AN EXPERIMENTAL INVESTIGATION OF ICT EXPERTISE BY GAZE ANALYSIS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

ERDİ KAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

JANUARY 2019

AN EXPERIMENTAL INVESTIGATION OF ICT EXPERTISE BY GAZE ANALYSIS

Submitted by **ERDİ KAYA** in partial fulfillment of the requirements for the degree of **Master of Science in The Department of Cognitive Science, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics**

Prof. Dr. Cem Bozşahin
Head of Department, **Cognitive Science**

Assoc. Prof. Dr. Cengiz Acartürk
Supervisor, **Cognitive Science Dept., METU**

Examining Committee Members:

Prof. Dr. Cem Bozşahin
Cognitive Science Dept., METU

Assoc. Prof. Dr. Cengiz Acartürk
Cognitive Science Dept., METU

Asst. Prof. Dr. Murat Perit Çakır
Cognitive Science Dept., METU

Asst. Prof. Dr. Umut Özge
Cognitive Science Dept., METU

Asst. Prof. Dr. Murat Ulubay
Business Dept., Yıldırım Beyazıt University

Date: _*30.01.2019*

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 : **Erdi Kaya**

Signature : ____________________

ABSTRACT

AN EXPERIMENTAL INVESTIGATION OF ICT EXPERTISE BY GAZE ANALYSIS

Kaya, Erdi

MSc., Department of Cognitive Sciences

Supervisor: Assoc. Prof. Dr. Cengiz Acartürk

January 2019, 40 pages

The aim of the present study is to contribute to the approaches used for the division of experts and novices according to the approaches and differences in question answering. For this purpose, an experiment was conducted with 41 participants. The subjects were divided into two groups as experts and novices according to their information test scores. Then the participants answered a set of questions based on ICT and their eye movements were recorded. The participants were then divided as high performers and low performers according to the experimental task scores. The results showed that the information test on paper was not successful to divide the participants as experts and novices. There was a significant difference between high performers and low performers in terms of gazing at the correct answers compared to incorrect ones. On the other hand, no significant differences were obtained in other gaze measures. The conclusion is that the investigation of the differences between high and low performers by eye tracking is subject to a limited set of measures.

Keywords: Information and Communication Technologies (ICT), human-computer interaction, expertise, eye-tracking

ÖZ

BİLGİ VE İLETİŞİM TEKNOLOJİLERİ UZMANLIĞININ GÖZ İZLEME YOLUYLA DENEYSEL BİR ARAŞTIRMASI

Kaya, Erdi
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Doç. Dr. Cengiz Acartürk

Ocak 2019, 40 sayfa

Bu çalışmanın amacı, soru cevaplamadaki yaklaşım ve farklılıklara göre uzman ve acemilerin belirlenmesi için kullanılan yaklaşımlara katkıda bulunmaktır. Bu amaçla 41 katılımcı ile bir deney yapılmıştır. Bilgi testi puanlarına göre denekler uzman ve acemi olarak iki gruba ayrıldı. Daha sonra katılımcılar bilgi iletişim teknolojileri ile ilgili sorulara cevap verdiler ve deneklerin göz hareketleri kaydedildi. Deneysel görev puanlarına göre denekler yüksek performanslı ve düşük performanslı olarak ikiye bölündü. Analiz sonuçları, kağıt üzerindeki bilgi testinin katılımcıları uzman ve acemi olarak bölme konusunda başarılı olamadığını göstermiştir. Yüksek performans gösterenler ile düşük performans gösterenler arasında, uyaranların gösterildiği tüm ekranlara kıyasla doğru cevaba bakma açısından önemli bir fark vardı. Öte yandan, diğer bakış ölçümlerinde anlamlı bir farklılık gözlenmedi. Yüksek ve düşük performanslılar arasındaki farkların göz takibi ile araştırılmasının sınırlı bir takım ölçümlere bağlı olduğu sonucuna varılmıştır.

Anahtar Sözcükler: Bilişim ve İletişim Teknolojileri (BİT), insan-bilgisayar etkileşimi, uzmanlık, göz izleme

To My Family…

ACKNOWLEDGMENTS

First of all, I owe a debt of gratitude to my supervisor Cengiz Acartürk for his support, encouragement, insights at every stage of my thesis and endless guidance and advices throughout the present thesis.

Besides my supervisor, I would like to express my sincere gratitude to Murat Perit Çakır for his assistance.

I would like to thank Mehmetcan Fal for his support and assistance in developing the C# application which is used for the eye-tracking experiment.

I would like to thank my friend Ayşenur Hülagü for her supports, assistance and conversations during coffee breaks at the library.

Also, I would like to thank Sibel Gülnar for her assistance in the Graduate School of Informatics.

I would like to thank all my friends, instructors from my graduate institute for helps with data gathering.

Finally, I would like to thank my family for supporting me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SA	Situational Awareness
ICT	Information and Communication Technologies
ANOVA	Analysis of Variance
px	pixel
pc	Personal Computer
API	Application Programming Interface
STM	Short-Term Memory
LTM	Long-Term Memory

CHAPTER 1

INTRODUCTION

1.1. Purpose of the Study

Individual differences between users have attracted attention in the research literature, within the framework of several dimensions, such as expertise. The issue of individual differences can be addressed as random factors by means of statistical sampling. On the other hand, users can be divided as experts and novices according to the domain expertise depending on the specific research field (Chi, Feltovich, & Glaser, 1981). In brief, the approaches to the division between different participant groups, such as experts and novices or high performers and low performers will be the major issue of this study.

Expertise can be expressed as the whole of knowledge and skills which recognize the differences between experts and novices and it belongs to a certain domain. Experts generally perform well in the representative tasks of a domain. At this point, a major question is how experts organize their knowledge and skills and how individuals' performance in a particular domain differ from each other.

In addition, a subtopic can be investigated as domain expertise in narrow scope, since different domains require different representations and categorizations of problems. The importance of information and communication technologies (*ICT*), which can be classified as a subdomain, have started to arise with the increase in the use of computer technologies in our daily lives. This study aims at contributing to the approaches on the division between experts and novices or the division between high performers and low performers based on expertise in ICT. In accordance with this purpose, we claim that the analysis of gaze measures during the process of question answering, in addition to test-based knowledge assessment can be used for the division of experts-novices or high-low performers. In accordance with this purpose, we investigated whether there are gaze indicators that are able to complement information test scores in ICT expertise. At this point, in order to determine these findings, an information pretest was conducted to divide a group of participants into two groups (experts and novices) according to their expertise in ICT. We then conducted an experimental task to divide a group of participants into two groups (high performers and low performers). During this experimental task, which is reported as the second experiment session in the chapters below, the participants answered a set of ICT questions while their eye movements were recorded.

1.2. Scope of the Study

Several research questions may be constituted about expected patterns of eye movements and performance of the participants. We used eye movement measures as dependent variables in so far as gaze data have an important role as an indicator of human information processing (i.e., visual extraction of information from a display). The performance of the participants was also measured in terms of the correct answers of the information pretest on paper and the experimental task where eye movements were recorded. Both of these tests had questions with network diagrams as subtopics in ICT. Accordingly, the participants were presented a two-step experiment. In the first step, before the eye tracking session that was executed on computer, the participants filled in a pretest on paper to determine their expertise in ICT. Then, in the second step, the participants were presented another test where their eye movements were recorded. The following research questions were devised:

- Is it possible to use pretest scoring for novice-expert discrimination in ICT to predict task performance in ICT question answering?

- Which gaze measures, such as gaze shift between questions and answers or gaze duration, are able to align with task performance accuracy?

In the direction of these main questions, we formulated two hypotheses. The first hypothesis is that a pretest-based division between the participants as experts and novices will align with task performance in ICT question answering (H1). The second hypothesis is that there is a relation between high performers and low performers (as identified by task performance) and their eye movements during the course of their answering ICT questions (H2). For instance, one may expect that the gazing rates at correct answers compared to all choices for the high performer participants may be different from the low performer participants, given the previous findings in the literature. The accuracy of the answers may also influence this relationship. The following section presents the outline of the thesis.

1.3. Thesis Outline

The follow-up chapters of this thesis was put in order as below. In Chapter 2, a literature review with annotations was given in line with four fields that are from the viewpoint of these main questions of this thesis: expertise, approaches from cognitive perspectives, knowledge assessment, and eye tracking. Experimental results and studies from these research fields were presented allied with conceivable relations between these fields. In Chapter 3, information about the experiment carried out as a part of this thesis was clarified. Then, in Chapter 4, the results of this experiment were analyzed. In the final chapter (Chapter 5), the results of the experiment of this present study were discussed and investigated whether our hypothesis was supported.

CHAPTER 2

LITERATURE REVIEW

In this chapter, a literature review is given about the topics connected with this present study. Firstly, studies related to expertise will be put across in this chapter. The literature review about expertise will generally focus on definition of expertise and characteristics. This topic is also related to theories and methods in relation to expertise. Therefore, the studies on this topic will be also showed. Thirdly, knowledge assessment which are the main theoretical background of this thesis will be investigated in terms of the structure of expertise. And finally, eye tracking that is used to collect data during the experiment process in the context of this thesis will be shown.

2.1. Expertise

Competent and knowledgeable people who perform consistently superior performance in representative tasks in a specific area are identified as experts. At this point, it may be useful to emphasize the reasons of expert performance. According to a widely accepted theoretical concept, innate mental capacities were the cause of superior performance in most areas of expertise (Galton, 1869). On the other hand, it has been observed that the measurement of basic mental capacities is not sufficient to examine the acquisition of expert performance in a field and the differences between the experts and others are mostly seen in the skills acquired by the experts (Ericsson & Lehmann, 1996). According to the same study, the superior performance of the experts is generally limited to the relevant domain.

Also, there are studies focusing on the content and organization of experts' knowledge. One of them has investigated categorization and representation of physics problems (Chi, Feltovich, & Glaser, 1981). For the related study, four experiments have put through with regards to existence of problem categories, differences in the categories used by experts and novices, differences in the knowledge in relation to the categories and features in problems. According to this study, experts and novices start their problem representations from different problem categories. On the other hand, experts can abstract physics principles to solve a problem while novices ground on their approaches and representations on the problem’s basic features. In another study, most of the information representations of adult experts were also observed in children (Chi & Koeske, 1983). In another study, in order to demonstrate the performance of the experts, the knowledge of experts was used in the process of building computer-based models (Hoffman, 1992).

Several different approaches are presented with the studies on defining the concept of expertise. In one of them, groups were formed according to their level of expertise (Hoffman, 1998). This classification is given as the following (Hoffman, Trafton, & Roebber, 2005):

- *Naïve* is someone who doesn't have any information about a domain.
- *Novice* is someone who has some minimal information about a domain.
- *Initiate* is novice who has more information than novices about a domain.
- *Apprentice* is someone who defined as a student.
- *Journeyman* is an experienced someone about a domain.
- *Expert* is brilliant journeyman who has special skills or knowledge.
- *Master* is someone who is from an elite group of experts.

Two approaches have been defined in the studies on the features of the experts: absolute approach and relative approach (Chi, 2006b). This assumption, which claims that excellence and creativity develop through innate talent and luck, is defined as absolute expertise in psychology. In the relative approach, expertise is defined as a level of mastery that novices can obtain. In the second approach, a more knowledgeable group can be defined as an expert, while a less knowledgeable group can be defined as a novice. Therefore, the novice concept can be expanded to mean non-expert persons. In other words, the novice concept can be used to define a large range from groups given in Table 1 such as from naives to journeymen. The "relative expertise" concept which experts are defined according to novices is adopted in the second approach. The purpose of relative expertise is to understand how people with less knowledge and skills become experts. In the present study, the second approach has been referenced.

This approach which experts are defined as more knowledgeable than non-experts has some basic theoretical assumptions. The first assumption is that experts are more knowledgeable in a certain domain (Bedard & Chi, 1992). In the second assumption, experts and non-experts have almost equivalent fundamental capacities and equivalent abilities to reason. According to the third assumption, the differences between experts and others are determined by the representation ways of their knowledge.

As a result of the studies, it was observed that the experts performed better than the novices in some subjects. For instance, experts saw some important shapes in X-ray films that novices couldn't see (Lesgold et al., 1988). In another study, it was observed that the experts comprehend the knowledge and strategies specific to the field related to minimal cognitive effort (Alexander, 2003).

There are also studies where experts cannot perform well (Sternberg, 1996). Expertise is limited to the relevant domain. In other words, experts cannot excel in domains where they do not have any expertise. In a study, it was observed that the chess master performed worse at remembering the chessboard positions placed randomly compared to remembering the real chessboard positions in chess games (Gobet & Simon, 1996).

There have also been studies that are made incorrect estimates by experts. In a study on weather forecasts, it was observed that meteorology experts were conservative (Robert R. Hoffman et al., 2005).

2.2. Approaches From Cognitive Perspectives

This section will focus on studies in related to expertise in cognitive sciences. Studies in the field of cognitive sciences have been based on three main sources: artificial intelligence, psychology and education (Feltovich, Prietula, & Ericsson, 2006). With the development of the early computer models, computers were enabled to support intelligent behavior and this allowed artificial intelligence to emerge as models of human cognition (Newell & Simon, 1973).

Studies on expertise in educational psychology have led to a new role of expertise. At this point, the cognition of experts was considered as the target stage for education. On the other hand, novice cognition was considered as an initial stage for the educational process.

Domain specific skills and knowledge were observed to include strong links with basic cognitive abilities (Newell & Simon, 1972). In another study, it was seen that the knowledge gained over time in a certain area was associated with changes in cognitive processing (Chi, 1978). In another study, it was observed that people with high knowledge showed higher performance in learning than low knowledge people (Chiesi, Spilich, & Voss, 1979). While the performance of the experts in the relevant field of expertise is explained by psychometric factors such as logical inference and memory, the performance of the novices is explained by the general ability factors (Horn & Masunaga, 2006).

One of the most known features in the field of expertise is that expertise combines larger and more cognitive units. Through practical and experience, the knowledge of the relevant domain begins to be organized cognitively by the experts in larger pieces (Glaser & Chi, 1988). In this study, it was seen that the representation of the problem related to the related field is made by experts deeper than the novices. Briefly, keeping the information of an domain as more complex allows for faster access to this information. By this means, experts are gaining skills for related tasks.

On the other hand, when it comes to expertise, the physical performance comes to mind. However, state awareness also has an important position in the concept of expertise.

Situation awareness (SA) is the perception of environmental elements and the comprehension of these elements and future estimations (Endsley, 1995). In other words, SA consists of three levels: perception, comprehension and projection. In the perception phase, novices may have serious difficulties in what information is important. In the comprehension phase, the information is retained, combined and interpreted. In the projection phase, humans make predictions about future situations. In other words, the third phase allows decision making. To make effective decisions, higher levels of SA are used and therefore SA affects the performance of experts. In SA, while novices experienced serious difficulties in recognizing and receiving key information and integrating this information, experts were observed to continue these stages with better performance (Endsley, 2006).

In another study, self-explanations have been investigated on students which have been categorized as successful and unsuccessful (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). At this point, the expression “good” represents successful students while the expression “poor” represents unsuccessful students. According to the study, successful students have generated many explanations and connected these to principles in the text while unsuccessful students have not been generated satisfactory self-explanations. In other words, this study has analyzed the role of self-explanations in terms of question-answering.

In another study, the effect of text and images on learning has been investigated (Leopold, Doerner, Leutner, & Dutke, 2014). Two experiments were carried out in this context. In the first experiment, four groups were used, the control group containing only the text, a group in which the illustrated text was given without any strategic instructions, the other group where the subjects identified and noted important concepts, and the last group that the subjects identified important concepts and took notes with the pictures. It has been observed that the groups using the pictures showed better learning performance than the groups where the pictures were not used.

One of Hegarty’s studies has investigated the reasons of individual differences in understanding mechanical systems (Hegarty, Just, & Morrison, 1988). According to the mechanical ability tests, three reasons have been put forward as the causes of individual differences in the sense of performance: the ability to accurately identify, the ability to use rules, the ability to combine information of two or more characteristics. In other words, this study supports that expertise combines larger and more cognitive units.

2.3. Knowledge Assessment

In this section, the main subject of this thesis, information evaluation process will be discussed. At this point, to understand how experts behave and why they are more talented than novices, it is necessary to understand the way in which information is represented. For this purpose, this section will discuss the methods used by the experts and novices to shed light on the process of knowledge assessment.

At this point, a representation can be defined as a model of the research performance of the person trying to solve a particular problem (Newell & Simon, 1972). A problem representation has two stages: an understanding stage and a search stage. In the understanding phase, the initial situation, the intended situation and the constraints are represented. In the search phase, the search path for the relevant problem is gradually represented. At this point it should be noted that the differences between the experts and the novices are obtained by comparing the differences at the search stage that is the second stage of a problem representation. In a study, the representation of experts was shaped through a progressive forward-working search, from the initial state to the final desired state (Simon & Simon, 1978). On the other hand, the representation of the novices was created with a backward-working search that proceeded from the last state to the first state. The key difference observed during the search phase where experts and novices differ in problem is the representation of the domain knowledge (de Groot, 1966).

The structure of the expertise can be revealed in two general ways: intrinsic and contrived tasks (Robert R. Hoffman, 1987). In the first method, it is observed how the experts perform in the tasks that are in the domain of expertise. In the second method, it is observed how the experts perform for different tasks from the related domain. For instance, the primary task of a chess master is to make the best move. At this point, the task that is based on detection the best move is defined as the intrinsic task. On the other hand, the task of remembering the positions of the chess pieces is defined as the contrived task (Chi, 2006a).

Contrived tasks can help identifying weaknesses of experts. In order to observe the differences in representation, four different contrived tasks are used in laboratory studies: recall, perception, categorization and verbal reporting (Chi, 2006a). Experts remember a property of their domain of expertise faster and more complete than novices. Experts' superiority in remembering can be seen in both visual and oral tasks. In an experiment based on the diagnosis of X-ray films and diagnosis of the disease, four radiologists with experience of 10 years or more were selected as experts and eight radiologists with experiences of one to four years were selected as novice (Lesgold et al., 1988). In this study, while verbal reporting of the disease diagnosis was determined as an intrinsic task, the contrived task was determined to draw the shapes of the region which were thought to be problematic areas on the films. According to these results, it was seen that the experts noticed the critical regions better than the novices. The third contrived task is the categorization stage which experts and novices try to sort samples in terms of category. Hence, the hierarchical structure of information representations can be analyzed. Three techniques are used in the oral reporting task which is the last of the contrived tasks: loud notification, answering interview questions (Cooke, 1994) and explanations (Chi, 1997).

2.4. Eye Tracking

The main tool that this thesis uses during the process of knowledge assessment is eye movements and there are several studies in the field of expertise that use eye movements. Knowledge is considered to have a key importance in the acquisition of expertise.

In 1968, the first study in which eye movements were used in the field of chess (Jongman, 1968). In this study, the differences between perceptual skills were studied by examining the eye movements of experts and non-experts. Then, the results in this study were re-analyzed in another study (de Groot & Gobet, 1996). These studies have shown that the distances between eye-fixations of less-expert players are less than those of expert players. This difference showed that the experts encode the fixations more widely. At the same time, it was observed that experts performed shorter-time eye-fixations. Shorter-term eye-fixation means faster encoding. In another study, it was observed that more successful chess players performed eye-fixations between larger areas than other players to determine changes in chess positions (Reingold, Charness, Pomplun, & Stampe, 2001). In another study, it was observed that expert chess players made simple decisions with fewer eye-fixations than other chess players (Fisk & Lloyd, 1988). The use of larger visual areas between eye-fixations is effective in these results.

In another study, it was observed that chess relations were formed in parallel by experts (Reingold, Charness, Schultetus, & Stampe, 2001). On the other hand, it was observed that less specialized players obtained the same relationships in series.

In this context, studies were conducted to investigate how the information in memory was used while the skills in the field of chess were acquired (Chase & Simon, 1973a), (Chase & Simon, 1973b). In these studies, eye movements were used to determine the chess information blocks. For this purpose, subjects' eye movements on the chessboard were analyzed. Then the subjects were given the task of recall to compare these results. Similar studies have continued (Gobet & Simon, 1998), (Gobet & Clarkson, 2004). According to these studies, expertise skills are not based on differences in short-term memory (STM) capacity, but on the number of chunks kept in long-term memory (LTM).

To summarize these studies on chess, expert chess players keep chess models in long-term memory, which provides a larger visual field during the coding of these chess positions (Ericsson & Kintsch, 1995). In this way, expert chess players can code the relevant chess information faster and more precisely than non-expert players.

There are other studies using eye movements. In two of these studies, the future estimations of the expert football players was analyzed by looking at the eye movement behaviors (Williams, Davids, Burwitz, & Williams, 1994), (Williams & Davids, 1998). In these studies, the process of being predicted the opponent's future movement by defense footballers was analyzed. It was concluded that eye movement data and expert research strategies were task oriented in this study. Experienced players have been observed to move faster than the less experienced players.

There are also studies on eye movements in the sports field. In these studies, the differences between the experienced athletes and the less experienced athletes in terms of eye movements for perceptual-cognitive were investigated.

In a follow-up study, the comprehension skills of humans who read the explanations of the pulley systems and the eye-fixations were examined in this process (Hegarty & Just, 1993). Data on the comprehension skills of the readers showed that the comprehension skill depends on the language of explanation and the person's capacity. In this study, the inputs of comprehension process are explanatory texts, pictures and previous information of the subject, while the result is the mental model. The use of pictures as well as explanatory texts facilitated the process of understanding how the pulley systems are moving. Eye-fixation data, on the other hand, showed that subjects could put together the information in the text and pictures in terms of both single and multi-reel pieces. In this process, it was observed that the readers generally read the information about these parts over and over again before forming the spatial mental model of the parts of the reel system given in the pictures. The picture controls of the subjects ranged from 2 to 3 parts to local inspections to global inspections that deal with more parts.

In another study which is related to multimedia learning, the use of eye tracking for cognitive processes in multimedia learning has been focused on (Hyönä, 2010). For this purpose, it has been tried to obtain information about learning time through global eye movement measures via important event moments in an animation.

There is also a study focuses on visual saliency (Borji, Sihite, & Itti, 2013). In this study, visual attention modeling has been tried to be created through eye movement datasets. Computational complexity analysis was performed on this dataset in order to create this model. It has been stated that a special visual attention model can be worked on for each person in the section of the article. Moreover, in this section, it has been stated that better performance can be achieved by combining different attention models (saliency models).

There are studies investigating expertise in many different fields with eye tracking. In a study investigating the expertise in the artistic field, an experiment was conducted based on eye movements which consisted of two parts on nine artists as experts and nine people who had never received art education as novices (Vogt & Magnussen, 2007). Within the scope of this experiment, participants looked at 16 pictures from different categories which are from everyday scenes to abstract ones. While 12 of these pictures contained recognizable objects, the remaining 4 were completely abstract. According to the results, the participants, who were artists, looked at the abstract paintings for a longer time, whereas the participants who had no artistic education looked at the other 12 paintings for a longer time. Participants were subjected to a verbal test for memory control. There was no important difference between the groups in terms of the number of pictures remembered. However, in terms of the number of correctly remembered features in the pictures, the painter group performed better than the novice group. In addition, no difference was seen between the groups in terms of fixation times. On the other hand, the novice group again and again looked at the related pictures within the less and longer fixation values compared to the painter group.

In another study, the effects of expertise on the representative and abstract paintings were investigated (Pihko et al., 2011). In this study, 20 experts from the field of art history and 20 novices analyzed a series of paintings from five different categories from representative to abstract. The eye movements of the participants were recorded for subjective aesthetic judgments and emotional evaluations. According to the results, the emotional and aesthetic levels of the novices were observed as the highest in the representative paintings and the least in the abstract paintings. At the abstraction level, the experts behaved independently. By the increase of the degree of abstraction, the number of eye-fixations and length of scan paths for both groups increased and eye fixation times decreased. At the same time, it was seen that experts and novices gave their attention to different fields on the paintings.

In a study based on the idea that tasks involving dynamic visual parts require perceptual skills, the differences that can be seen between experts and novices in the perception and interpretation of complex and dynamic visual triggers during a task process were investigated (Jarodzka, Scheiter, Gerjets, & van Gog, 2010). In the experiment performed with seven experts and 14 novices, eye movements and verbal report data were obtained from the participants. According to the results obtained, the experts were more involved in the relevant aspects of the trigger than in the novices. At the same time, it has been observed that experts exhibit more heterogeneous task approaches than novices. It has also been seen that experts use information-based shortcuts.

In another study investigating the differences between experts and other less-expert groups in understanding the visuals, 819 experts, 187 less-experts and 893 novices were investigated and eye movements and performance data of these participants were collected (Gegenfurtner, Lehtinen, & Säljö, 2011). According to the results, experts made shorter eye fixations compared to the novices. Another result was that the experts carried out more eye fixations compared to the novices in the area related to the task. On the other hand, in areas not related to the task, experts made less eye fixations than the novices. In addition, it has been observed that experts have longer saccades than novices and that they fix the relevant information in a shorter time. This situation has shown that experts had superiority in terms of parafoveal processing and selective attention.

An eye tracking study was conducted to understand visual diagnosis in the field of radiology (van der Gijp et al., 2017). It is important to define visual search models and effective teaching strategies that are reflections of educational performance. Visual search models, such as searching which is made systematically in chest X-rays, have been associated with high-level expertise. There was no important effect on perceptual performance in teaching visual search methods. In other words, according to this study, different visual search methods should be developed in order to improve students' image perception skills.

CHAPTER 3

3. METHODOLOGY

3.1. Participants

In total, 41 humans have participated in the within-subjects experiment as part of the present study. In this within-subjects experiment, each of these subjects participated in both pretest and eye-tracking sessions. The subjects were taken one by one at each session of the experiment. Each participant completed the pretest session on paper and then the eye-tracking session on computer. All the participants had normal or corrected-to-normal views.

Participants were from METU, Ankara University, Çankaya University, Pamukkale University and Gazi University and they were graduate students or PhD students in different departments or humans that completed their doctorates. 21 of the participants were female and 20 were male. The mother tongue of all participants was Turkish and the experiment was done in English for the pretest and Turkish for the eye-tracking test. The participants' degree of education recorded. Participants were given a present as incentive for their participation.

3.2. Experiment Procedure

Ethics committee approval was given by METU Ethics Committee for the experiment. Before the experiment, the participants received an approval and voluntary participation form. The experiment started after the participant read and approved the approval form.

This experiment has two phases. In the first phase, a pretest was carried out on the paper to determine the levels of the users. In the second phase, eye-tracking test was performed on the computer. Participation in this experiment took approximately 30 minutes.

3.2.1. The First Experiment Session

The first stage of the experiment is called the pretest phase. At this session, which started after the participant read and approved the experimental approval form, a test paper consisting of 15 questions was given to the participants. The questions mainly contained basic information about computer networks and these questions were presented to the subjects in multiple choice format. Each question consisted of different answer options where one of them was correct answer while others were incorrect answers.

In the first phase of the experiment, it was observed whether participants were experienced ICT or not. With this pretest session, it was aimed to classify the participants as expert and novice according to their knowledge levels.

3.2.2. The Second Experiment Session

In the second session of the experiment, subjects were asked to answer a total of 10 questions, which were presented on a computer screen, and eye movements of the subjects were recorded by an eye-tracking device.

Due to the use of the eye-tracking device, users were given a calibration test before starting this second session. The participant was allowed to start answering the questions when excellent calibration was achieved by taking 5 stars in this calibration test, which corresponds to good calibration in Eye Tribe eye tracker terminology.

The questions were presented as multiple choices questions. Each question had four answer options, and only one of these options was the correct answer.

Seven questions were composed of visual figures and texts while the remaining three questions were text-only. Questions that had visual figures and their answer screens were given in different screens. The question screen of the questions with figures was created through a picture and a text which describes the question, while the answer screens of these questions were designed with four pictures, one for each option. All visual figures used in this session were designed by a drawing program called ConceptDraw. The following figure shows the screen of a question with a visual diagram.

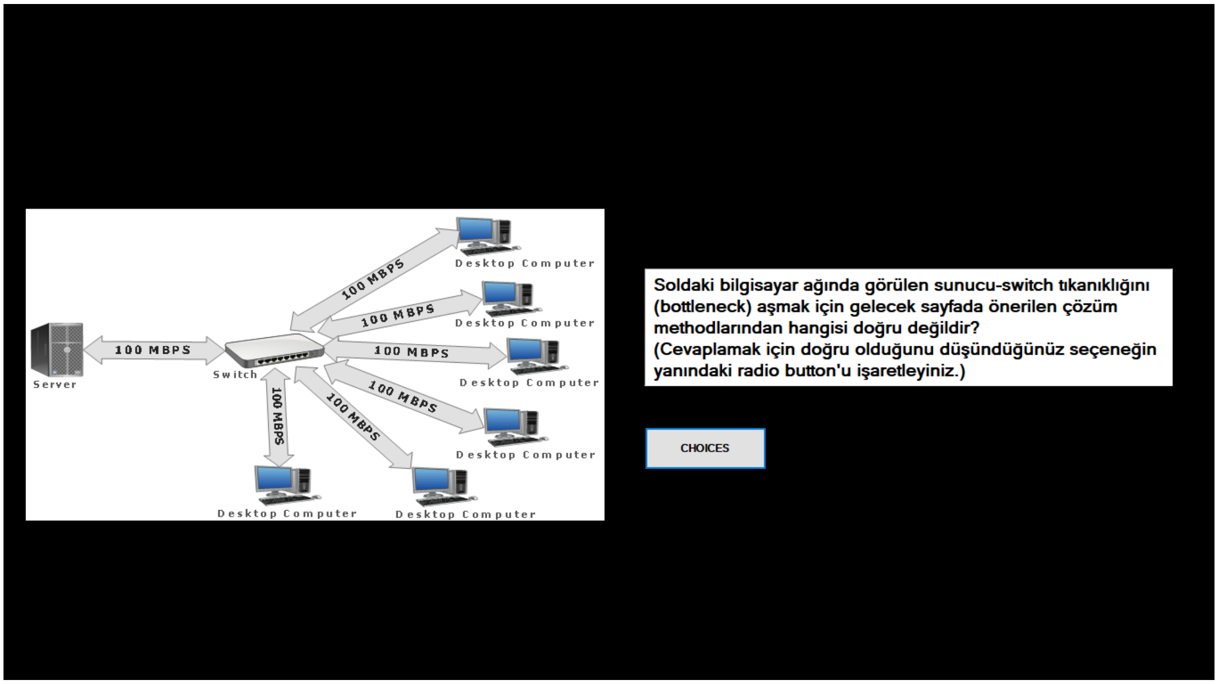

Figure 1: Question Screen Example

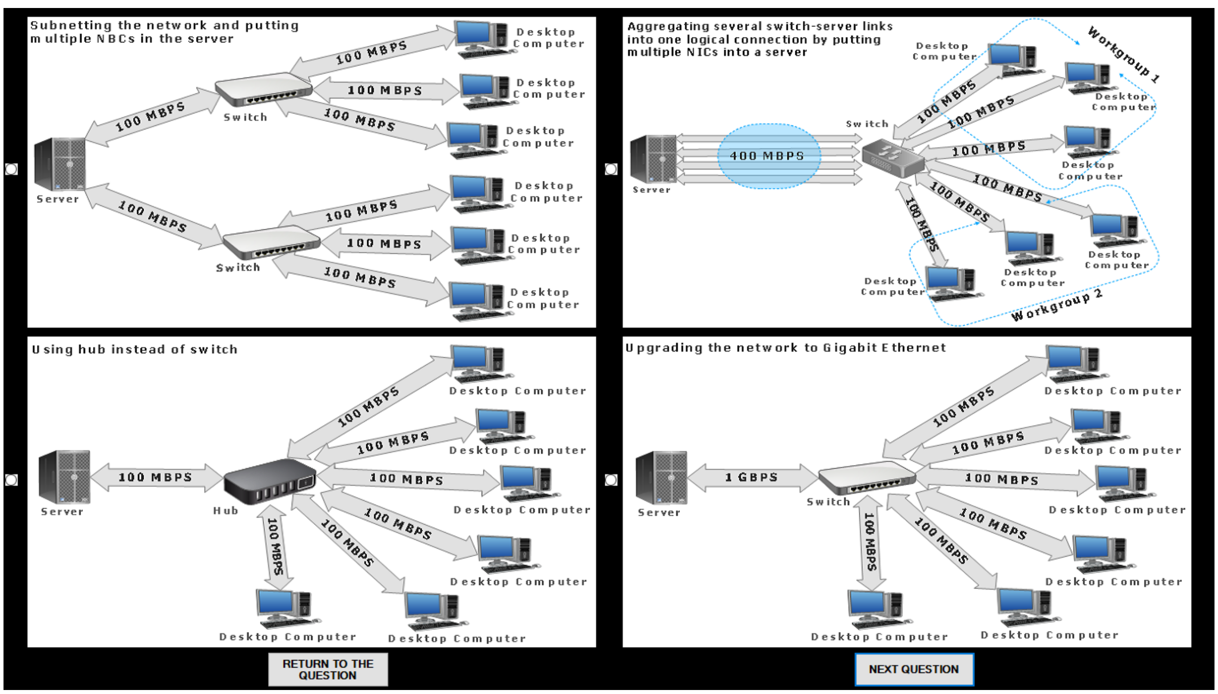

Figure 2: Question Choices Screen Example

In Figure 2, the answer screen of the question given in Figure 1 is shown.

In questions that did not include any figure, the question and answer options were shown to the participants on a single screen. These questions consisted of only texts. The following figure shows one of such questions.

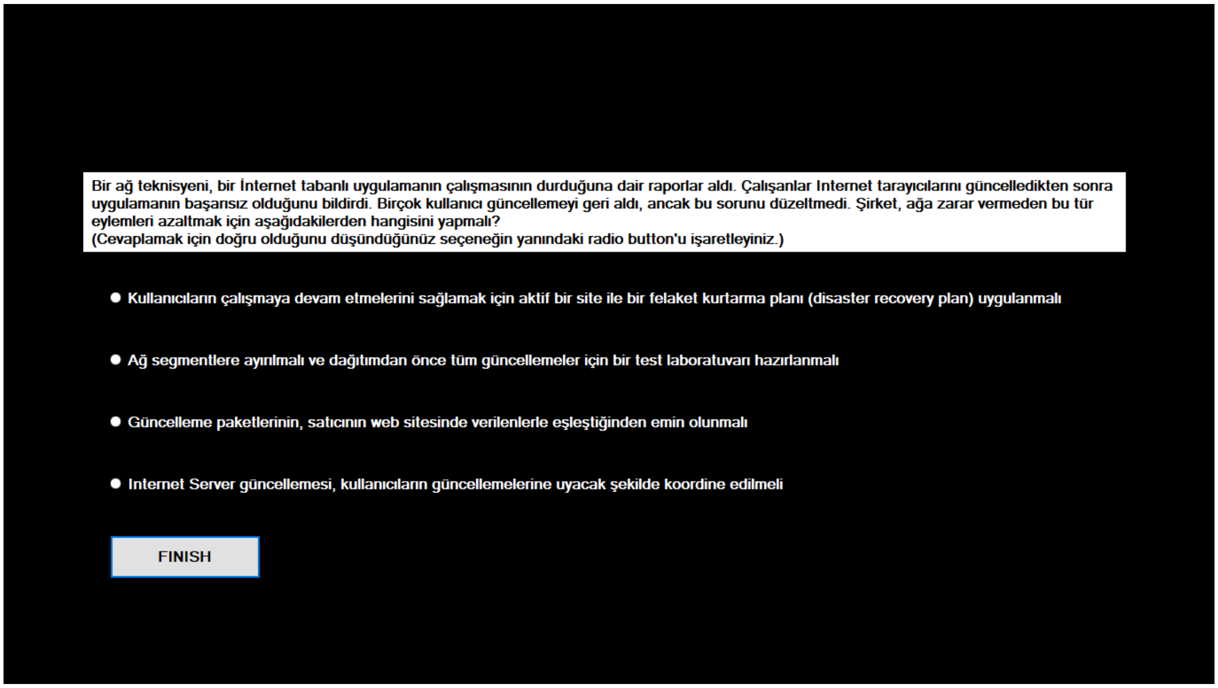

Figure 3: Question with Choices Screen Example

In the answer parts of all questions, the answer options were selected via radio button. Participants were allowed to switch between the question screen and the screen that includes answer options, while screen shifts between the questions were not allowed.

3.3. Eye Tracking Equipment

At this stage which is the second session of the experiment, a laptop was used to provide the experimental environment. The screen resolution of this computer is 1366 x 768. The upper left corner of the screen represents the first pixel (px) showed as x = 0, y = 0 in the coordinate system. The top left corner of the screen starts as the origin because the used eye-tracking device receives x and y coordinates in this way. The last pixel on the screen is shown as x = 1366, y = 768, which corresponds to the lower right corner of the screen.

In this second stage, two devices were used. The first device is the EyeTribe eye tracking device which records the eye movements of the participants. The other device is a Logitech mouse which is used to switch between screens and select the answer option. In order to use during this stage of the experiment, an application was developed by using C# which is an object-oriented programming language. In this application which is a Windows Forms application, the screen coordinates are taken from the

System.Forms.Inputs library of .NET resources. Also, a chin rest was used to prevent the subject from moving head while answering the questions on the screen. Figure 4 shows the experimental setup consisting of laptop, EyeTribe eye-movement device, mouse and chin rest.

Figure 4: Experimental Setup

In this present study, raw eye movement data was obtained from the EyeTribe API and collected after a 9-point calibration at a sampling rate of 30 Hz.

3.4. Procedure and Data

In the eye-tracking phase of the experiment, the subjects answered the questions on the computer while the C # application developed was recorded the raw eye movements of these participants on the text files. The format of this raw eye movement data is prepared in the form of which question screen, the X coordinate value of the eye movement, the Y coordinate value of the eye movement and the time of this eye movement, respectively. The time intervals between the raw eye movement data is 33.3 milliseconds. This is because EyeTribe device operates at a sampling rate of 30 Hz. Figure 5 shows a part of the text file in which this raw eye movement data is recorded.

```
State :Q1 Question Screen* eyeX :*780,5162* eyeY :*702,8659* Time :*63657994751590
State :Q1 Question Screen* eyeX :*781,5541* eyeY :*689,4124* Time :*63657994751594
State :Q1 Question Screen* eyeX :*947,1412* eyeY :*515,311* Time :*63657994751595
State :Q1 Question Screen* eyeX :*1039,5671* eyeY :*439,5748* Time :*63657994751598
State :Q1 Question Screen* eyeX :*1040,7653* eyeY :*449,7158* Time :*63657994751598
State :Q1 Question Screen* eyeX :*1025,5847* eyeY :*410,2613* Time :*63657994751607
State :Q1 Question Screen* eyeX :*1042,4192* eyeY :*430,1905* Time :*63657994751611
State :Q1 Question Screen* eyeX :*1037,2234* eyeY :*429,4838* Time :*63657994751620
State :Q1 Question Screen* eyeX :*1054,3436* eyeY :*294,9724* Time :*63657994751621
State :Q1 Question Screen* eyeX :*1083,9213* eyeY :*190,1646* Time :*63657994751624
State :Q1 Question Screen* eyeX :*1088,036* eyeY :*175,4744* Time :*63657994751624
State :Q1 Question Screen* eyeX :*1088,4717* eyeY :*181,209* Time :*63657994751627
State :Q1 Question Screen* eyeX :*1086,9463* eyeY :*173,7485* Time :*63657994751628
State :Q1 Question Screen* eyeX :*1079,1406* eyeY :*182,5435* Time :*63657994751629
State :Q1 Question Screen* eyeX :*1086,7043* eyeY :*172,8176* Time :*63657994751630
State :Q1 Question Screen* eyeX :*1045,619* eyeY :*201,0583* Time :*63657994751634
State :Q1 Question Screen* eyeX :*1053,4364* eyeY :*199,3343* Time :*63657994751635
State :Q1 Question Screen* eyeX :*1063,7203* eyeY :*200,1491* Time :*63657994751638
State :Q1 Question Screen* eyeX :*1053,4128* eyeY :*207,9419* Time :*63657994751638
State :Q1 Question Screen* eyeX :*1059,0198* eyeY :*209,3781* Time :*63657994751639
State :Q1 Question Screen* eyeX :*1057,9546* eyeY :*188,2104* Time :*63657994751644
State :Q1 Question Screen* eyeX :*701,385* eyeY :*297,8508* Time :*63657994751644
State :Q1 Question Screen* eyeX :*630,6888* eyeY :*342,541* Time :*63657994751661
State :Q1 Question Screen* eyeX :*620,6151* eyeY :*334,5754* Time :*63657994751701
State :Q1 Question Screen* eyeX :*624,6568* eyeY :*322,4592* Time :*63657994751731
State :Q1 Question Screen* eyeX :*611,8362* eyeY :*331,2257* Time :*63657994751761
State :Q1 Question Screen* eyeX :*619,4312* eyeY :*331,4581* Time :*63657994751811
State :Q1 Question Screen* eyeX :*536,2949* eyeY :*351,9926* Time :*63657994751821
State :Q1 Question Screen* eyeX :*422,1492* eyeY :*343,2763* Time :*63657994751851
State :Q1 Question Screen* eyeX :*404,3319* eyeY :*346,4549* Time :*63657994751881
State :Q1 Question Screen* eyeX :*413,5591* eyeY :*354,665* Time :*63657994751917
State :Q1 Question Screen* eyeX :*408,8008* eyeY :*345,8185* Time :*63657994751947
State :Q1 Question Screen* eyeX :*408,6952* eyeY :*339,9742* Time :*63657994751997
State :Q1 Question Screen* eyeX :*336,568* eyeY :*350,7241* Time :*63657994752027
State :Q1 Question Screen* eyeX :*334,103* eyeY :*340,3107* Time :*63657994752057
```

Figure 5: Content of Eye Movement Data File

In a second text file, participants' screen shifts and answers given to these questions were recorded. The text file is given below.

```
DesicionTime :*29336* Screen :*Q1 Question Screen* ID :*Q1
Response :*2* DesicionTime :*62406* Image Names :*Q1A.png*Q1B.png*Q1C.png*Q1D.png* ID :*Q1
DesicionTime :*13284* Screen :*Q2 Question Screen* ID :*Q2
Response :*0* DesicionTime :*7257* Image Names :*Q2 Choices Screen*Q2A.png*Q2B.png*Q2C.png*Q2D.png* ID :*Q2
DesicionTime :*10320* Screen :*Q2 Question Screen* ID :*Q2
Response :*0* DesicionTime :*16977* Image Names :*Q2 Choices Screen*Q2A.png*Q2B.png*Q2C.png*Q2D.png* ID :*Q2
DesicionTime :*1782* Screen :*Q2 Question Screen* ID :*Q2
Response :*3* DesicionTime :*27996* Image Names :*Q2 Choices Screen*Q2A.png*Q2B.png*Q2C.png*Q2D.png* ID :*Q2
DesicionTime :*11521* Screen :*Q3 Question Screen* ID :*Q3
Response :*4* DesicionTime :*25040* Screen :*Q3 Choices Screen* Image Names :*Q3A.png*Q3B.png*Q3C.png*Q3D.png* ID :*Q3
Response :*3* DesicionTime :*11486* Screen :*Q4 Question & Choices Screen* Image Names :*Q4A writing*Q4B writing*Q4C wr
DesicionTime :*42017* Screen :*Q5 Question Screen* ID :*Q5
Response :*1* DesicionTime :*39332* Screen :*Q5 Choices Screen* Image Names :*Q5A.png*Q5B.png*Q5C.png*Q5D.png* ID :*Q5
Response :*2* DesicionTime :*36666* Screen :*Q6 Question & Choices Screen* Image Names :*Q6A writing*Q6B writing*Q6C wr
DesicionTime :*7823* Screen :*Q7 Question Screen* ID :*Q7
Response :*4* DesicionTime :*35602* Screen :*Q7 Choices Screen* Image Names :*Q7A.png*Q7B.png*Q7C.png*Q7D.png* ID :*Q7
DesicionTime :*12806* Screen :*Q8 Question Screen* ID :*Q8
Response :*1* DesicionTime :*54145* Screen :*Q8 Choices Screen* Image Names :*Q8A.png*Q8B.png*Q8C.png*Q8D.png* ID :*Q8
DesicionTime :*3500* Screen :*Q8 Question Screen* ID :*Q8
Response :*1* DesicionTime :*2751* Screen :*Q8 Choices Screen* Image Names :*Q8A.png*Q8B.png*Q8C.png*Q8D.png* ID :*Q8
DesicionTime :*5167* Screen :*Q9 Question Screen* ID :*Q9
Response :*0* DesicionTime :*44493* Screen :*Q9 Choices Screen* Image Names :*Q9A.png*Q9B.png*Q9C.png*Q9D.png* ID :*Q9
DesicionTime :*4641* Screen :*Q9 Question Screen* ID :*Q9
Response :*3* DesicionTime :*7221* Screen :*Q9 Choices Screen* Image Names :*Q9A.png*Q9B.png*Q9C.png*Q9D.png* ID :*Q9
Response :*2* DesicionTime :*83220* Screen :*Q10 Question & Choices Screen* Image Names :*Q10A writing*Q10B writing*Q10
```

Figure 6: Subject Answer File

3.5. Data Cleansing

All of 41 participants answered 10 questions and this resulted in 410 screen data in total. However, it is necessary to perform cleaning process on the 410 screen data where raw eye movements were recorded. At this point, 25 screen data which has less raw eye movement record samples than 200 was excluded. Then, during the experiment, 4 screen data with no eye movement record was also excluded because of the cases that the C # application did not take any record or no record has in the region of the correct answer. Thus, 29 out of 410 screen data has been removed from the analyses. This ratio corresponds to 7.07% of the total screens.

CHAPTER 4

4. RESULTS

In this chapter, the results of the experiment are reported. All analyzes were obtained using JASP 0.9.2.0 which is an open source statistical tool.

First of all, the pretest scores and accuracy (i.e., task performance) will be reported in this chapter. Then, the duration of the eye movements of the subjects classified according to the experimental tests will be analyzed. The results will be summarized in the last section.

4.1. Pretest Scores and Accuracy

The experiment consisted of two stages. In the first stage, a 15-question pretest was given to the participants to assess the expertise of the subjects on computer networks and ICT in general, and then the second stage was conducted where the eye movements of the participants were recorded during answering 10 ICT questions, which were different than the pretest questions.

In the pretest stage, which was the first stage of the experiment, 15 multiple choice questions were asked to the participants and the correct answer to each question was accepted as 1 point. In other words, the participants were subjected to a test that would receive a score between 0 and 15. The box graph of the pre-test results of the 41 participants is shown below. According to Figure 7, the mean of the participants’ scores was 8,13 for the pretest which gives between 0 and 15. Also, the highest score was seen as 13 while the lowest score was seen as 3,00.

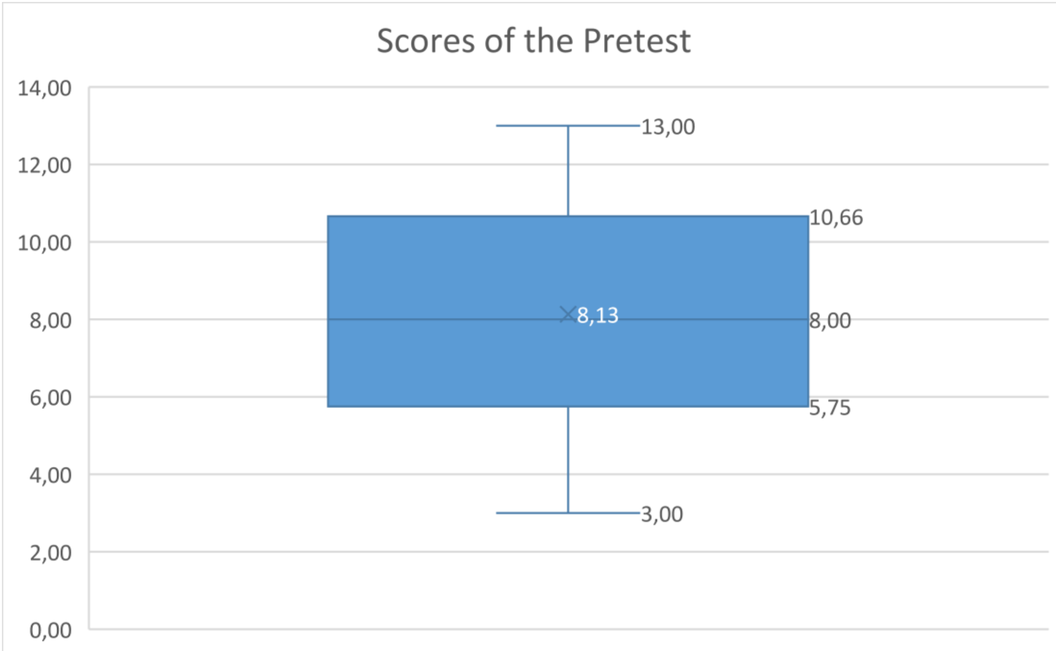

Figure 7: Score Distribution of Participants in the First Experiment Session

In the eye-movement session which was the second phase of the experiment, another set of 10 questions were asked to the participants on a computer screen. When each question was answered correctly, the participant got 1 point. In other words, the participants received a score of 0 to 10 from this session. The results of 41 participants in this second stage test are shown in the following box graph. According to Figure 8, the mean of the participants' scores was 4,54 for the eye-tracking test which gives between 0 and 10. Also, the highest score was seen as 7 while the lowest score was seen as 0,00.

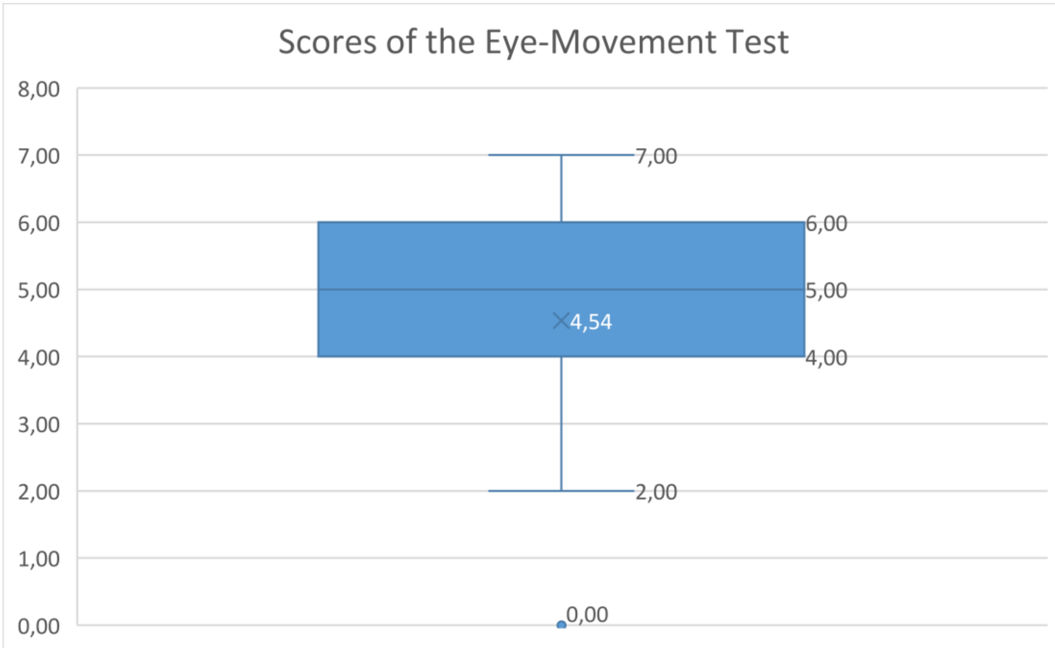

Figure 8: Score Distribution of Participants in the Second Experiment Session

In dividing the participants into two groups, as experts and novices, the scores of the pretest were taken into consideration. According to the pretest results, the subjects were identified as either an expert or a novice. After then, a high performer-low performer division was carried out according to the experimental task in the eye-movement test which was the second experimental session. In other words, the subjects were divided in two separate ways for each session test. Briefly, the participants were divided as experts and novices according to the pretest scores. Then, the same participants were divided as high performers and low performers according to the experimental task in the second experimental session.

The eye-movements of the participants were recorded during the second experimental session so that we used these eye-movement measures for gaze shifts between screens during the experimental task and gaze durations on the screens. During the first experiment session, the participants responded the 15 questions related to ICT on paper. In other words, there was no eye movement recording during the pretest.

4.2. Duration Analysis for Groups

The participant divisions were performed according to the results from two different tests. This has led to differences in groups of subjects. All of these groups were gathered under a single roof and an analysis was conducted to observe group differences.

EyeTribe which is used for eye-movement measures runs at a sampling rate of 30 Hz. In other words, the EyeTribe records a gaze sample every 33.3 milliseconds and the duration within the range of two gaze samples in succession is 33.3 milliseconds. Thus, the number of gaze samples of participants also indicates to the gaze duration of the participants. We can obtain the gaze duration by the following simple equation:

$$Gaze\ Duration\ of\ Participant = Number\ of\ Gaze\ Samples\ \times 33.3\ milliseconds$$

4.2.1. Findings on the Division According to the First Experiment Scores

In this section, the results were related to the participants, who were divided into experts and novices according to the pretest scores obtained in the first stage of the experiment. The first 20 participants were identified as experts, while the remaining 21 were identified as novice. After this division, it was measured how long the participants gazed at the correct answer as well as other three incorrect answers during the second experiment task. The following figure shows the number of gaze samples of experts in terms of looking at correct and incorrect answers for every question.

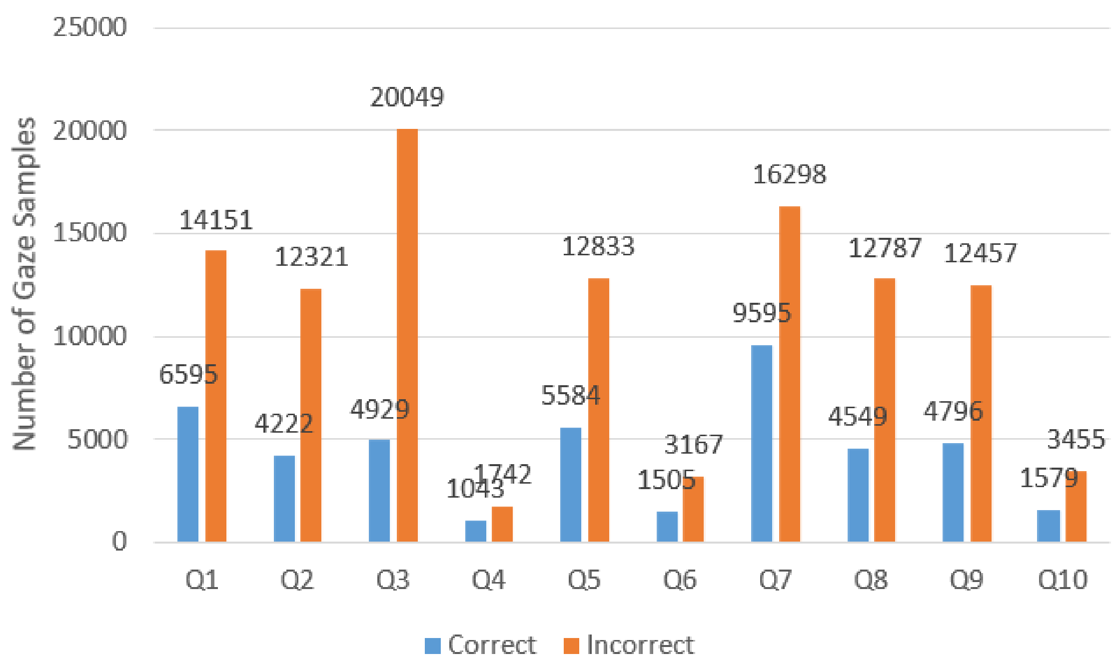

Figure 9: Number of Gaze Samples of Experts According to the First Experiment

The following figure shows the number of gaze samples of novices in terms of looking at correct and incorrect answers for every question.

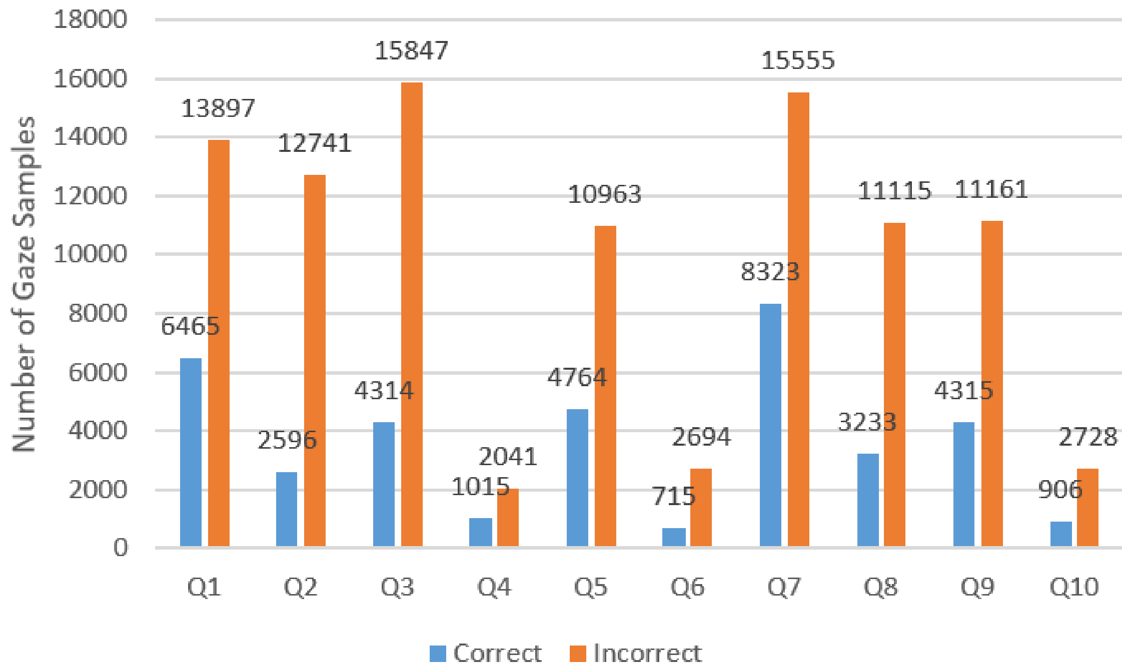

Figure 10: Number of Gaze Samples of Novices According to the First Experiment

At this point, it should be noted that the questions numbered as 4, 6 and 10 didn't contain any visual figures. In other words, they were composed of only texts. Therefore, the answer options covered much smaller areas than other seven questions on the screen. The time to look at the options in these questions was less.

Number of gaze samples on correct and incorrect answers according to the total look-up times are given in the following figure.

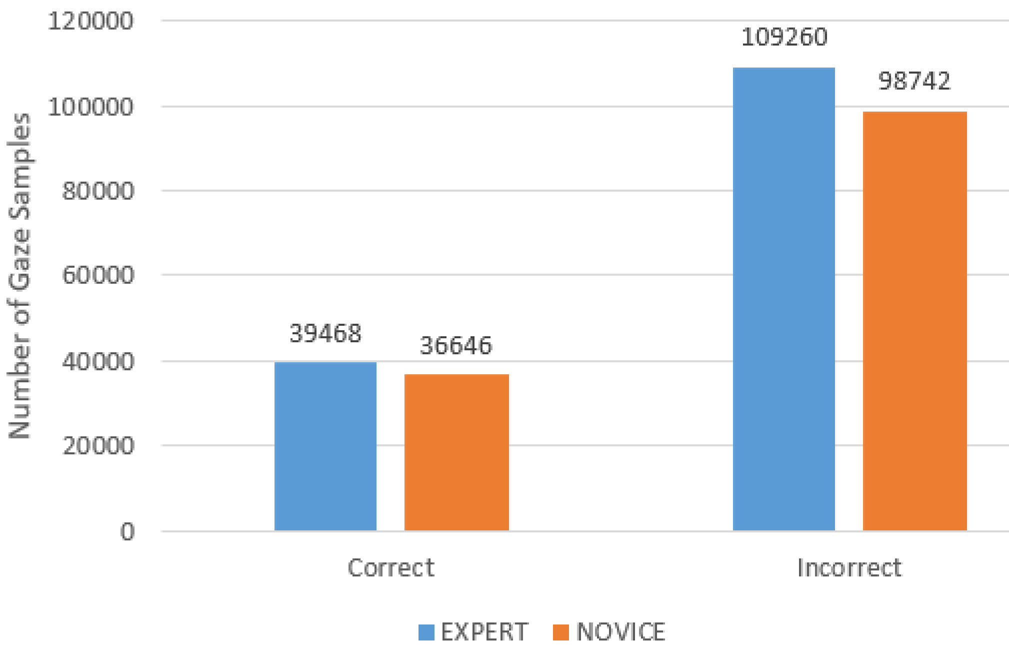

Figure 11: Total Number of Gaze Samples of Experts and Novices According to the First Experiment

According to these results without ANOVA, there was no significant difference between expert and novice groups in gazing at the correct answer.

After these raw findings, the data of the subject groups were observed by Analysis of Variance (*ANOVA*). Firstly, the percentages of correct answers and incorrect answers of the participants were obtained. In other words, the number of correct answers of a participant was divided by the total number of the questions in the experimental task in the second experiment session and the correct answer percentage of the related participant was obtained. In here, the total number of the questions was ten for the experimental task. The same procedure was applied to obtain the percentage of incorrect answers. At this point, it must be known that the questions which left empty by the related participant were not included in the incorrect answers.

After the determination of the percentages of correct and incorrect answers belonging to the participants, a repeated measures ANOVA was implemented. At this repeated measures ANOVA process, the percentages of correct and incorrect answers were added into repeated measures factors, namely accuracy and the group variable included in experts and novices was selected as between subjects variable. According to the repeated measures ANOVA results, there was statistically no significant difference between expert and novice groups with regards to the percentages of correct and incorrect answers, $F(1,39) = 2{,}880$, $p = 0.098 > 0.05$. This result can be seen the table below.

Table 1: ANOVA Results in terms of Correct-Incorrect Answer Percentages of Experts and Novices

Within Subjects Effects ▼

	Sum of Squares	df	Mean Square	F	p
Accuracy	0.113	1	0.113	2.953	0.094
Accuracy ✻ Group	0.050	1	0.050	1.302	0.261
Residual	1.498	39	0.038		

Note. Type III Sum of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Group	0.025	1	0.025	2.880	0.098
Residual	0.338	39	0.009		

Note. Type III Sum of Squares

To obtain the looking ratios at correct answer, the number of gaze samples on the correct answer was divided by the number of gaze samples on all the screen where the stimuli were displayed. The same procedure was applied to obtain the ratios of looking at incorrect answers. Repeated measures ANOVA was used to analyze the ratios of gaze durations for experts and novices which were determined according to pretest scores in the first experiment session. The ratios of gaze durations on correct and incorrect answers were added into Repeated measures factors, namely GazeDurationRatio. The pretest groups, namely PretestGroup, consisting of experts and novices were used as the between subjects variable. According to the results, there was statistically no significant difference between expert and novice groups with regards to the ratios of looking at correct and incorrect answers in terms of duration, $F(1,39) = 1,323$, $p = 0.257 > 0.05$. This result can be seen at Table 2.

Table 2: ANOVA Results of the Ratios of Looking at Correct and Incorrect Answers of Experts and Novices

Within Subjects Effects ▼

	Sum of Squares	df	Mean Square	F	p
GazeDurationRatio	0.003	1	0.003	1.334	0.255
GazeDurationRatio ✻ PretestGroup	0.002	1	0.002	0.737	0.396
Residual	0.096	38	0.003		

Note. Type III Sum of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
PretestGroup	0.003	1	0.003	1.323	0.257
Residual	0.085	38	0.002		

Note. Type III Sum of Squares

Repeated measures ANOVA was used to analyze the number of shifts between question and choices screens in terms of experts and novices. The number of shifts between question and choices screens were added into Repeated measures factors, namely GazeShift. The pretest groups which are called as PretestGroup and consist of experts and novices were used as the between subjects variable. There was statistically no significant difference between expert and novice groups with regards the number of shifts between question and choices screens, according to correct and incorrect answers given by the participants, F (1,39) = 0,761, p = 0.389 > 0.05. This result can be seen the table below.

Table 3: ANOVA Results of the Number of Shifts Between Screens in terms of Experts and Novices

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
GazeShift	0.167	1	0.167	1.673	0.204
GazeShift ✻ PretestGroup	5.512e −4	1	5.512e −4	0.006	0.941
Residual	3.782	38	0.100		

Note. Type III Sum of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
PretestGroup	0.103	1	0.103	0.761	0.389
Residual	5.143	38	0.135		

Note. Type III Sum of Squares

To sum up, the pretest on paper which was used to divide the participants as experts and novices couldn't provide any significant difference between experts and novices on the subject of correct and incorrect answer percentages, the ratio of gaze duration and the number of shifts between screens. In other words, the pretest remained incapable to divide the participants as experts and novices by oneself.

4.2.2. Findings on the Division According to the Second Experiment Scores

In this section, the participants were divided into two groups as high and low performers, according to the scores obtained from the experimental test which the ten questions related to ICT were asked and their eye movements were recorded during this experimental task. Similar to the previous division in the first experiment session, the first 20 participants were labeled as high performers, while the remaining 20 were labeled as low performers. According to this division, the number of gaze samples on correct and incorrect answers on the screen in the eye-movement test of the subject groups is in Figure 12.

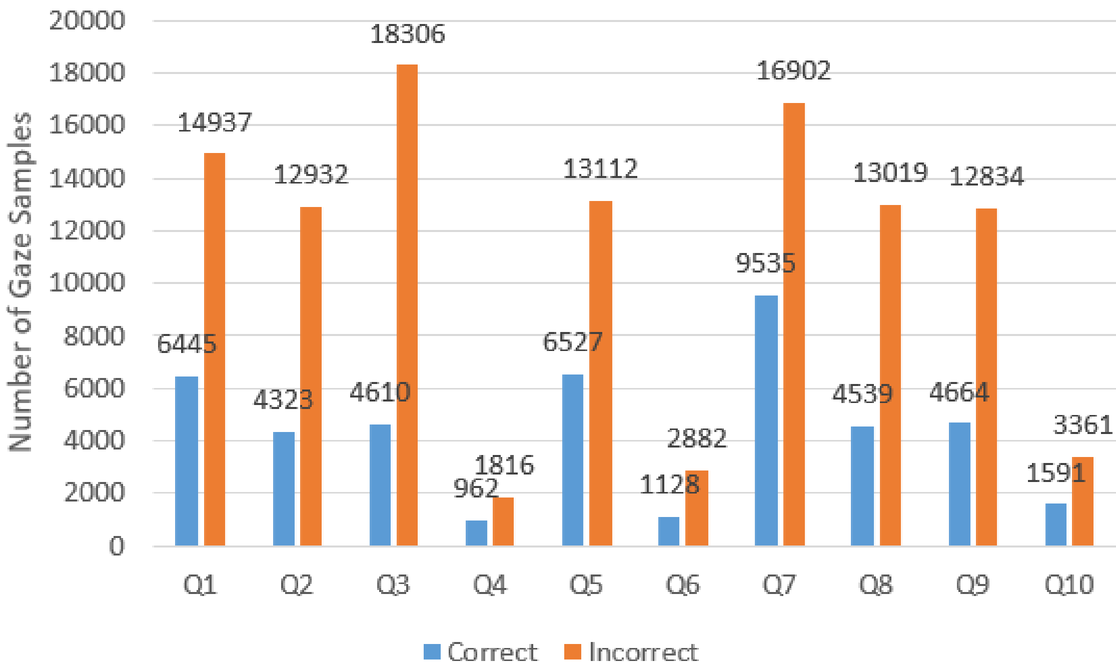

Figure 12: Number of Gaze Samples of High Performers According to the Second Experiment

The following figure shows the number of gaze samples of novices in terms of looking at correct and incorrect answers for every question. At this point, the gaze duration can be evaluated as 30 times of the number of gaze samples by the reason of that EyeTribe runs at a sampling rate 30 Hz.

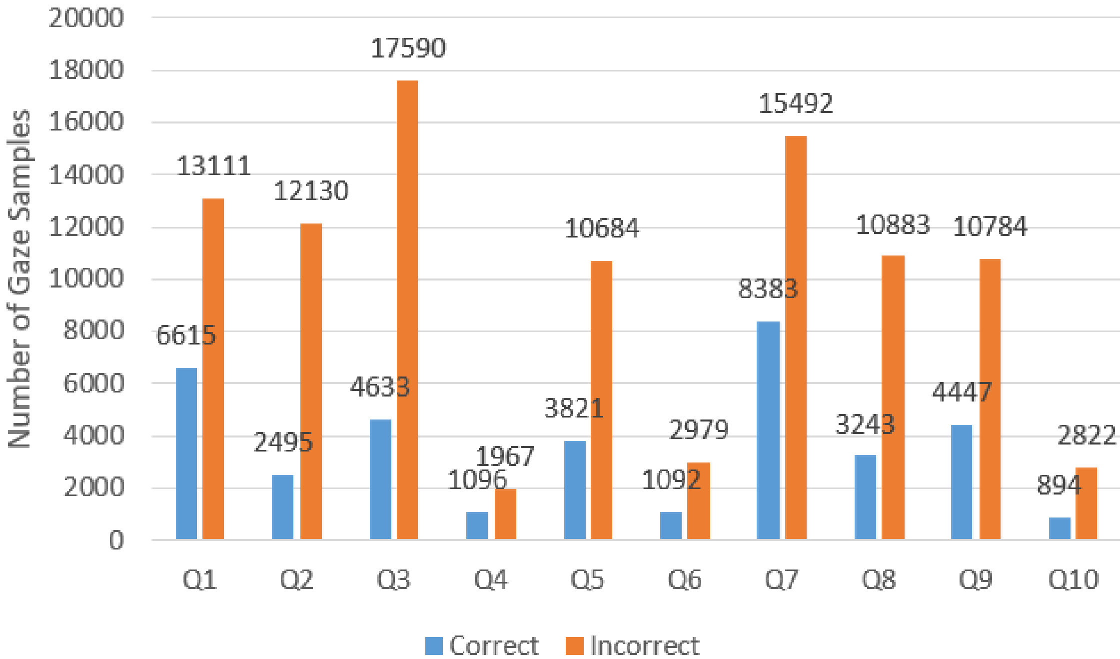

Figure 13: Number of Gaze Samples of Low Performers According to the Second Experiment

Number of gaze samples on correct and incorrect answers according to the total look-up times are given in Figure 14.

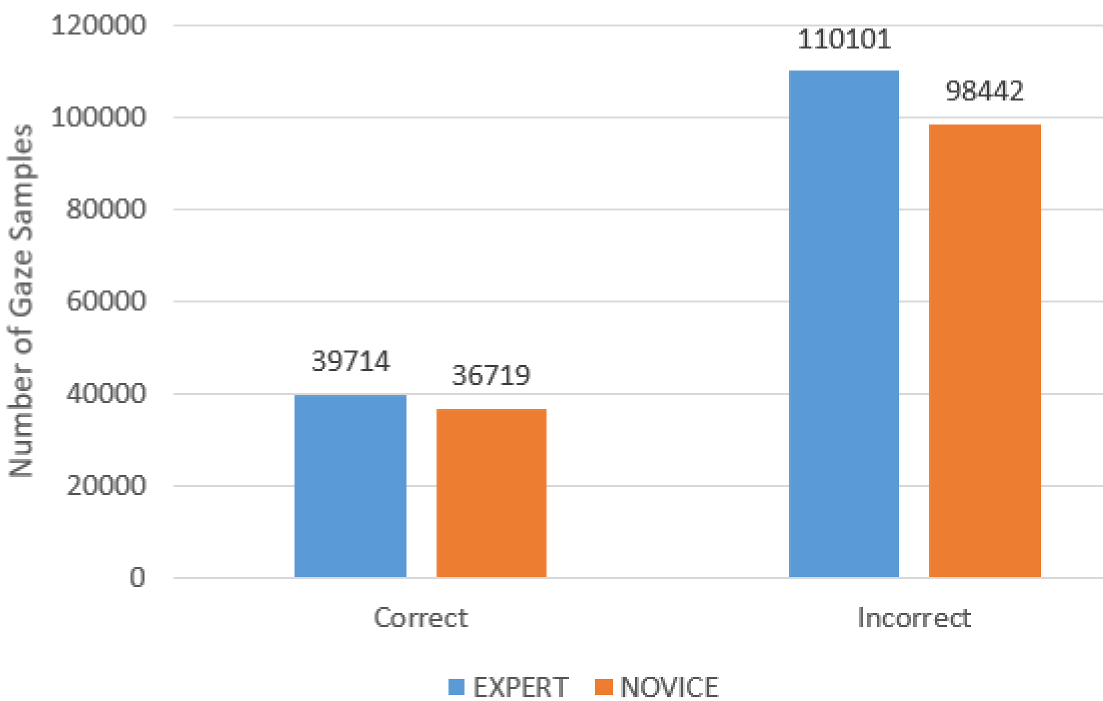

Figure 14: Number of Gaze Samples of High and Low Performers According to the Second Experiment

As it has been mentioned before, the number of gaze samples on the correct answer was divided by the number of gaze samples on all the screen to obtain the ratio of looking at the correct answer in terms of duration. The same procedure was applied to obtain the ratios of looking at incorrect answers. Repeated measures ANOVA was used to analyze the ratios of gaze durations for high and low performers which were determined according to the experimental task scores in the second experiment session. The ratios of gaze durations on correct and incorrect answers were added into Repeated measures factors, namely GazeDuration. The experimental task groups, namely TaskPerformanceGroup, consisting of high and low performers were used as the between subjects variable. According to the results, there was statistically significant difference between high and low performer groups with regards to the ratios of looking at correct and incorrect answers in terms of duration, F (1,39) = 4,441, p = 0.042 < 0.05. This result can be seen the table below.

Table 4: ANOVA Results of the Ratios of Looking at Correct and Incorrect Answers of High and Low Performers

Within Subjects Effects ▼

	Sum of Squares	df	Mean Square	F	p
GazeDuration	0.003	1	0.003	1.223	0.276
GazeDuration ✻ TaskPerformanceGroup	6.050e −4	1	6.050e −4	0.237	0.629
Residual	0.097	38	0.003		

Note. Type III Sum of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
TaskPerformanceGroup	0.009	1	0.009	4.441	0.042
Residual	0.079	38	0.002		

Note. Type III Sum of Squares

Repeated measures ANOVA was used to analyze the number of shifts between question and choices screens in terms of high and low performers. The number of shifts between question and choices screens were added into Repeated measures factors, namely GazeShift. The experimental task groups which are called as TaskPerformanceGroup and consist of high and low performers were used as the between subjects variable. There was statistically no significant difference between high and low performer groups with regards the number of shifts between question and choices screens, according to correct and incorrect answers given by the participants, F (1,39) = 1,726, p = 0.197 > 0.05. This result can be seen the table below.

Table 5: ANOVA Results of the Number of Shifts Between Screens in terms of High and Low Performers

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
GazeShift	0.167	1	0.167	1.673	0.204
GazeShift ∗ TaskPerformanceGroup	1.250e −6	1	1.250e −6	1.256e −5	0.997
Residual	3.783	38	0.100		

Note. Type III Sum of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
TaskPerformanceGroup	0.228	1	0.228	1.726	0.197
Residual	5.018	38	0.132		

Note. Type III Sum of Squares

As can be seen from the above results, the high performer subjects had higher values than the low performer subjects in terms of the looking rates at the correct answer options. On the other hand, there was no significant difference between high and low performers in terms of the number of shifts between screens according to the second experiment session.

4.3. Summary of the Results

As a summary, according to the first analysis results, it was observed that there was no significant difference between experts and novices which were divided according to the first experiment session, pretest. Furthermore, no significant difference was statistically observed between these two groups in the way of looking longer at the correct answer option with the repeated measures ANOVA analysis of the second experiment session.

Then, we divided the participants as high performers and low performers according to the second experimental session scores and analyzed the eye movements of high and low performers in the experimental task. it was observed that there is statistically no significant difference between high and low performers in terms of the number of shifts between the question and answer screens according to the ANOVA results. On the other hand, high performer participants passed low performer participants in terms of the rates of looking durations at the correct answer option in comparison to all the screen, where the stimuli were displayed ($p = 0.042 < 0.05$, according to ANOVA results given in Table 4). In other words, no significant differences were statistically observed in other gaze measures except looking rates in terms of the experimental task scores. These results gave that the investigation of the differences between high and low performers by eye tracking is subject to a limited set of measures.

As a conclusion, the first hypothesis was not supported. This is because that there was no relation between the pretest and the experimental task based on the division of the participants according to their expertise levels at the field of ICT. Also, it was observed that only pretest scoring was not enough to discriminate between novices and experts in terms of ICT. On the other side, the second hypothesis presented in the present thesis has been supported partially. It was observed that the experimental task scores showed better performance than the pretest scores in the matter of the division of participants according to their knowledge levels. There was statistically a significant difference between high performers and low performers in terms of the ratios of looking durations at the correct answer option in comparison to all the screen. No significant differences were statistically observed in other gaze measure (gaze shifts between screens) in the experimental task session.

CHAPTER 5

5. CONCLUSION AND DISCUSSION

Within the scope of this thesis, the approaches of different subject groups in question answering related to ICT were examined. At this point, four main studies have been effective in the preparation of the experimental process. In the first of these studies, the approaches of different user groups in answering of physics problems was examined (Chi, Feltovich, & Glaser, 1981). In the second related study, the comprehension process of subjects with text and diagrams were analyzed (Hegarty & Just, 1993) and the second study has given an idea about the creation of the eye-movement test in the second session of the experiment in the present thesis. In the third related study, the expertise in the artistic field was investigated (Vogt & Magnussen, 2007). In the final related study, an eye tracking study was conducted to understand visual diagnosis in the field of radiology (van der Gijp et al., 2017).

In this present study, a two-stage experimental study and observations were made in respect of which knowledge assessment methods can be used about the division of participants as experts and novices in the pretest, as high performers and low performers in the experimental task.

In this chapter, firstly, the pretest results were evaluated. Then, the results of the eye-movement test were evaluated. Finally, the restrictive conditions encountered during the present thesis study and the future studies in this field were discussed.

5.1. Discussion of Pretest Results

According to two separate tests conducted in the scope of this thesis, two separate subject divisions were carried out and the performance of the participants in these two tests was observed. In the process of determining the questions addressed to the participants in the tests, the questions, which the participants who do not know much about ICT can solve by only reasoning, are tried to be selected. On the other hand, the questions requiring expertise and knowledge at a certain level beforehand were also asked to the participants.

It is thought that the changes or improvements that will be made on test questions asked to the participants can help to get more accurate results in relation to the observation of the approaches based on eye-movements of the groups.

According to the analyzes on the pretest scores of the participants, the pretest couldn't be a successful method to divide the participants as experts and novices. A pretest-based division between the participants as experts and novices did not align with task performance in ICT question answering.

5.2. Discussion of Eye Movement Results

One of the main questions of this present study was whether the pretest scores on paper were enough to meet needs based on the division of experts and novices. In other words, it was expected that only pretest scores would be useful in terms of the division of experts and novices. However, it was observed that there was no significant effect of the pretest scores in the matter of the division of experts and novices. The only difference between these groups was observed when the same participants are divided into high and low performers according to their scores in the eye tracking test. High performer subjects had a longer looking rates at the answer screens than the novice group members.

These results obtained with ANOVA analysis showed that there was statistically significant difference between high performers and low performers in terms of the looking rates at the answer screens. Thus, the hypothesis presented in this present thesis has been supported partially and it was observed that the experimental task scores showed better performance in the matter of the division of participants according to their knowledge levels. In other words, there is a relation between expertise groups (as identified by task performance) and their gaze durations on correct and incorrect answers during the course of their answering ICT questions. On the other hand, there is no relation between expertise groups and their gaze shifts between screens.

5.3. Limitations and Future Work

Within the scope of this thesis, a two-stage experimental process was conducted on 41 participants. The aim was to differentiate the subjects into two different groups in terms of their expertise levels and to observe the differences and approaches of these different groups in question answering. It should be noted that questions used in this study were connected with ICT. The aim is to observe benefits of combination of pretest and accuracy test based on eye movement measures in terms of the division of the participants as expert and novice. At this point, if more participants could have been tested, more healthy inferences about the processes of understanding, interpreting and answering the questions among these groups would have been made. Also, if the participants have been divided into more than two groups instead of two separate groups, they could have been observed different details with regards to the approaches in question answering.

In the context of this thesis, the eye-movement analysis was carried out in terms of only looking durations. A more detailed eye-movement analysis would have led to more accurate results in terms of the division of groups by different analyzes such as eye fixations, saccades.

On the other side, the usage of a computer with higher resolution could have helped to draw more healthier px values in terms of the coordinates of the eye-movement data.

Another challenge in the study of ICT was that ICT is a large area. If we were able to focus on a more local subject at the field of ICT, we could have more clear results. Because the definition of ICT takes in a lot of factors. Otherwise, it is difficult to interpret the results obtained from the eye measurements of the participants.

This study may provide a positive step for future studies in the matter of division of humans in terms of expertise at a certain domain. If the restrictive factors mentioned here can be improved, the details about the division of participants according to their expertise levels can be captured more clearly.

REFERENCES

Alexander, P. A. (2003). Can We Get There From Here? *Educational Researcher*, *32*(8), 3–4. https://doi.org/10.3102/0013189x032008003

Bedard, J., & Chi, M. T. H. (1992). Expertise. *Current Directions in Psychological Science, 1*(4), 135–139. https://doi.org/10.1111/1467-8721.ep10769799

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study. *IEEE Transactions on Image Processing*, *22*(1), 55–69. https://doi.org/10.1109/tip.2012.2210727

Chase, W. G., & Simon, H. A. (1973). *The mind's eye in chess*. *Visual information processing*. New York: Academic Press.

Chi, M. T. H. (1978). Knowledge structures and memory development. In *Childrens' Thinking: What develops?* (pp. 73–96). Hillsdale, N. J.: Erlbaum.

Chi, M. T. H. (1997). Quantifying Qualitative Analyses of Verbal Data : A Practical Guide. *Journal of the Learning Sciences*, *6*, 271–315.

Chi, M. T. H. (2006a). Laboratory Methods for Assessing Experts and Novices Knowledge. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 167–184). https://doi.org/10.1017/cbo9780511816796.010

Chi, M. T. H. (2006b). Two Approaches to the Study of Experts Characteristics. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 21–30). https://doi.org/10.1017/cbo9780511816796.002

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Chi, M. T. H., & Koeske, R. D. (1983). Network Representation of a Child's Dinosaur Knowledge. *Developmental Psychology*, *19*(1), 29–39. https://doi.org/10.1037//0012-1649.19.1.29

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 257–274.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, *41*(6), 801–849. https://doi.org/10.1006/ijhc.1994.1083

De Groot, A. (1966). Perception and memory versus thought: Some old ideas and recent findings. In *Problem solving: Research, method and theory* (pp. 19–50). New York: Wiley.

De Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess*. Assen (The Netherlands): Van Gorcum.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *The Journal of the Human Factors and Ergonomics Society*, *37*(December), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2006). Expertise and Situation Awareness. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 714–742). https://doi.org/10.1017/9781316480748.037

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and Exceptional Performance: Evidence Of Maximal Adaptation To Task Constraints. *Annual Review of Psychology*, *47*(1), 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of Expertise from Psychological Perspectives. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 41–68). https://doi.org/10.1017/cbo9780511816796.004

Fisk, A. D., & Lloyd, S. J. (1988). The role of stimulus to role consistency in learning rapid application of spatial rules. *Human Factors*, *30*, 35–49.

Galton, F. (1869). *Hereditary genius: an inquiry into its laws and consequences*. London: Julian Friedman Publishers.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Glaser, R., & Chi, M. T. H. (1988). Overview. In *The nature of expertise* (pp. xv–xxviii). Hillsdale, N. J.: Erlbaum.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four... or is it two? *Memory*, *12*, 732–747.

Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3*(2), 159–163. https://doi.org/10.3758/bf03212414

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, *6*, 225–255.

Hegarty, M., & Just, M. A. (1993). Constructing Mental Models of Machines from Text and Diagrams. *Journal of Memory and Language*, *32*(6), 717–742.

Hegarty, M., Just, M. A., & Morrison, I. R. (1988). Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning. *Cognitive Psychology*, *20*(2), 191–236.

Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *The AI Magazine, 8*, 53–67.

Hoffman, R. R. (1992). *The psychology of expertise: cognitive research and empirical AI*. New York: Springer-Verlag.

Hoffman, R. R. (1998). How Can Expertise be Defined? Implications of Research from Cognitive Psychology. *Exploring Expertise*, 81–100. https://doi.org/10.1007/978-1-349-13693-3_4

Hoffman, R. R., Trafton, G., & Roebber, P. (2005). *Minding the Weather: How Expert Forecasters Think*. Cambridge, MA: MIT Press.

Horn, J., & Masunaga, H. (2006). A Merging Theory of Expertise and Intelligence. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 587–612). https://doi.org/10.1017/cbo9780511816796.034

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, *20*(2), 172–176. https://doi.org/10.1016/j.learninstruc.2009.02.013

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*(2), 146–154. https://doi.org/10.1016/j.learninstruc.2009.02.019

Jongman, R. W. (1968). *Het oog van de meester (The eye of the master)*. Assen (The Netherlands): Van Gorcum.

Leopold, C., Doerner, M., Leutner, D., & Dutke, S. (2014). Effects of strategy instructions on learning from text and pictures. *Instructional Science*, *43*(3), 345–364. https://doi.org/10.1007/s11251-014-9336-3

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a Complex Skill: Diagnosing X-Ray Pictures. *In The Nature of Expertise*, 1–17. https://doi.org/10.1017/S000748530002229X

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A., & Simon, H. A. (1973). Artificial Intelligence and the concept of mind. In *Computer Models of language and thought* (pp. 1–60). San Fransisco: W. H. Freeman.

Pihko, E., Virtanen, A., Saarinen, V.-M., Pannasch, S., Hirvenkari, L., Tossavainen, T., Hari, R. (2011). Experiencing Art: The Influence of Expertise and Painting Abstraction Level. *Frontiers in Human Neuroscience*, *5*(September), 1–10. https://doi.org/10.3389/fnhum.2011.00094

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye-movements. *Psychological Science*, *12*, 48–55.

Reingold, E. M., Charness, N., Schultetus, R. S., & Stampe, D. M. (2001). Perceptual automaticity in expert chess players. *Pscyhonomic Bulletin and Review*, *8*, 504–510.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In *Childrens' Thinking: What develops?* (pp. 325–348). Hillsdale, N. J.: Erlbaum.

Sternberg, R. J. (1996). *Costs of expertise*. *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. Hillsdale, N. J.: Erlbaum.

van der Gijp, A., Ravesloot, C. J., Jarodzka, H., van der Schaaf, M. F., van der Schaaf, I. C., van Schaik, J. P. J., & ten Cate, T. J. (2017). How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education*, *22*(3), 765–787. https://doi.org/10.1007/s10459-016-9698-1

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception*, *36*(1), 91–100. https://doi.org/10.1068/p5262

Williams, A. M., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research Quarterly for Exercise and Sport*, *69*, 111–128.

Williams, A. M., Davids, K., Burwitz, L., & Williams, J. G. (1994). Visual search strategies of experienced and inexperienced soccer players. *Research Quarterly for Exercise and Sport*, *65*, 127–135.