson could be compared. The comparison of C and CNV can provide information about modality principle. However, in both condition the existence of subtitles in CWS and CNV lesson made the comparison of eye tracking results meaningless. The only condition to make comparisons with eye tracking metrics was the comparison between CWS and CNV The mean of eye fixation durations were higher in CWS lesson than CNV lesson. That difference is a sign of learners use narration to focus the presented materials on screen more. The difference in second study even higher than first study that showed that having familiarity about content make the use of narration more (Figure 33).

Additionally the differences between first and second study could provide information about pre-training principle. In second study, learners' familiarity can be accounted as pre-training. Pre-training increased mean of fixation durations in all lessons except MSP (self paced mechanism). On the other hand, the difference between M (system paced mechanism) and MSP lessons were point segmenting principle in that self paced segments used to foster learning. On the contrary to expectations, the scores for M lesson were higher than MSP lesson in which segmenting principle applied. In addition to test scores, the decrease in mean of fixation durations could be attributed to something else in MSP lesson, which is unknown yet (Figure 33). To solve these issue further studies should be conducted.

The generative processes or germane load is related to some design characteristics of presentation (multimedia, personalization, voice, and image principle). In current study all these principles are kept constant. There was no only words or pictures option in lessons to fulfill to multimedia principle. The narration in C and CWS lesson was the same for personalization and voice principle. And there was no talking heads in lessons (image principle).

In this study the multimedia materials were developed for the sake of experiments. But these multimedia materials were not produce coherent results with classic lessons of CTML. So, designers should aware of factors that affect extraneous, intrinsic, and generative processing could be lead different results than expectations.

The last but not least outcome of current study came from independent variables which were not included to research questions. These variables are gender and university. Traditionally, the gender and demographics of the participants were collected in studies, but they are not reported in papers in multimedia leaning. However, analysis of eye tracking and test scores showed that the results might be changed profoundly according to gender and university. In line with current study Cowards, Crooks, Flores & Dao (2012) found a significant modality by gender. The scores of females were higher for animations compared to males' better performance on static images. They argue that this is an important contribution to literature. The designers should be aware of gender differences in design issues.

The other variable that should be taken into account in design of multimedia instruction is the participants' general academic aptitudes. Although the university entrance exam results were not recorded in study, the differences between universities produced differences in both test and eye tracking results. Designers should consider the academic aptitudes of targeted learners.

5.4. Researcher "How I can Design a Better Study by Using Results of Current Study"

Multimedia learning and cognitive load attracts researchers' attention for a long time. This interest on the area resulted many studies from different domains. Recent articles on this area concentrated on online measuring of multimedia advantages and testing boundary conditions of multimedia learning. The studies conducted mostly in the frameworks of Cognitive Theory of Multimedia Learning (Mayer, 2001, 2009) and Cognitive Load Theory (Sweller, 2005).

These two theories have impact mainly on designing practice. Instructional designer follow their principles and rules to maximize learning experience. Although many studies conducted in that area, principles still boundary conditions for different groups of learners and domains.

This study integrates a new methodology with an online measuring tool into multimedia studies. The new methodology was examining of multimedia principles in the case of multiple study of same content. In this study the interaction between multiple studies of same content and principles of multimedia learning was observed by a relatively new online instrument, eye tracking. Since the methodology of the study was new and it can be easily applied into all multimedia principles. But, in this study boundary conditions of CTML preferred. These were pace and domain variability.

The design of study has several strengths with weaknesses. The repeated nature of the treatment was one of the strengths of study. It combines traditional measuring methods with new instruments. The traditional way of assessing learning was retention and transfer test, and the new instruments was online eye tracking. The use of different domains of the study was another contribution of this study leading to a more ecologically valid environment for multimedia learning. The last strength of the study was the number of participants that was comparably higher than other eye tracking studies.

Current study has some weaknesses, because it has many independent variables, which are not controlled by design. These variables were age, university and gender. These variables were included into study but not treated as factors of ANOVA design due to sampling of participants. Another weakness of the study is the number of multiple readings. All participants study lessons only twice, however the repetition could have a range other that two; such as three or four study per lesson. Additionally, the time lag between studies was another weakness of study, which was kept fixed about 15 minutes. This time lag can be changed from several hours to week. The change in the number of re-study and the varied time lag between studies could have been made the prior knowledge assumption more comprehensive. But including those variables into the study made the volume of study unmanageable.

The weaknesses of the study define limitations of the study. Possible further studies were suggested in following paragraphs in the light of results and limitations.

In this study, too many variables make it very difficult, if not impossible, to manage that. Each study has many independent variables [Gender, University, Study (First, Second, Pace, Image Type (animated, static), load (materials, experiment)] since age is eliminated from analyses. For all IV's six eye tracking metrics could be computed as dependent variable). Since there was 5 lessons and their interactions, the interpretation of the results became complicated. If those variables analyzed in one experimental design, the interactions between those variables and covariance could be observed, but the risk of type II error increased. If they analyzed independently, the readability could have been increased but in this case, the risk of type I error might be increased. This was a double edged sword.

Divide and conquer strategy could be worked well in this study. All studies could be handled as independent experiments to simplify results and to increase readability. Further studies should take one principle and one learner characteristic as an IV, and should use either retention test results or eye tracking result as DV, if they want to have concise conclusions and implication. Additionally, including many variables have been increase the confounding risk of variables in experiments.

For example, in pace study of M and MSP lesson should be simplified by using only static or animated images. And, it might be conducted on only one university. On the other hand, the speed of system paced mechanism lesson can be changed into fast-medium-slow options to create different levels of cognitive load as another experiment.

In chemistry lesson, the video-narration, video-narration-on screen text, and video-on screen text options were chosen as medium of content. However, the chemistry lesson has include different sections such as materials and parts in which chemistry experiments told. The content of the chemistry lesson could have been more unique in terms of load.

The learning experience in multimedia literature is based on retention and transfer tests mainly. Those tests are developed by researcher. Since the domain of multimedia studies diversifying, the measurement types should be diversify accordingly. In multimedia learning, the dependence on paper-pencil based tests developed in cognitive domain restricted the design issues only cognitive ability. But, in education, other than cognitive domain; motor and affective domains should be concentrated on as tribute to legendary figure Bloom (1956). Fortunately, new studies emerge on the affective components of multimedia learning, which makes the picture more complete.

Another suggestion for methodology, using best worked examples as content of lesson might be a smarter choice for new methodology in which multiple studies of same content occurs. If best worked examples were chosen as content, the methodology of the current research could have been compared with literature more strongly. To achieve this, following method can be tracked. As a first step eye tracking metrics collected as a baseline for best worked examples. Later the trends in change might be collected for eye tracking metrics in restudy sessions. With the comparisons between first, second, and may be third studies, the methodology could have been tested at the very beginning, whether it was worked or not for using in further studies. Best worked examples can be applied to observe individual differences such as gender, memory capacity, attitude etc. in multimedia also.

To sup up, the current study provides a new methodology to study multimedia learning. This methodology tested for the boundary conditions of the multimedia learning. Results showed that this methodology could be used as a tool to control prior knowledge, and to test interaction with prior knowledge and other multimedia principles. Interestingly, this method is quite sensitive to individual differences such as gender and cognitive capacities. But, further studies should be conducted to find out the interaction between prior knowledge and other principles of multimedia learning. Those studies should not include only boundary conditions but best worked examples also.

REFERENCES

Ainsworth, S. (2008). How should we evaluate multimedia learning environments? In J.-F. Rouet & R. Lowe & W. Schnotz (Eds.), *Understanding Multimedia Comprehension*. pp 249-265 New York: Springer.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.) *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 2, pp.89-195). London: Academic Press

Ayres, P. & Paas, F. (2012), Cognitive Load Theory: New Directions and Challenges. *Appl. Cognit. Psychol.*, 26: 827-832. doi: 10.1002/acp.2882

Baddeley, A. (1999). *Essentials of Human Memory*. Hove: Psychology Press.

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

Baddeley, A. D., Eysenck, M., & Anderson, M. C. (2009). *Memory*. Hove: Psychology Press.

Bakeman, R. (2005). Recommended Effect Size Statistics for Repeated Measures Designs. *Behavior Research Methods, Instruments, and Computers, 37, 379-384* .

Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance matters: Spatial contiguity effects as trade-off between gaze-switches and memory load. *Applied Cognitive Psychology.* 26: doi: 10.1002/acp.2887

Beymer, D., Orton, P.Z. & Russel, D.M. (2007). An Eye Tracking Study of How Pictures Influence Online Reading. In Baranauskas et.al. (Eds): INTERACT 2007, LNCS 4663, Part II, 456-460.

Bloom, B. (ed.) (1956). *Taxonomy of Educational Objectives, the classification of educational goals – Handbook I: Cognitive Domain.* New York: McKay

Bransford J.D. & Schwartz D.L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P.D. Pearson (Eds.) Review of research in education (Vol. 24, pp. 61-100). Washington, DC: American Educational Research Association.

Brünken, R., Plass, J.L. & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist,* 38(1), 53-61.

Cleeremans, A.& Dienes, Z.(2008).Computational Models of Implicit Learning In Sun, R. (Ed.), *The Cambridge Handbook of Computational Modeling*,Cambridge, UK: Cambridge University Press, pp. 396-421

Cohen, J. (1992) A Power Primer. Psychological Bulletin, 112(1), 155-159

Cowards, F.L.; Crooks, S.M.; Flores, R. & Dao, D. (2012). Examining the Effect of Gender and Presentation Mode on Learning from a Multimedia Presentation. *Multidisciplinary Journal of Gender Studies.* 1 (1), 4869. doi: 10.4471/generos.2012.03

Clark, J. M. & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149-170.

Crow, D. (2006). *Left to Right / the cultural shift from words to pictures*. Switzerland: AVA Publishing.

D'Amico, E. J., Neilands, T. B., & Zambarano, R. (2001). Power analysis for multivariate and repeated measures designs: a flexible approach using the SPSS MANOVA procedure. *Behavior research methods instruments computers a journal of the Psychonomic Society Inc*, *33*(4), 479-484.

Engelkamp, J. (1998). *Memory for actions*. Hove, UK: Psychology Press.

Grimley, M. (2005). Learning From Multimedia Materials: The relative impact of individual differences. *Educational Psychology*, 27(4), 467-485.

Grissom, R. J. & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York: Routledge

Halford, G.S., Baker, R., McCredden J.E., and Bain, J.D. (2005). How Many Variables Can Humans Process? *Psychological Science*, 16(1), 70-76

Hoogerheide, V., & Paas, F. (2012). Remembered utility of unpleasant and pleasant learning experiences: Is all well that ends well? *Applied Cognitive Psychology*. 26: doi: 10.1002/acp.2890

Huang W., & Eades, P. (2005). How People Read Graphs, *Asia Pacific Symposium on Information Visualization*, 45, 51-58.

IBM Corp. (2011). *IBM SPSS Statistics for Windows*, Version 20.0. [Computer Software]. Armonk, NY: IBM Corp.

Jacoby, L.L. (1991). A Process of Dissociation Framework: Seperating Automatic from Intentional Uses of Memory. Journal of Memory and Language, 30(5), 513-541

Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches (2nd Ed.)*. Boston, MA: Pearson Education Inc.

Johnson, C.I., & Mayer, R.E. (2012) An Eye Movement Analysis of the Spatial Contiguity Effect in Multimedia Learning. *Journal of Experimental Psychology: Applied*, 18(2), 178-191.

Kalyuga, S. (2005). Prior Knowledge Principle in Multimedia Learning. In R. E. Mayer, *The Cambridge Handbook of Multimedia Learning*, New York: Cambridge University Press, pp. 325-337.

Kalyuga, S. (2007). Expertise Reversal Effect and Its Implications for Learner-Toilered Instruction. *Educational Psychology Review*, 19, 509-539.

Kalyuga, S. (2009) *Managing Cognitive Load in Adaptive Multimedia Learning.* Information Science Reference. Hershey New York

Kalyuga, S. (2011) Cognitive Load Theory: How Many Types of Load It Really Need? *Educational Psychology Review,* 23, 1-19

Kennedy, G. (1983). *Invitation to Statistics*. Oxford: Basic Blackwell.

Khosrowjerdi, M., & Iranshahi, M. (2011). Prior knowledge and information-seeking behavior of PhD and MA students. *Library & Information Science Research*, 33(4), 331-335

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

Kirschner, P.A., Ayres, P., & Chandler, P. (2011) Contemporary cognitive load theory research: The good, the bad and the ugly. *Computers in Human Behavior,* 27, 99-105

Kirscher, P.A., Kester, L., & Corbalan, G. (2011) Cognitive load theory and multimedia learning, task characteristics and learning engagement: The current State of the Art. *Computers in Human Behavior,* 27, 1-4

Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2012). Looking in the same manner but seeing it differently: Bottom-up and expertise effects in radiology. *Applied Cognitive Psychology*. 26: doi: 10.1002/acp.2886

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

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

Mayer, R. E. (2009). *Multimedia Learning* (2nd ed). New York: Cambridge University Press.

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In: Mayer, R.E. (Ed.), *The Cambridge Handbook of Multimedia Learning*, New York: Cambridge University Press, pp. 31-48.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167-171.

Mayer, R. & Moreno R (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43-52.

Miller, G.A. (1956). The Magic Number Seven, plus or Minus Two: Some limits on our Capacity for Processing Information. *Psychological Review*, 63, 81-97.

Moreno, R. (2006a). Learning in High-Tech and Multimedia Environments. *Current Directions in Psychological Science*, 15(2), 62-67.

Moreno, R. (2006b). 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. (2010). Techniques that Increase Generative Processing in Multimedia Learning: Open Questions for Cognitive-Load Research. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory*, Ch. 8. New York: Cambridge University Press pp. 153-177.

Moreno, R. & Park, B. (2010) Cognitive Load Theory: Historical Development and Relation to Other Theories.. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory*, Ch. 1. New York: Cambridge University Press pp. 9-28.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.).Mahwah, NJ: Lawrence Erlbaum Associates

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra- word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9, 329–389.

Nathan, M. J., & Resnick, L. B. (1994). Less can be more: Unintelligent tutoring based on psychological theories and experimentation. In S. Vosniadou, E. De Corte, & H. Mandl (Eds.), *Technology-based learning environments* (pp. 183–192). Berlin: Springer-Verlag.

Ozcelik, E., Arı-Arslan, I, Cagiltay, K. (2010) *Why does signaling enhance multimedia learning? Evidence from eye movements*. Computers in Human Behavior, 26, 100-117.doi.org/10.1016/j.chb.2009.09.001,

Paivio, A. (1969).Mental imagery in associative learning and memory. *Psychological Review*, 76, 241–263.

Paivio, A. (1975). Coding distinctions and repetition effects in memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 179–214). Orlando, FL: Academic.

Plass, J.L., Moreno, R., & Brünken, R. (2010) *Cognitive Load Theory*, New York: Cambridge University Press

Peterson, C. (2009). Minimally sufficient research. *Perspectives on Psychological Science*, 4, 7–9.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, 64, 916-924

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.

Rayner, K., Chace, K.H., Slattery, T.J, & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading*, 10(3), 241-255.

Reed, K. S. (2006). Cognitive Architectures for Multimedia Learning. *Educational Psychologist*, *41* (2), 87-98.

Roodenrys, K., Agostinho, S., Roodenrys, S., & Chandler, P. (2012). Managing one's own cognitive load when evidence of split attention is present. Applied Cognitive Psychology. 26: doi: 10.1002/acp.2889

Sawicka, A. (2008). Dynamics of cognitive load theory: A model-base approach. *Computers in Human Behavior*, 24, 1041-1066.

Schnitzer, B.S., & Kowler, E. (2006). Eye movements during multiple readings of the same text. *Vision Research*, 46, 1611-1632.

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, 34(5), 1264-1288. doi:10.1037/a0012943

Schnotz, W. (2005). An Integrated Model of Text and Picture Comprehension In R. E. Mayer, *The Cambridge Handbook of Multimedia Learning*, New York, Cambridge University Press, pp. 49-69.

Schnotz, W. & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, 19, 469-08.

Shipstead, Z., Redick, T.S.; & Engle, R.W. (2010) Does working memory training generalize?, Psychologica Belgica, 50 (3-4), 245-276

Singh, A., Marcus, N., & Ayres, P. (2012). The transient information effect: Investigating the impact of segmentation on spoken and written text. *Applied Cognitive Psychology*. 26: doi: 10.1002/acp.2885.

Skuballa, I. T., Schwonke, R., & Renkl, A. (2012). Learning from narrated animations with different support procedures: Working memory capacity matters. *Applied Cognitive Psychology*. 26: doi: 10.1002/acp.2884

Slykhuis, D.A., Wiebe, E.N., & Anetta, L.A. (2005). Eye tracking students' attention to PowerPoint Photographs in a science education setting. *Journal of Science Education and Technology*, 14(5/6), 509-520.

Sorden, S.D. (2005). A Cognitive Approach to Instructional Design for Multimedia Learning. *Informing Science Journal*, 8, 263-279.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12 (1), 257-285.

Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer, *The Cambridge Handbook of Multimedia Learning*, New York, Cambridge University Press, pp. 19-30.

Sweller, J, (2010). Cognitive load theory: Recent theoretical advances. In Plass, J., Moreno, R, & Brunken R (Eds.), *Cognitive Load Theory*, Ed. Ch 1., New York, Cambridge University Press pp. 29 – 47

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

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

Tabachnick, B. G. & Fidell, L. S. (1996). *Using Multivariate Statistics (3rd ed.)*. New York: Harper Collins College Publishers.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.).* Boston, Massachusetts: Allyn and Bacon.

Tabbers, H. K., Paas, F., Lankford, C., Martens, R. L., & Van Merriënboer, J. J. G. (2008). Studying eye movements in multimedia learning. In J.-F. Rouet, R. Lowe, & W. Schnotz (Eds.), Understanding multimedia documents (pp. 169-184). New York: Springer.

van Gog, T., Kirschner, F., Kester, L. & Paas, F. (2012), Timing and Frequency of Mental Effort Measurement: Evidence in Favour of Repeated Measures. *Applied Cognitive Psychology*, 26: 833–839. doi: 10.1002/acp.2883

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review,* 17*, 147–177*

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *The American Psychologist*, 54 (8), 594-60

Yecan, E., Sumuer, E., Baran, B. & Cagiltay, K. (2007). Tracing users' behaviors in a multimodal instructional materials: An eye-tracking study. In J.Jacko (Ed) HCI, Part III, LNCS 4552, pp. 755-762

Youssef, A., Ayres, P., & Sweller, J. (2012). Using general problem solving strategies to generate ideas in order to solve geography problems. *Applied Cognitive Psychology.* 26: doi: 10.1002/acp.2888

APPENDIX A

INFORMED CONSENT (IN TURKISH)

Bu araştırma ODTÜ Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü doktora öğrencisi tarafından yapılmaktadır. Araştırmanın amacı genel olarak çoklu öğretim ortamlarının öğrenme ile ilişkisini gözlemlemektir. Araştırmada elde edilen verilerin tamamı araştırma amaçlı toplanmakta olup, bireysel veriler gizli tutulacaktır. Sonuçlar ortalamalar ve dağılımlar olarak verilecek, katılımcılar arasında başarı sıralamaları yapılmayacaktır.

Araştırmaya katılım tamamen gönüllüdür. Araştırmanın herhangi bir aşamasında, devam etmek istemediğinizi belirtip, ayrılabilirsiniz.

Katılımınız için şimdiden teşekkür ederim.

Serkan Alkan

Kişisel Bilgiler:

Yaş :_________________________________

Bölüm ve Sınıf :_________________________________

Cinsiyet :________________________________

Arastirma sonuclari ile ilgili geribildirim istiyor musunuz? Evet ( ) Hayır ( )

e-posta :_________________________________

Katılımcı Kodu:

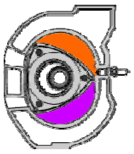
Katıldığı deneyler:

APPENDIX B

ACHIEVEMENT TEST (IN TURKISH)

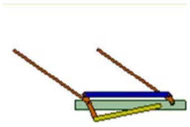
Lütfen aşağıdaki soruları cevaplayınız.

1) Aşağıdaki verilen mekanizma örneklerinin yanına isimlerini yazınız.

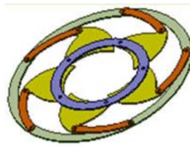


____

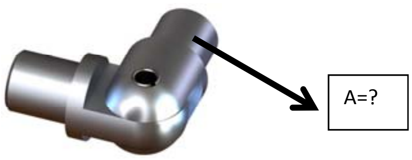
____



____



____



2) Yandaki şekilde (**A**) nedir?
   A) Mafsal
   B) Kinematik eleman
   C) Kinematik çift
   D) Mekanizma
   E) Makina

3) Makina ve mekanizma ile ilgili olarak aşağıdakilerden hangisi doğrudur?
   A) Makina daha geneldir ve farklı kullanım amaçları vardır.
   B) Mekanizma belirli bir amaç için üretilmiştir.
   C) Her mekanizma bir makina için özel olarak tasarlanmıştır.
   D) Makina ve mekanizma terimleri birbirlerini kapsar.
   E) Makinalarda birden fazla mekanizma olabilir.

____A_____, kendisini inceleyerek ____B___ yapısını analiz ve sentez edebileceğimiz bir ____C___ bir sistemi; ___B_____ ise ____D_____ bir sistemi ifade eder.

4) Yukarıdaki cümlede boşluklara gelmesi gereken seçeneklerin karşısına uygun harfleri yazınız.

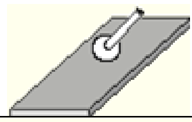
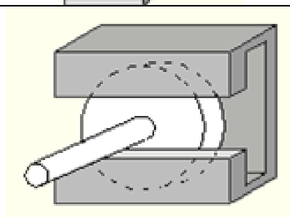
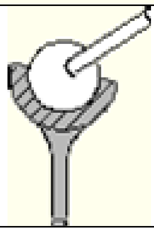
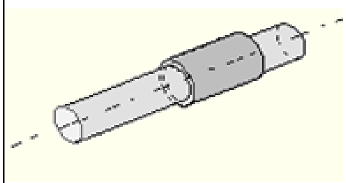
gerçek (reel)	( )
ideal	( )
mekanizma	( )
makina	( )

Aşağıda verilen boşluklara uygun kelimeleri yazınız.

______ kinematik çiftlerde, iki kinematik eleman arasında temas, mekanizmanın tüm hareketi süresince mevcuttur. ________kinematik çiftlerde, kinematik elemanlar hareketin tümü boyunca temas etmeyebilirler ve bu temas kontrol edilebilir.

Kapalı kinematik çiftler ayrıca temas şekline göre sınıflandırılabilir. ______ kinematik çiftlerde kinematik elemanlar bir yüzey boyunca temas ederler. ______ kinematik çiftlerde ise temas, geometrik olarak bir nokta veya bir çizgi üzerindedir.

Aşağıdaki kinematik çiftlerin karşısındaki hücrelerdeki uygun olan tanımlara (X) işareti koyunuz.

	Açık Kinematik Çift	Kapalı Kinematik Çift	Kuvvet Kapalı Kinematik Çift	Şekil Kapalı Kinematik Çift	Basit Kinematik Çift	Yüksek Kinematik Çift
						
						
						
						
						

Lütfen izlediğiniz ders ile ilgili aşağıdaki soruları cevaplayınız.

Asağıdaki malzemelerin isimlerini yanlarına yazınız

 ___________________

 ____________________

 _______________________

____

 _______________________

 ____________________

Derste yer alan deneylerin isimlerini yazınız.

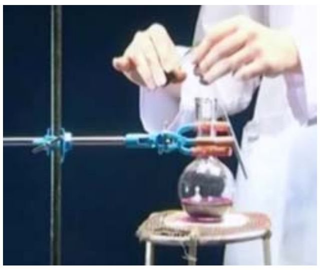
1.______________________________________
2.______________________________________
3.______________________________________

Aşağıdaki ekran görüntülerinin yanına derste yer alıyorlarsa (evet) yer almıyorsa (hayır) işaretleyiniz

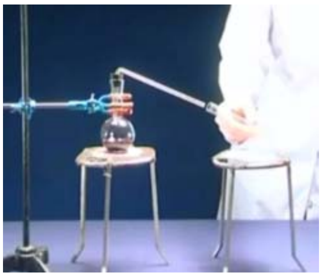
 Evet ( ) Hayır ( )

 Evet ( ) Hayır ( )

Evet ( ) Hayır ( )

 Evet ( ) Hayır ( )

Aşağıdaki ekran görüntülerinin yanına derste anlatılan cümleyi yazınız.



APPENDIX C

PARAMETERS RECORDED BY EYE TRACKER

1. Timestamp	22. Event
2. DateTimeStamp	23. EventKey
3. DateTimeStampStartOffset	24. Data1
4. Number	25. Data2
5. GazePointXLeft	26. Descriptor
6. GazePointYLeft	27. StimuliName
7. CamXLeft	28. StimuliID
8. CamYLeft	29. MediaWidth
9. DistanceLeft	30. MediaHeight
10. PupilLeft	31. MediaPosX
11. ValidityLeft	32. MediaPosY
12. GazePointXRight	33. MappedFixationPointX
13. GazePointYRight	34. MappedFixationPointY
14. CamXRight	35. FixationDuration
15. CamYRight	36. AoiIds
16. DistanceRight	37. AoiNames
17. PupilRight	38. WebGroupImage
18. ValidityRight	39. MappedGazeDataPointX
19. FixationIndex	40. MappedGazeDataPointY
20. GazePointX	41. MicroSecondTimestamp
21. GazePointY	42. AbsoluteMicroSecondTime-stamp

APPENDIX D

METRICS CALCULATED BY TOBII STUDIO

	N	Mean	Max	Min	Sum	Std Dev
Time to First Fixation						
Fixations Before						
First Fixation Duration		√	√	√		√
Fixation Duration	√#	√	√	√	√*	√
Total Fixation Duration		√*	√	√	√	√
Total Fixation Duration (Include Zeros)						
Fixation Count		√	√	√	√#	√
Fixation Count (Include Zeros)						
Visit Duration	√-	√	√	√	√+	√
Total Visit Duration		√+				√
Total Visit Duration (Include Zeros)						
Visit Count		√-	√	√		√
Visit Count (Include Zeros)						
Percentage Fixated						
Percentage Clicked						
Time to First Mouse Click						
Time to First Mouse Click (Across Media)						
Time from First Fixation to Next Mouse Click						
Time from First Fixation to Next Mouse Click (Across Media)						
Mouse Click Count						
Mouse Click Count (Include Zeros)						
***,+,-,#, and – refers same values with different denotation**						

CURRICULUM VITAE

Personal Information

Surname, Name:	Alkan Serkan
Nationality:	Turkish (TC)
Date and Place of Birth:	13 December 1969, Dereköy
Phone:	+ 90 312 210 4125
E-mail:	serkana@metu.edu.tr

Education

Degree	Institution	Year of Graduation
MSc	Middle East Technical University, Department of Computer Education and Instructional Technology	2006
BSc	Middle East Technical University, Department of Pyschology	1994
High School	Izmit High School	1986

Work Experience

Year	Place	Enrollment
2010-	Middle East Technical University	Psychologist
2000-2010	Middle East Technical University	Research Assistant
1996-2000	Ileri Education and Publication Inc.Co.	Specialist, Administrator
1994-1996	Kocaeli University	Instructor

Foreign Languages

English, Advanced

Publications

1. **Serkan Alkan**, Kürşat Çağıltay, Studying computer game learning experience through eye tracking, British Journal of Educational Technology, 38(3), pp 538-542, 2007
2. Karakaş, Erdoğan, E., Sak, L., Soysal, A.Ş., Ulusoy, T., Yüceyurt Ulusoy, İ., Alkan, S. (1999). Standardization of Stroop Test – TBAG Form in Turkish Culture; reliability, validity. Klinik Psikiyatri Dergisi. 2(2), 75-88.