Exploring and quantifying web user experience with continuous response measurement

Altunyurt, L.

Award date:
2009
Exploring and Quantifying Web User Experience with Continuous Response Measurement

by
This study aims at going beyond the classical usability perspective and understanding users’ experience by combining a variety of different quantitative methods. These methods include questionnaire data, eye-tracking and continuous response measurement. By using these methods, several aspects of users’ experience and overall judgements about three Dutch car insurance websites are analyzed in an experimental setting with a student sample from Eindhoven University of Technology. The results of Continuous Response Measurement indicate that participants’ global assessments about the websites are dominated by the average of the feedbacks provided and also the last feedback. Eye-tracking method also suggests that average eye fixation duration of the participants decrease before they give a negative feedback about the website. These results and others those are found during this study suggest implications for designing successful websites by understanding and differentiating several dimensions of user experience as well as investigating determinants of subjects’ global assessments about websites.
“Take roads. A Usable road is one that is wide and straight (less mental effort), with no oncoming traffic (less mistakes, less mental effort). One that enables you to get from A to B as fast as possible (more powerful) and one that has a consistent and clear use of signs (high learnability). In short the most usable road is a freeway. But, a freeway is also directly boring in terms of user-experience. A road with a high level of user-experience is completely different. It is a twisting mountain road (visual). Now you got great scenery (visual, emotional), the smell of nature, and the excitement from the climb (and the sheer cliff only feet away). You got little friendly signs put out by the local, who sells fruits along your way (show-off effect). Every city is slightly different (branding, emotional, environment). You feel happy when you see the local people wave when you pass by, and you stop let a sheep pass (emotional, trust, coexistence). But a mountain road is far from a usable road. It is much harder to drive on, it is difficult to learn, you can’t go as fast and the risk of making a mistake (taking a wrong turn or cashing into a sheep) is much greater. But, a mountain road will give you a much better user-experience than any freeway could ever do.”

Thomas Baekdal
# Table of Contents

Acknowledgements ........................................................................................................... 5

1. Introduction .................................................................................................................. 6
   1.1 User Experience ......................................................................................................... 8
   1.2 Continuous Response Measurement ....................................................................... 16
   1.3 Eye-Tracking ........................................................................................................... 21
   1.4 Combining the data ............................................................................................... 23

2. Data and Methods ........................................................................................................ 24
   2.1 Design .................................................................................................................... 24
   2.2 Participants ............................................................................................................ 24
   2.3 Dependent Measures ............................................................................................ 24
   2.4 Equipment ............................................................................................................. 24
   2.5 Questionnaire ....................................................................................................... 26
      2.5.1 Pre-task Questionnaire .................................................................................... 26
      2.5.2 Post-task Questionnaire ................................................................................ 27
   2.6 Procedure ............................................................................................................... 27

3. Results ........................................................................................................................ 29
   3.1 Processing the data ............................................................................................... 30
   3.2 Analyzing the data ............................................................................................... 31
   3.3 All in one Regression Table .................................................................................. 48
   3.4 Post-task Interview .............................................................................................. 49

4. Discussion and Conclusion ......................................................................................... 49

5. References .................................................................................................................... 56

6. Appendices ................................................................................................................... 59
   6.1 Appendix 1 – Pre-Task Questionnaire ................................................................. 59
   6.2 Appendix 2 – Post-Task Questionnaire ............................................................... 60
   6.3 Appendix 3 - Instructions ..................................................................................... 62
   6.4 Appendix 4 – Example Java Code ....................................................................... 63
   6.5 Appendix 5 – Sample from raw slider file ........................................................... 66
   6.6 Appendix 6 – Sample from fixation file ............................................................... 67
   6.7 Appendix 7 – Visualizations ............................................................................... 68
Acknowledgements

It is a pleasure to thank those who supported and encouraged me while I was writing this thesis. It really meant a lot to me. There is a group of people out there who made this project possible with either their academic contributions or personal support. Firstly, I want to thank Wijnand Ijsselsteijn for accepting supervision of this project even though he had such a busy schedule. At the beginning, I had difficulty in finding my way on this large and vague field but his ideas and great experience contributed this project a lot by giving it a direction that I enjoyed a lot in the end. This thesis would not have been possible without help of Chris Snijders. He spent so much effort to guide this project and he was always there available to help. I learned a lot from his critical and sceptic point of view and different perspectives that he brings to the problem at hand. I am so grateful to Yvonne van Laarhoven for her ongoing encouragement throughout the project. She has always been available her support in a number of ways like sharing her experiences and clearing away obstacles from my way. I am really happy that she was part of my thesis. I would like to thank Bob Nieme for making this project possible in Adversitement B. V. environment. I really enjoyed a lot being a part of Adversitement during this period since my colleagues always showed their concern about me and encouraged me all along. I owe my deepest gratitude to Ozan Aksoy and Sara Van Butsel for their continuous support during this project. They put a lot of effort and their peer reviews add so much to this project. Lastly, I want to thank my family and friends who were supporting me during the long days I spent in the library or the laboratory. I am so grateful to each and every one of you.
1 Introduction

As soon as you start reviewing the literature of User Experience, it does not take long to notice that this field is quiet chaotic with its definitions, measures, factors pronounced by different authorities. As well as the articles published on internet, even scientific papers lack a consensus. When I had first seen the “roads” example, as presented above (by Thomas Baekdal), I simply loved it. In my opinion, it is a very appropriate metaphor which makes it easier to visualize User Experience and emphasize its difference with usability. Nowadays there is a consensus among Human technology interaction specialists that classic “Usability” approach is no longer sufficient to understand the interaction between systems and the users. As Karat et. al. (2001) states, “we are no longer designing systems just to enable well-specified tasks: we are designing systems with the primary objective of providing users with entertaining experiences.” In order to answer basic questions like “Why people do or do not use a particular website?”, “What makes a good or bad impression about a website?”, “What are the determinants of the experience that users have through the time they spend on a website?” we should go one step beyond usability and try to understand the User Experience in a broader perspective. Nowadays, technology provides us various opportunities to achieve this understanding. Two of them are Continuous Response Measurement (CRM) and Eye-tracking techniques. In this study, we will make use of these techniques to have a deeper understanding of the users’ experience while surfing the websites.

Apart from shifting the classical Usability perspective to User Experience perspective, another important shift is on its way concerning evaluation methodology of websites. The most widely used technique to understand to what extent websites are accepted by users is post-task questionnaires or interviews. These methods are simply based on letting the users “experience” the website and asking their opinion as soon as they have finished. This methodology is an effective way of understanding the users’ overall impression “after” using the website but falls short in explaining what is really going on “during” the interaction. Discovering how people’s impression change over time and what causes these changes is a valuable insight for user experience researchers. It can be claimed that think aloud method also provides continuous feedback from the users during their experience with websites. But the problem here is the quantification of their assessments and also the distracting effect of thinking aloud. Data collected with this method is qualitative rather than quantitative. Moreover, it is a time intensive technique while analysing the gathered data.

Another possible continuous response that can be collected whilst users are exploring the websites is collecting psychophysical data. In this case, the disadvantage is the ambiguity while evaluating the collected data. It may appear that the user was aroused
in some certain time periods but we cannot know for sure what caused this arousal. It can be caused by frustration, excitement or some other emotional response. Same phenomenon holds for behaviour coding as well. Even if the data collected from behaviour coding is continuous, the evaluation part is problematic. Besides being time-intensive, the facial expressions for instance can be interpreted in different ways. A user may laugh because of being pleased or it may be an ironic laugh because of frustration. Unless self-report is provided by the user, these techniques fall short in providing accurate quantitative data. By introducing CRM in this domain, we are attempting to solve these problems.

Biocca et. al. (1994) mention that, “cognitive states of individuals change continuously as they attend to, understand and react to messages”. As they suggest, the interaction between users and websites is a dynamic phenomenon. Their impression changes over time while they achieve or fail to realize their goals. They may like a website or hate it depending on whether or not they can find what they are looking for. We believe that, if the interaction between websites and the users is a dynamic phenomenon, the methods of measuring it should also be dynamic. Several other research domains like advertisements benefited from Continuous Response Measurement in order to figure out the predictors and determinants of the global assessment of an experience (It should be noted that “Overall Impression” and “Global assessment” are being used interchangeably during this article). We believe that in our domain of websites, CRM can broaden our horizons about the way we understand the interaction that users’ experience. It can help us finding the patterns of affective reactions to websites and its relation with overall judgements. Furthermore, we will have chance to explore the feasibility of this technique while applied on websites and figure out the possible problems that may occur. Therefore, we believe that future research may benefit from our findings from the integration of CRM systems to website evaluation.

Various quantitative and qualitative techniques have been used in user experience research in website context such as questionnaires, analyzing click-through behaviour, eye-tracking, think-aloud method, interviews, web analytics, heuristic evaluation etc. These techniques have been applied either alone or together with other techniques. To get the most out of these techniques, Nielsen et al. (2002) states that using only one does not suffice. Rosenfeld & Moville (2002) also advise integration of multiple research methods in order to get better results. In this study we will make an attempt to explore “User Experience in Websites” by collecting data from three different sources, namely continuous response measurement, eye-tracking and questionnaires. The data gathered from different sources will be combined to see the benefits of introducing such an approach. For instance, eye-tracking data may help to see how problematic is that to use the CRM tool for users while interacting with the website. Do they have to look at it each time they use it? By combining the data, we can also figure out the patterns of eye
fixations before users provide a negative or a positive feedback about the website. Finding such patterns like increasing or decreasing eye fixation durations may enable us to notice automatically if users are getting frustrated or pleased. We believe that future adaptive systems have a lot to benefit from such quantification of combined user data coming from different techniques integrated with users’ emotions.

When it comes to user experience, it is more than just “liking” or “hating” a website. Users’ judgement does not have only one single dimension but various dimensions. The literature suggests that there exist different components of our assessments about a technological product such as Perceived Usefulness, Perceived Ease of Use, Perceived Hedonic Quality, Perceived Visual Attractiveness (Davis, et. al., 1989). With this study, we will try to confirm that these dimensions also exist for website user experience domain. Furthermore, we will use these different dimensions to make a better sense of our Continuous Response Measurement data. It is believed that this investigation will help us to understand how websites differ in separate dimensions and how these differences affect the global assessments of the participants.

To sum up, we will try to come up with an understanding of user experience which will help us to design successful websites. In order to achieve this goal, we will be following an integrative approach using Continuous Response Measurement, Eye-Tracking and Questionnaires. In the following paragraphs, we will first introduce User Experience and provide the definitions of the concepts those are being pronounced in the literature. Afterwards, we will discuss User Experience in context of World Wide Web and talk about how these concepts are addressed in WWW. Later on, we will provide a brief overview for each method that we used for this study and present our research questions.

1.1 User Experience

Even if there are different definitions depending on the perspective of the author, let us introduce one of the most explanatory ones. Hassenzahl and Tractinsky, (2006) define user experience as follows; "User Experience is about technology that fulfils more than just instrumental needs in a way that acknowledges its use as a subjective, situated, complex and dynamic encounter. User Experience is a consequence of a user’s internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.)". As this definition suggests, User Experience is more than just “Usability”. Since “Usability” and “User Experience” are commonly confused terms which are being used interchangeably, a
separation is thought to be necessary before going into more detail. Furthermore, we will try to comprehend this separation and try to understand this buzzword (Hassenzahl and Tractinsky, 2006) in depth by reviewing the literature on User Experience. We will also be more specific and try to understand User Experience in World Wide Web context.

As technology matures, usability is no longer adequate to explain the interaction between users and systems. Hence, in order to reach a broader understanding of Human-Technology Interaction issues, the notion of “User Experience” was introduced. Before it came into usage, Usability was already recognized as an important phenomenon. International Standards Organization (ISO) defines usability in Part 11 of the ISO 9241 standard (BSI, 1998) as; “The extent to which products can be used by specified users to achieve specified goals with effective, efficiency and satisfaction in a specified context of use”. As we have provided definitions which comprise only a very limited fraction of the documentation on User Experience, it goes without saying that there are many more options presenting different concepts and definitions regarding these two words. With respect to the concepts that are mentioned in different, yet overall accepted definitions, we see that usability has dimensions such as (Zhou, 2007); usefulness, learnibility, effectiveness, flexibility, efficiency, memorability, satisfaction, understandability, operability, attractiveness etc. On the other hand, User Experience does not only focus on the task related issues but takes it one step further.
Figure 1: Three Aspects of Technology Usage

As the figure above illustrates, Mc. Namara and Kirakowski (2006) divides the aspects of technology usage into three. Functionality refers to the technical issues and explains the systems characteristics. It is stated that usefulness of the device features, maintainability and reliability can be addressed in this aspect. On the other hand, usability is characterized by interaction between users and the products. It refers to the threshold to use and accessibility of the functionality. Finally, User Experience presents a wider relationship between products and the user. It has its own dimensions like flow, challenge, fun, excitement, perceived attractiveness, perceived usefulness etc. This is exactly the point where the idea behind this study was elicited. Advertisement B.V. Netherlands, the company that co-operates in this study, is an e-commerce consultancy company which came up with the goal of understanding User Experience in depth. As Advertisement puts it, “clients are pronouncing User Experience more and more in last couple of years”. Even though not everyone is sure about what User Experience is, the concept is growing more and more popular. In order to provide a better consultancy which goes beyond the Usability, Advertisement decided to cooperate in this study to have a better insight of User Experience in web domain. Thus, for the rest of this literature overview, we will be focusing on User Experience.
In the following paragraphs, we will provide an overview of the concepts and their definitions which are pronounced with respect to User Experience and Web Acceptance Model;

*Perceived Usefulness:* It is the subjective probability of the prospective users indicating to what extent using a specific application system will increase his or her job performance within an organizational context (Davis, Bagozzi and Warshaw, 1989). It has been verified that perceived usefulness is positively related to the frequency of internet usage, daily internet usage and diversity of internet usage (Teo, Lim and Lai, 1999).

*Perceived Ease of use:* It is the degree to which the prospective user expects the target system to be free of effort (Davis et al., 1989).

*Playfulness:* As a human characteristic, it is the tendency of a person to have fun. In this context, individuals who are more playful are hypothesized to be more intrinsically motivated to use the World Wide Web (Atkinson and Kydd, 1997).

*Perceived Enjoyment:* The perceived enjoyment was found to be positively related to frequency and the daily usage of the internet (Teo, Lim and Lai, 1999).

*Relevance:* An estimate of appropriateness that exists between information provided and information used as judged by a person (Shih, 2003). Relevance was found to be a strong determinant of perceived ease of use, perceived usefulness, user attitudes toward the internet and perceived performance.

*Perceived Hedonic Quality:* Comprising quality dimensions with no obvious relation to the task user wants to accomplish with the system, such as originality and innovativeness (Mahlke, 2005).

*Perceived Visual Attractiveness:* As the name suggests, it is the overall feeling of the user about the website in terms of aesthetics (Mahlke, 2005).

*Experiential flow:* The optimal and enjoyable experience in which we feel “in control of our actions, masters of our own fate. We feel a sense of exhilaration, a deep sense of enjoyment” (Huang, 2003).

*Engagement:* A state of mind that we must attain in order to enjoy a representation of an action. Additionally, it entails a kind of playfulness (Karat, Pinhanez, Arora and Vergo, 2001).

*Play:* Behaviour motivated by the need to avoid boredom and maintain arousal (Karat, Pinhanez, Arora and Vergo, 2001).

*Challenge:* A demanding or stimulating situation.
Research so far demonstrates that perceived usefulness appeared to be the main factor behind the intention to continue visiting a website (Castaneda, Leiva and Luque, 2007 and Davis, Bagozzi and Warshaw, 1989). Perceived ease of use was also found to affect intention significantly, even if it is not as a strong factor as perceived usefulness (Davis et al., 1989). Furthermore another study indicated that different dimensions which influence “the intention to use” are found to be independent such as perceived usefulness, perceived ease of use, perceived hedonic quality and perceived visual attractiveness (Mahlke, 2002 and Mahlke, 2005). Heijden (2003) conducted an experiment using not only students but with real website visitors to understand the factors behind the intention. He found the same results concluding that intention is most dominantly influenced by attitude. It was stated that “Based on certain beliefs, a person forms an attitude about a certain object, on the basis of which he or she forms an intention to behave with respect to that object.” Moreover, ease of use, enjoyment and usefulness are shown to contribute equally to the attitude towards using and visual attractiveness is also proven to be a factor in addition to the factors above.

To summarize, the most commonly pronounced and empirically validated factors of User Experience appear to be as following: Perceived Usefulness, Perceived Ease of Use, Perceived Hedonic Quality, Perceived Visual Attractiveness and Intention to Use. Since these concepts occur in different combinations in different studies, our first research question will consist of the attempt to figure out the significant factors from the list, as presented above, in the context of our study. All the phenomena which are provided with definitions are set of concepts collected from User Experience or Technology Acceptance Model literature. In this case, we will be specifically looking for the dimensions in User Experience of websites.

The World Wide Web, a source which provides access for all sorts of information, presents a wide variety of experiences in terms of emotions and impressions. This interaction with the WWW can be described as useful, enjoyable, challenging, frustrating, entertaining etc. While surfing the internet we are not only performing our tasks, but we also have an experience which is influenced by the aspects of the website, the conditions of the room that we are in, the computer that we use, our psychological state, the inherent knowledge that we bring, our ability to use computers, a former experience about the website etc. As we can relate user experience to other domains such as video games, software, specific products etc., this list can be prolonged easily. Each domain has its own factors, own concepts and emotions. Hence, we will focus on User Experience in the World Wide Web domain. Firstly a short history and a summary about the WWW will be presented in the following paragraphs.

About 25 years ago, the internet as we know it was developed by the National Science Foundation, and the World Wide Web was created 20 years ago by an English
physicist Tim Berners Lee. After all those years, the internet has come to play a big role in the lives of millions of people and it has caused profound changes with respect to entertainment and the way we communicate with each other. Nowadays it even seems impossible for people to live without it since it is represented in almost every aspect of our life. We frequently hear that the world is a “global village” and its impact in our lives is considered as a revolution. It has modified the ways we share information and how we make business; we discuss on the internet, we shop, and we have access to information. In short, internet has made our lives easier, and the key to its success is that it is available almost anywhere and anytime. Millions of websites are out there providing different functionalities and opportunities, waiting for us to interact. To put it with a cliché expression, it is just a few clicks far from us to make our lives easier. Or is it? We can see that this is not always the case. Even after its 20 years of history, most of the websites still have lots of problems. Websites administrators, who are aware of the importance of optimizing their websites, are looking for new ways to develop them in order to keep up with their intended goals.

“Motivations behind internet usage” is another main direction of World Wide Web research. Since motivation is the prerequisite of an action, various studies have been conducted in this field. Some people use internet as a part of their “work” and some use it for “fun”. Internet usage behaviour is based on different motivations and these motivations strongly affect the users’ behaviour. Usage behaviour also differs among individuals and the conditions. In general, there are two basic types of motivation mentioned in literature: Extrinsic and Intrinsic Motivation (Castaneda, Leiva and Luque, 2007). Extrinsic motivation focuses on the functionality of the website and it is more goal-directed. With regard to intrinsic motivation, usage of website is an aim in itself and it is more hedonistic and exploratory. Depending on the website, the motivation behind the usage may differ. For example, the motivation to use an e-commerce website is more extrinsic than intrinsic (Castaneda et al., 2007). An entertainment website on the other hand works the other way around. Nevertheless, some of the websites that we use to reach information can combine of both extrinsic and intrinsic motivations. In addition, different aspects of the motivation are related to different factors behind the usage (Castaneda et al., 2007). For example, extrinsic motivation is more related to perceived usefulness but, on the other hand, intrinsic motivation is tightly coupled with perceived ease of use. Teo, Lim and Lai (1999) found that perceived usefulness for local internet users is the main reason for using the World Wide Web. Secondary reasons appeared as perceived enjoyment and ease-of-use. There exist various reasons to use internet and each of them put a different weight in the scale with respect to usage behaviour. Furthermore these weights are different for different types of websites. As we have tried to emphasize so far, user experience on websites is a broad and multi-faceted
phenomenon. We will try to achieve a profound understanding by approaching the problems from different perspectives and combining different techniques.

For the purposes of this study, three websites which are active in Car Insurance business in the Netherlands has been selected (namely Unive, Ditzo and AllSecur). There are several reasons that these websites have been selected. First of all, these websites are the customers of Advertisement B.V and they are operating at the same interesting business. Their functionality is more or less the same which makes them compatible with each other. Since our comparisons will be among the websites, these websites provide as enough features in common. Apart from the fact they perform the same function, these websites differ in their appearance, content, functionality, percentages of text and images used etc. This provides us enough variance which is needed for the comparison. Another advantage of using these websites is that they are not very well known by the participant group who are students in Eindhoven University of Technology. Using well-known websites like social networks would reduce the validity of the experiment. The figure below presents the homepages of all three websites;

Figure 2: Websites used for the experiments: Ditzo, Unive and AllSecur.

All the factors mentioned in the paragraph above can change the impression of the users in different ways. Even if users have an overall impression about the website after using it, still this overall impression has separate components like perceived ease of use and
perceived hedonic quality. Next step of this study is figuring out in which dimensions did the websites we have used differ, if they differ at all. First of all, we will be investigating if any of the website is superior to others in terms of overall impression ratings provided by users. Then we will try to figure out if any dimension causes this difference. In conclusion, we formulate the following research question:

**Research Question 1: What are the dimensions in User Experience of websites and to what extent the websites can differ in these dimensions?**

The way we think and feel about a product, a person or a website is highly influenced by our interaction with these concepts in reality, and this may change our impressions in a positive or negative way. When we drive the car that we have just bought for the first time, our impressions and expectations turns into a real experience. Imagine that you meet a fellow who you had been seeing around but have not met before. As soon as you get to know each other, real interaction replaces the prejudices and stereotypes. It is a fact that real interaction is the most important part of our attitudes towards a person or a product. However, this is not the entire story. It occurs frequently that we already possess an image of certain websites before we have even used them. As we are daily confronted by commercials of certain companies as presented in the media, it is highly probable that we are already familiar with them. In our attempt to understand the User Experience, it is therefore considered interesting to see the effect of so-called *familiarity*. We will be investigating to see whether or not the familiarity of the website *(before starting the experiment)* affected users’ global assessments.

Another interesting phenomenon would be the effect of the order of seeing the websites. While searching for something on the web, even the average contemporary internet user takes a look at multiple web pages. When we want to buy a flight ticket for instance, we do not usually decide on one single offer from one single company. Instead, we tend to take a closer look at the alternatives to see what our options are. Therefore, it is expected that the same attitude will hold while choosing the right insurance for your car. In this study, participants are being presented three alternative car insurance websites where they can select the most suitable and reasonable option for them. The scenario we present to the participants is supposed to recreate real life circumstances, except for the fact that they cannot choose the order in which they will visit the websites. They are obliged to follow the order that is presented to them. At this point, we should consider the effect of the order of presentation. There are several phenomena which suggest that the order of the items presented to the participants may influence the final outcome. The *serial position effect* (Hermann Ebbinghaus) for example states that the recall accuracy varies according to the serial position of the item. The *Novelty effect* suggests that the performance or the liking of the user can initially improve when a new
technology is presented, not because of the enhanced performance of the technology but because of the increased interest of the user about the new technology. This may result in a positive bias towards the first website shown to users. As there are enough reasons to believe that order can influence the outcome, we will be investigating whether or not \textit{the order of the presentation of the websites has an effect on users’ global assessments.} Putting the methodological questions about order and familiarity together, we formulated the following question;

\textbf{Methodological Question 1:} Can we use familiarity or the order of presentation to predict the global assessments of the users about the websites?

1.2 Continuous Response Measurement

Due to interaction with a website, internet users experience a change in their cognitive and emotional state of mind. Everything starts with a first impression which further continues to change and develop from then on. It is very common for researchers to ask users to evaluate the products after they have used them. On one hand this is an effective method which gives an overall impression users have of this particular product, but, on the other hand this same method fails to understand the processes users went through.

The interaction between users and the websites is not static; it changes over time depending on what happens during the interaction. Users for example get frustrated and even annoyed, or they see something they like. They read, click around and try to make an accurate decision about the credibility of the company. They either achieve their goals or fail to achieve. Their attitudes and moods change over time continuously. Thus, it is believed that a dynamic measure distributed within a certain amount of time can provide a better insight in a dynamic interaction between users and the websites. As Biocca et.al. (1994) states, continuous response measurement (CRM) systems allow participants to constantly provide feedback about their changing mental states, evaluations and opinions. This feedback is valuable especially if the researcher is interested in exploring the theoretical issues regarding the cognitive processing of continuous messages. In the following paragraphs, a brief summary will be provided about Continuous Response measurement systems and their history.

The history of CRM can be traced back as far as 1932 when Millard invented this concept. Later on, "program analyzer" was introduced by Paul Lazarsfeld in 1945 which is considered as the first mature form of such a system. Ever since, new research areas where CRM was used emerged, such as interpersonal communication (Palmer and
Cunningham, 1987), public speaking and performance (West, Biocca and David, 1991), advertising (Hughes, 1992), television programming (Biocca et. al., 1992), political communication and health communication. The purpose of CRM in these research areas vary firstly, in obtaining a self-report of the users concerning their fluctuating psychological state, secondly in measuring the constantly changing message content, and thirdly, in coding communication behaviour (Biocca, et. al., 1994). As previously mentioned, the purpose and the measures used depends on the area of interest to a great extent. Baumgartner et al. (1997) used CRM systems to understand consumers’ global assessments of advertisement episodes. The purpose of the research was to explore the structure of the consumer’s continuous emotional responses towards advertisement episodes, in order to investigate whether some patterns of affective feedbacks lead to an enhanced positive impression with respect to the advertisements or not. Therefore, it is relevant to ask the following question: do people focus on certain key moments to make an overall judgment about the advertisement? For this purpose, participants were shown advertisements and were asked to provide continuous feedback about their liking of the advertisements. After collecting the data, the following predictor variables have been used: the peak experience, the final experience, the sum of momentary experiences, coefficient of the linear trend, duration of the advertisement, time spend until peak point is reached and the time after peak point is reached. Results verified the fact that consumers focus on certain key moments of their interaction. Peak impression and the last impression appear to be the best predictors for the overall impression of the customers. Even though the domain of advertising and websites do not overlap for one hundred percent, the principles of these two studies are interchangeable. Both watching advertisements as well as exploring websites create different mental states in course of time. Hence the results of Baumgartner et al. (1997) are to a great extent promising for our research. For the purposes of our study, we will be using the following predictor variables to understand the determinants of the overall impression of the websites: first impression, average impression, weighted average of the impressions, negative peak impression, positive peak impression and the last impression.

Another set of possible predictors of the global assessments of the users is the quantifiable website usage metrics. The definition of usability depends heavily on how we measure it (Zhou, 2007). We can use binary task completion or errors made as a unit of measurement in a website. It is believed that the same holds for User Experience as well, therefore, the concept that we come up with as User Experience depends on how we measure it. In their extensive study, Hornbaek (2004) reviewed 180 papers from HCI journals which propose several quantitative metrics. Some of the most frequently used quantifiable measures are: Binary task completion, Number of Clicks, Accuracy, Error rates, Recall, Page Views, Unique Page Views, Completeness, Quality of outcome, Task completion time, Time spent until a certain event, Usage patterns, Use frequency,
Information accessed etc. There are various tools and methods that can be used to measure each or subset of the measures listed above. Considering the tools that have been used for this study and their capabilities, we will focus on “Unique Page Views”, “Number of Clicks” and “Time spent during the task”. These metrics have already been introduced as part of the so-called “web analytics” tools. At this point, we will try to examine whether or not quantifiable web usage metrics can predict users’ overall assessments.

Besides the straight-forward measures that can predict the overall impression of the users, we can bring the analysis one step further and use derivatives to see if the amount of the increase or the decrease also matters. Imagine two patterns of continuous response data in which one of them is slightly going up and down, the other wildly varying around the mean reaching the borders of the physical device that measures the response. Or let us consider the two figures presented below. It is obvious that the experience of the participants were quite different in terms of number of data points, smoothness of the lines etc.

Figure 3: Two examples of continuous response measurement data where vertical axis represents the slider position (going up means a positive evaluation) and horizontal axis is the time.

Eventually, these two patterns in the end may result in the same average, weighted average, the first or the last impression. When we only consider these as the predictors, we will not see any difference among the patterns, even if the conduct of the users is clearly different. Therefore one of our goals is to explore whether or not the magnitude of the changes of users’ momentary assessments can predict their global assessments. All in one, we formulate the following research question:
**Research Question 2:** Based on the “continuous feedback” provided by the users, what are the significant predictors of the global assessments of users?

The users’ personal characteristics can deviate greatly. Website users are not just one bunch of people who decide exactly with same reasoning and react to the same stimuli in the same way. They differ in their background, age, computer ability, internet usage etc. Instead of labelling every user in the same category, trying to figure out different patterns of their decisions may provide us a deeper insight about their experience. Considering the different variables predicting the global assessment of the participants, there may exist different types of users whose final evaluations about the website are either determined by their last impression or the first impression. Moreover, there may be a group of participants who has a tendency to create an overall judgement based on their positive or negative peak experience throughout the time they spent on the website. In conclusion, it is believed that seeing whether or not such a categorization is possible would be an interesting research question for the project.

Another participant characteristic would be the tendency of users towards compensation behaviour. Compensation is a strategy in psychology which can be summarized as follows: One can consciously or unconsciously conceal his or her feelings, desires, weaknesses or frustrations. By doing so, the negative effect of the compensated emotions is diminished. Traces of compensational behaviours can be seen in our daily lives. If we overreact to a friend after having been disappointed about a certain issue, we may compensate our behaviour by behaving nicer than we used to do. If we realize that we are about to cross a personal border while interacting with a fellow person, we may take a step back to compensate this behaviour. We are hypothesizing that the same compensation behaviour may hold for our research as well. It may be the case that, as the user provides a positive or a negative feedback, the next behaviour may be compensation behaviour in the opposite direction. In order to reach a balance, participants may tend to give a positive feedback just after providing a negative one (or the other way around). Thus, testing this phenomenon will be one of our interests in this project.

People tend to respond differently to different situations. We smile more often to the people we like, our pupils get larger when we see them. We wear the shirt we like more often than the one we like less. In short, the frequency of our affective responses to people or daily life objects differ depending on whether we like them or not. The question that needs to be posed is, *can it also be applied to websites?* While participants use the websites presented for this study, they are not obliged to give feedback about the website in certain time intervals. Therefore, the number of feedback they provide may be determined by their liking about the website. In order to see to what extent this
reasoning is correct, we will investigate whether or not people provide more or less feedback with respect to their like or dislike of websites. All put together, characteristics or behaviour patterns of users in our website user experience domain are one of our primary concerns. Thus, we formulate the following research question;

**Research Question 3 (User Characteristics):** Can we find behavioural or characterstic patterns of users by making use of the Continuous Response Measurement in website context?

One of the most important considerations in human-technology interaction research is the use of non-obstrusive techniques. If the effort spent on the use of the tools distracts the users, then the validity of the experiment should be questioned. Therefore, effort should be put on preparing the experimental set up in such a manner that users engage with the measurement tool in the hand and perform their main task without any disturbance. To find out whether or not participants are interrupted while using the slider is important for both our study and future research on continuous response measurement systems. For example, looking at the slider while they interact with it can be a possible disturbance for the participants of this experiment. Since the eye-tracker provides us accurate information about where participants are exactly looking at, there exist the possibility to check if they are looking at the slider while they are interacting with it. Thus, we will be making use of this opportunity to see if looking at the slider is a distraction for the participants.

Another potential distraction of the slider that we are going to investigate is its limited range. Types of devices that can be used to gather continuous response can vary, depending on the field of interest and the setting of the experiment. The device can be either digital or analogue; the latter even contains different types such as sliders, buttons or dials. For this study, slider type CRM system has been used. Each device has its own advantages and disadvantages and sliders are no exceptions. One of the disadvantages of the slider is its physical boundaries. There is an upper and the lower limit that slider can reach. This limit creates a problem when one of the boundaries is already reached by the participant. The user cannot provide a negative feedback when the lower limit has already been reached. The same logic can be applied the other way around as well. In order to examine the extent to which this physical limit is problematic for participants, we will be investigating whether or not participants are reaching the borders of the slider in upper or lower directions. In conclusion, we formulate the following question;

**Methodological Question 2: Is the Continuous Response Measurement tool (namely Slider) obtrusive or distracting for the users?**
1.3 Eye-Tracking

Eye tracking is a sophisticated technique which monitors the eye movements, fixation durations, the order of the fixations etc. The collected data is analyzed through software in order to detect metrics such as mean fixation duration, gazing time, saccade rate, scan path etc. (Pan et al., 2004). The software usually provides data visualizations, like heat maps that show the parts of the system which gets more eye-fixation. Heat maps are supposed to indicate the extent to which attention has been directed on that part of the screen. Additionally the order of the fixations and the spatial arrangement is another topic of interest. There are various domains such as perception, cognitive linguistics and product design which make use of eye-tracking visualizations and raw data for research purposes.

In order to provide a cognitive basis for the interface design of web pages, it is more important to understand how users view the pages (Pan et al., 2004). The history of eye-tracking dates back from almost 100 years ago, when computers did not even exist yet (Jacob et al., 2003). Even if the promises of the eye-trackers are not fully achieved for the moment, it is getting more and more popular for researchers who are working on web page viewing behaviour. These studies include analysis of scan paths on web pages (Josephson & Holmes, 2002), tracking down the determinants of web page viewing behaviour (Pan et al., 2004) and analyzing ocular behaviour, to see whether or not web users complete their tasks on a web portal (Goldberg et al. 2002).

As the eye-trackers get cheaper and become easier to use, they are getting more and more popular in human-technology interaction research. Most of the eye-trackers on the market are still expensive but there is a substantial improvement compared to previous years and it is reasonable to believe that they will get more affordable in the near future. This will increase their availability for user experience researchers and enable them to have a deeper insight about the behaviour patterns of web users. In addition to that, even webcams may become available extensively to be used as eye-trackers. If this becomes reality, it will be possible to collect a large amount of online data from real users. Moreover, the ecological validity of collecting gaze data will also increase, since users will be providing data from their actual working environments. This collected data can be used together with web analytics to answer ‘when’, ‘what’ and ‘how’ questions. To give a satisfactory answer to the ‘why’ question, we still need subjective report from the users. Atkinson (2007) states that eye tracking can be used to provide more detailed information which makes it easier to understand the user experience better. The information about how users interact with the web page is not available in web logs (Atterer et al. 2006), and eye-tracking may help to fill this gap in making a meaningful understanding of web analytics data.
Eye-tracking provides detailed and quantitative data for user experience research. On top of the traditional usability metrics, eye-tracking gives the opportunity to explain the results of other user tests better. Users do not need to be trained to use the equipment because it is quite straightforward for them. It can detect both where the attention of the user is focused as well as the intensity of the focus. Moreover, from another point of view, eye-tracking is a useful technique to see what types of content do users find compelling or simply ignore. In short anything that is important for a website owner to know.

We believe that the eye-tracking data combined with questionnaire and continuous feedback data will convey different perspectives of website user experience that have not been studied before in this combination. For instance, fixation duration has been a subject of interest in many different domains like usability research, cognitive science, advertising etc. Fixation duration is an indicator of processing time of the stimulus and can also be an indicator of participants’ cognitive states. Velichowsky et. al. (2001) has shown that, when participants are presented two stimuli, one of which is the target and the other one is the distractor, the fixation duration of the participants increased significantly. An indirect conclusion from this research may be the possibility that frustration can have an effect on the fixation duration and furthermore this can provide us some insight about the cognitive states of the users. We can assume that there is a change in participants’ mental state when they decide to provide a new feedback using the CRM system. They may see something they like or they may be frustrated, in any case, they change their previous evaluation about the website. There is reason to believe that something happens with respect to the interaction between users and the website, before they change their assessment (literally moving the slider). In order to see if the reason that makes the user change his or her mind about the website has an effect on the average fixation duration, we will be investigating whether or not average eye fixation duration of the users differ before they provide any feedback about the website.

On the other hand, giving a positive or a negative feedback may also differ in terms of its effects on fixation duration. The possible effect on fixation duration may appear before the slider is used to provide a negative feedback, but this is not necessarily the case for the positive feedback (or the other way around). Therefore, it would be interesting to see whether or not the average eye fixation duration of the users differ before participants provide positive (or negative) feedback about the website. In order to test these phenomena, we formulated the following research question;

**Research Question 4:** Can we find any clue from eye-fixation behaviour indicating that participants are about to give a negative or a positive feedback about the website?
As it is the case for the average fixation duration, another possible indicator that signals a change in the mental states of the users may be the “total” or the “average distance travelled by the eyes”. It may be the case that users who are looking for something and fail to find it can become frustrated. While they keep on searching, it is possible that the total travel distance of the eyes will increase and hence this will result in an increase of the average eye-travelling distance. If the presence or absence of such a frustration affects the variables mentioned above, they may be a predictor of the overall impression of the users as well. To test this phenomenon, the following question is formulated:

**Research Question 5: Can we predict overall impression from quantitative eye-fixation metrics?**

1.4 Combining the data

This study has three different sources of data for each particular subject, namely questionnaire, slider and eye-tracker. Each data source by itself brings research opportunities for our context of user experience in websites but moreover, the combination of these data is in my opinion one of the most interesting parts of this research. For instance, the dimensions of the user experience that appeared after exploring the questionnaire data can be combined with the data gathered from the slider. As it was mentioned before, we assume that our overall impression about a website has some components such as perceived ease of use, perceived hedonic quality etc. On the other hand, slider data provides us data about the experiences that users go within a certain amount of time. We have already formulated the question “What are the significant predictors of the overall impression based on the continuous feedback provided by the users?” Even if we find such significant predictor variables, we can take it one step further and ask if different dimensions of user experience have different predictor variables. For example it may be the case that perceived ease of use is very well determined by a predictor value (first impression, last impression etc.) however, perceived hedonic quality shows a totally different pattern. In order to test this phenomenon, the following question is formulated, which can be answered by combining continuous response measurement data and the factor analysis of the post-task questionnaire:

**Research Question 6: Is it possible that different predictors affect different dimensions of User Experience in different magnitudes?**

In the next chapters, we will discuss the experimental setting and provide details of the design. Equipments will be introduced as well as the flow and parameters of the study.
2 Data and Methods

2.1 Design

A within-subject experiment was conducted where participants viewed three websites and were asked to rate them using Continuous Response Measurement post-task self-report questionnaires. In addition, eye-tracking technology was used while participants were interacting with the websites. The order of the presentation of the websites was randomized for each participant.

2.2 Participants

30 Students of the Eindhoven University of Technology participated in our experiment (8 Female, 22 Male). The ages of the participants was varying between 18 and 30 with an average of 22. All of the students were either undergraduate or graduate students. Most of the participants reported that they use internet for several times a day. All the participants in this study were Dutch but they all spoke English fluently. The latter language was used to read the instructions; Dutch was used for the websites. A compensation of 10 Euros was handed to each participant for their participation.

2.3 Dependent Measures

The dependent variable for the continuous response measurement tool (namely slider) was supposed to be a number between 0 and 1000 where 0 indicates the negative extreme and 1000 indicates the positive extreme. Due to the physical limitations of the tool in hand, the active range was actually between 30 and 990.

Dependent variables for the eye-tracking tool were fixation duration and the number of fixations. These variables (together with the slider data) have been time stamped in order to synchronize them in the data analysis phase.

2.4 Equipment

Three main technologies were used for this study. The Tobii x120 eye-tracker technology was one of the tools. It is a new generation automatic and highly accurate
eye-tracking tool compared to the early versions of such tools. Providers claim to have solved major problems of eye-tracking like cumbersome equipment, poor precision and limited tolerance to head motion. Additionally the tool is not head-mounted but located on the desktop instead. The experiment was not obtrusive for the participants although they were required to stay in a frame and ordered not to move too much to avoid deviations of the equipment (See figure 4 below).

![Tobii eye-tracker x120](image)

**Figure 4 : Tobii eye-tracker x120**

The second tool used for this experiment was the slider which enabled us to collect continuous response from users. Slider has a five-point scale in Dutch namely, Uitstekend (excellent), Goed (good), Voldoende (sufficient), Matig (moderate) and Slecht (bad). It has the sampling rate of 100 msec. The position of the slider (ranging from 1 to 1000 where 1 is the bottom and 1000 is the upper point) and the time stamp was being recorded in a text file.

The third tool used for this study was UserZoom a remote usability testing tool. It is a web-based tool which enables the researchers to quantify and conduct user experience research by providing links to the customers (to be used in a browser with certain features) and ask them to complete the tasks by following the instructions. Besides various functionalities of the tool, we used it for combining our pre-task and post-task questionnaires together with the eye-tracking equipment. By doing this, we enabled the users to be able to fill in the questionnaires and perform their tasks one after another sequentially without any breaks in between.

The tools listed above were connected to a desktop computer environment. The whole experimental setting can be seen in the figure below. In order to synchronize and process the data coming from different sources, Java programming language was used.
2.5 Questionnaire

In order to understand the characteristics of the participants and their opinions about the websites, two questionnaires were applied. The details and the purposes of the questionnaires will be explained in following section.

2.5.1 Pre-task Questionnaire

Before the participants started exploring the websites, they were asked to fill in a pre-task questionnaire which aims to collect data about participant characteristics. The points of interest for this questionnaire are;

- Age
- Gender
- Education level
- Frequency of internet usage
• Daily amount of time spent on internet
• Diversity of internet usage
• Ability to use internet
• Familiarity about the websites presented
• Owning a driver licence or not

The frequency of internet usage, the daily amount of time spent on internet, the diversity of internet usage and the ability to use internet comprise more than one questions. These questions were taken from the early validated questionnaire by Teo, Lim and Lai, (1999). The full version of the questionnaire can be seen in Appendix 1.

2.5.2 Post-task Questionnaire

After exploring each website, participants were asked to fill in a post-task questionnaire about that website. With this questionnaire, the main goals were (a) to get an overall impression of the website which will be later on compared to the continuous response measurement data and (b) to examine the dimensions (e.g. perceived usefulness) of website user experience. The list of topics in the questionnaire were:

• Perceived ease of use
• Perceived visual attractiveness
• Perceived usefulness
• Perceived hedonic quality
• Intention to use
• Overall impression about the website

These topics also consist of multiple questions. Perceived ease of use and perceived usefulness were defined in Davis, Bagozzi and Warshaw (1989). Hedonic quality is defined in Hassenzahl (2001) and Visual attractiveness is from Heijden (2003). Also, all of the factors were presented in the questionnaire from Mahlke (2005). Full version of the questionnaire can be seen in Appendix 2.

2.6 Procedure

Sessions were held in general purpose lab in Human-Technology Interaction Department at Eindhoven University of Technology in September 2009. Within subject design was employed among 30 participants. Data was collected from 3 different data
sources, namely: Questionnaire, eye-tracker and the slider. The flow of the experiment was as follows:

![Flowchart image]

**Figure 6: Flow of the experiment**

First, participants were informed about the experimental setting and the flow of operation. Afterwards, they were also informed about how to use the equipment. They were verbally assured that their personal data will not be used for any purpose and verbal consent was requested. Afterwards, eye-calibration was performed to see whether participant is suitable for the experiment. During the calibration, also the distance and the range of the screen are also calculated by the eye-tracker to get more accurate results. Participants were asked to follow moving points on the screen and calibration results were examined if the success rate is good enough to continue with the experiment. Participants with special conditions like the ones who wear special eye-glasses were not able to participate in the experiment since the eye-tracking tool was not able to locate their eyes. As soon as the eye-tracker equipment and the slider were started, the experiment room was left.
As soon as the experimental session starts, participants were asked to fill in the initial pre-task questionnaire which aims to collect data about participant characteristics. The content and the aim of the questionnaire are explained in the introduction part and the full content is available at Appendix 1. After filling in the questionnaire, first website was presented to the users. Users were requested to imagine themselves as they have just bought a new car and the specifications (model, licence plate etc.) were provided to the user. They are told that they are going to be presented 3 car insurance websites and requested to evaluate the websites according to their preferences. They were asked to click around the website, search whatever is interesting for them and they were given the freedom to decide based on their own preferences. They were also told that they will be asked to pick the best option for them after completing the experiment. In the meantime, participants were requested to provide feedback using the slider as frequent as possible. It was emphasized that the feedback expected from them is the overall impression about the websites.

After evaluating each website, a post-task questionnaire was presented to each user in order to understand the dimensions of the experience they had and also receive their overall impression about the website. The aim of the post-task questionnaire is explained above and full version can be seen at Appendix 2.

After being finished with evaluation of all the websites and filling in the questionnaires, a final post-task interview has been applied to all participants. Questions were aiming to see what the important considerations in their decisions are and what make them move the slider. Their opinion about each website is asked. Also interaction of the participants with the slider tool was one of our interests.

3 Results

In order to start with, it should be noted that not all collected was usable for the data analysis because of some technical problems. One of the questionnaire samples was corrupted and could not be saved. The eye-tracking data collected from 7 participants were not usable for different reasons. In two cases, the session had to be interrupted and the data was not usable any more since it could not be synchronized with the other data sources. For 5 cases, the quality of the recording was not good enough due to the reasons like special kind of eye-glasses worn by the participants. In the case of the continuous response measurement tool, namely slider, it was not possible to use its data gathered from two of the participants. After all, the number of usable participants from all data sources was as follows:

- Questionnaire: 29
• Eye-tracker: 23
• Slider: 28

Note: Please note that this study was based on repeated measures design. Same subjects participated for the evaluation of 3 different websites. For some of the research questions that we have asked, the three different conditions that participants were being presented appeared to be independent from each other (like their overall impressions about websites as a dependent variable). But for some of the cases, the data we collected failed to pass this test and it seemed that measurements were interdependent (eye-fixation duration measures). We will mention when this repeated measure design becomes problematic so that we will avoid making hard conclusions from the results of corresponding analysis. Also it can be noticed that N (number of cases included in the analysis) changes from one analysis to another. That is because of the difference at removal of the corrupted data for each analysis.

3.1 Processing the data

Data collected from the slider and the eye-tracker was in text file format which were raw and needed to be processed before use. The first step of processing was to synchronize the data sources. Since technically it was not possible to start the slider and the eye-tracker exactly at the same time, synchronization was held programmatically by coding in Java language. To do that, screen recordings were analysed first to find the delay between two data sources and then, old times were replaced with the new ones.

After synchronization, most of the other measures have been aso calculated using Java programming language otherwise it would be too much time consuming. One of the samples Java files can be seen at appendix 4. Raw files which have not been processed yet can also be seen at appendix 5 and appendix 6. After processing the raw data, some visualizations have also been applied to have a better overview and moreover to see whether some patterns can be recognized. Here is an example:
The position of the slider (between 0 and 1000) is represented in the vertical axis and time is represented in the horizontal axis. In should be noted that this graph belongs to one participant visiting one website. Each participant has three different set of data of this type. There are also reference lines which show the first, peak, average, weighted average of the slider positions respectively and the overall impression provided in the post-task questionnaire. From now on, we will call the slider position as impression (e.g. peak position will be called peak impression). For those of the readers who are interested, two different kinds of visualizations can be seen at appendix 7.

The methodology that we will follow in the following paragraphs is; firstly presenting the research question as the title and then explaining the data analysis respectively.

3.2 Analyzing the data

**Research Question 1: What are the dimensions in User Experience of websites and to what extent the websites can differ in these dimensions?**

Since the dimensions of the user experience that we come up with will be used for the rest of the data analysis, we start with this part. Exploratory Factor Analysis was applied to 21 questions presented in the post-task questionnaire (seen after each website that user had explored). Factor loadings were calculated by Principle Component Method and Direct Oblimin rotation. Based on the inspection of the Screeplot and Eigenvalues, **4 factors were extracted explaining the 71% of the total variance** (see Table 1).
Table 1: Variance explained by factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>% of Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>47</td>
</tr>
<tr>
<td>Factor 2</td>
<td>12</td>
</tr>
<tr>
<td>Factor 3</td>
<td>6</td>
</tr>
<tr>
<td>Factor 4</td>
<td>5</td>
</tr>
</tbody>
</table>

This solution has a KMO statistics of .884, which is above the acceptance level and significant (p < 0.00). Based on the Structure Matrix, questions with a load over 0.7 were included in the corresponding factor. Based on this criterion, the factors and the questions enclosed in each factor are as follows;

**Perceived Usefulness (Factor 1):**
- I find the website to be useful completing the task.
- Using the website is helpful completing the task.
- Using the website makes it possible to complete the task.
- Using the website enhances my effectiveness in completing the task.
- I would recommend friends to use the website.
- In a similar situation I would use the website again.
- I find it easy to get the website to do what I want it to do.
- If possible I intend to use the website again.
- If possible I predict that I would use the website again.

All the questions those are expected to load in Perceived Usefulness appeared to be so. But also questions which are expected to appear in “Perceived Ease of Use” and “Intention to Use” factors also appeared in this factor.

**Note:** All of the questions and their expected factors based on the literature can be seen in Appendix 2.

**Perceived Hedonic Quality (Factor 2)**
- The website is standard.
- The website is innovative.
- The website is exciting.
- The website is original.
- The website is impressive.
All the questions loaded on this factor were expected to do so. Only the question “The website is boring” was also expected to appear in this factor but this expectation was not satisfied.

**Perceived Visual Attractiveness (Factor 3)**

- Overall, I find that the site looks attractive.
- The design of the website is unattractive.
- The colours that are used on the website are attractive.
- The layout of the site is attractive.

All the questions appeared in this factor were being expected to do so. This factor was the most accurate one in that sense.

**Perceived Ease of Use (Factor 4)**

- I find the website to be easy to use.
- Interacting with the website requires a lot of mental effort.
- My interaction with the website is clear and understandable.
- I find it easy to get the website to do what I want it to do.

All the questions appeared in this factor were also exactly the way it was expected. It should be noted that the question “I find it easy to get the website to do what I want it to do” appeared in both “Perceived Ease of Use” and “Perceived Usefulness” factors. This was the only question that appeared in two factors.

**Note:** The results of this factor analysis were saved as a regression factor score in order to be used in the further analysis.

Before investigating whether or not websites differ regarding the factors presented above, in the first place we investigated if they differ in terms of overall impression. Based on the question “What is your overall rating of your experience with the website?”, Linear Regression Analysis was performed using the dummy variables for websites as predictor variables and overall impression as the dependent variable. Results can be seen at Table 2 below.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.483**</td>
<td>.195</td>
</tr>
<tr>
<td>Website</td>
<td>(Reference: Unive)</td>
<td></td>
</tr>
</tbody>
</table>
As Table 2 suggests, *Unive had significantly higher ratings than AllSecur (p < 0.05)*. Considering other combinations of websites, none of them was significantly superior to others regarding overall impression. After finding that there exist a difference in terms of overall impressions, the next step of the analysis was to figure out which dimension might have caused this divergence. In order to do that, Linear Regression Analysis was performed where Regression Factor Score of Perceived Usefulness factor is used as the dependent variable and the websites as the predictor variables. Results can be seen in the following tables:

**Table 3: Regression analysis predicting Perceived Usefulness by websites**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.121</td>
<td>.182</td>
</tr>
<tr>
<td>Website (Reference: Unive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AllSecur</td>
<td>-0.544**</td>
<td>.258</td>
</tr>
<tr>
<td>Ditzo</td>
<td>.090</td>
<td>.258</td>
</tr>
</tbody>
</table>

**Model Summary**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>87</td>
</tr>
<tr>
<td>R²</td>
<td>0.57</td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05

As Table 3 indicates, *Unive had significantly higher ratings than AllSecur in terms of Perceived Usefulness scores. There is no other significant difference among the websites regarding this dimension.*

The same kind of analysis was held with Perceived Hedonic Quality (factor 2).

**Table 4: Regression analysis predicting Perceived Hedonic Quality by websites.**
As it is seen at Table 4, AllSecur has significantly higher ratings than Unive in terms of Perceived Hedonic Quality dimension. There was no other significant difference among the websites regarding this dimension. For Perceived Ease of Use and Perceived Visual Attractiveness factors, the same analysis was performed but there was no significant difference among the websites as well. On the other hand, it should be noted that regression factor score variables of these four dimensions were significantly correlating with each other (p < .05) which may affect the robustness of the conclusions.

**Research Question 2:** Can we use familiarity or the order of presentation to predict the global assessments of the users about the websites?

Linear Regression Analysis was held to see if the presentation order of the websites is a predictor of the overall impression. Regarding our context First stands for being first website that is presented to the participants. As Table 5 indicates, there is no evidence to believe that order of the website can be a predictor for the overall impression of the participants.

![Table 5: Regression analysis predicting Overall Impression by the order of the websites](image)
Depending on the results from pre-task questionnaire, the average familiarity of the websites appeared to be as follows;

**Table 6: Familiarity of the websites before starting the experiment**

<table>
<thead>
<tr>
<th>Familiarity (mean)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AllSecur</td>
<td>1.8</td>
</tr>
<tr>
<td>Ditzo</td>
<td>2.2</td>
</tr>
<tr>
<td>Unive</td>
<td>2.4</td>
</tr>
</tbody>
</table>

**Note:** The question was “How familiar are you with the website?” and 5 point likert scale was used where 1 stands for “not at all” and 5 stands for “to a great extend”.

In the next step, Linear Regression Analysis was held where familiarity was used as the predictor variable and Overall Impression as the dependent variable. The results are as follows;

**Table 7: Regression analysis predicting Overall Impression by familiarity**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.414**</td>
<td>.196</td>
</tr>
<tr>
<td>Measure (Reference: Unive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ditzo</td>
<td>-.207</td>
<td>.278</td>
</tr>
<tr>
<td>AllSecur</td>
<td>-.586**</td>
<td>.0278</td>
</tr>
</tbody>
</table>

**Model Summary**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>86</td>
</tr>
<tr>
<td>R²</td>
<td>0.052</td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05

At it is seen, participants were significantly more familiar with Unive than Allsecur (p < .05). When Ditzo was set as the reference variable, none of the websites was superior to others. Therefore, only significant difference was between Unive and AllSecur websites.

Next step of the analysis was using familiarity as a predictor variable together with the effect of websites to see if it cancels out the difference among websites in terms of overall rating (see Table 2). There results were as follows;
Table 8: Regression analysis predicting Overall Impression by familiarity and the websites

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.298**</td>
<td>.328</td>
</tr>
<tr>
<td>Measure (Reference: Unive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ditzo</td>
<td>-.053</td>
<td>.278</td>
</tr>
<tr>
<td>Allsecur</td>
<td>-.541</td>
<td>.284</td>
</tr>
<tr>
<td>Familiarity</td>
<td>.077</td>
<td>.109</td>
</tr>
</tbody>
</table>

Model Summary

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>86</td>
</tr>
<tr>
<td>R²</td>
<td>0.066</td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05

These results suggest that when we take familiarity into consideration, the significant difference between Unive and AllSecur disappeared. Unive was rated more than AllSecur regarding Overall Impression but familiarity eliminated this effect.

Research Question 3: Based on the continuous feedback provided by the users, what are the significant predictors of the global assessments of users?

As another set of predictor variables that come from a different kind of data source, firstly the correlations between the predictor variables were calculated. The results were as follows;

Table 9: Correlations among predictor variables gathered from the slider

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>High Peak</th>
<th>Low Peak</th>
<th>Average</th>
<th>Last</th>
<th>Weighted Average</th>
<th>Overall Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>1.00</td>
<td>.530*</td>
<td>.364**</td>
<td>.490**</td>
<td>.165</td>
<td>.510**</td>
<td>.405**</td>
</tr>
<tr>
<td>High Peak</td>
<td>.530**</td>
<td>1.000</td>
<td>.224**</td>
<td>.708**</td>
<td>.436**</td>
<td>.638**</td>
<td>.542**</td>
</tr>
<tr>
<td>Low Peak</td>
<td>.364**</td>
<td>.224**</td>
<td>1.000</td>
<td>.782**</td>
<td>.654**</td>
<td>.699**</td>
<td>.532**</td>
</tr>
<tr>
<td>Average</td>
<td>.490**</td>
<td>.708**</td>
<td>.782**</td>
<td>1.000</td>
<td>.743**</td>
<td>.896**</td>
<td>.698**</td>
</tr>
<tr>
<td>Last</td>
<td>.165</td>
<td>.436**</td>
<td>.654**</td>
<td>.743**</td>
<td>1.000</td>
<td>.668**</td>
<td>.649**</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>.510**</td>
<td>.638**</td>
<td>.699**</td>
<td>.896**</td>
<td>.668**</td>
<td>1.000</td>
<td>.647**</td>
</tr>
<tr>
<td>Overall Rating</td>
<td>.405**</td>
<td>.542**</td>
<td>.532**</td>
<td>.698**</td>
<td>.649**</td>
<td>.647**</td>
<td>1.000</td>
</tr>
</tbody>
</table>

N = 77, *p<0.1; **p<0.05
It should be noted that variables are correlating with each other to a great extent. In order to see to what extent we are measuring one underlying variable, factor analysis was applied and it was seen that there is only one factor which explains the 66% of the total variance. All the variables appeared to load highly on this single factor (see Table 10 below)

Table 10: Factor Analysis Component Matrix for First, High Peak, Low Peak, Average, Last and Weighted Average of the impressions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>.601</td>
</tr>
<tr>
<td>High Peak</td>
<td>.742</td>
</tr>
<tr>
<td>Low Peak</td>
<td>.001</td>
</tr>
<tr>
<td>Average</td>
<td>.781</td>
</tr>
<tr>
<td>Last</td>
<td>.792</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>.932</td>
</tr>
</tbody>
</table>

Linear Regression Analysis was first held with all the predictor variables. It appeared that Last Impression was the only significant predictor in this scenario. Considering the fact that the correlations among the variables are significant, High Peak, Low Peak, Weighted Average and First Impressions are left from the Regression Analysis. After the second run, the results were as follows;

Table 11: Regression analysis predicting Overall Impression by Average and Last impressions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.482</td>
<td>.376</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>.003**</td>
<td>.001</td>
</tr>
<tr>
<td>Last</td>
<td>.002**</td>
<td>.001</td>
</tr>
</tbody>
</table>

*Model Summary*

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.540</td>
<td></td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05

Now, it is seen that the Average of the impressions also appeared to be a significant predictor of the overall impression.

Another set of possible predictor variables comes with the analysis of the slopes among assessment points. In order to find the possible effects of the slope of the change from one slider point to the next one, the following variables were;
• Average of the absolute values of slopes
• Average of the positive slopes
• Average of the negative slopes
• Highest positive slope
• Highest negative slope

Next step was calculating the correlations between the variables to see to what extent they are measuring the same thing (See Table 12 below)

**Table 12: Correlations among slope variables.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Average</td>
<td>1.000</td>
<td>.690*</td>
<td>-.886**</td>
<td>.544**</td>
<td>-.764*</td>
</tr>
<tr>
<td>Av. Pos. Slopes</td>
<td>.690**</td>
<td>1.000</td>
<td>-.339**</td>
<td>818**</td>
<td>-.142*</td>
</tr>
<tr>
<td>Av. Neg. Slopes</td>
<td>-.886**</td>
<td>-.339**</td>
<td>1.000</td>
<td>-.166</td>
<td>.920**</td>
</tr>
<tr>
<td>Peak Pos. Slope</td>
<td>.544**</td>
<td>.818**</td>
<td>-.166</td>
<td>1.000</td>
<td>-.063</td>
</tr>
<tr>
<td>Peak Neg. Slope</td>
<td>-.764**</td>
<td>-.142**</td>
<td>.920**</td>
<td>.063</td>
<td>1.00</td>
</tr>
</tbody>
</table>

N = 77, *p<0.1; **p<0.05

As the correlation table indicates, these variables are highly correlating with each other. For further investigation, factor analysis was applied for this set of variables. There were two components appeared where first one was explaining 64% of the variance and the second one was explaining 30% of the variance. Considering the factor analysis loadings and the correlations, Peak of Negative Slopes, Average of Positive Slopes and Average of Negative Slopes measures were left out from the regression analysis. Overall Impression was the dependent variable of this analysis and the results were as follows;

**Table 13: Regression analysis predicting Overall Impression by Absolute Average and Peak of Positive Slopes**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Measure</td>
<td>2.995**</td>
<td>.173</td>
</tr>
<tr>
<td>Absolute Average</td>
<td>.118</td>
<td>.158</td>
</tr>
<tr>
<td>Peak of Positive Slopes</td>
<td>.070</td>
<td>.055</td>
</tr>
</tbody>
</table>

**Model Summary**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>71</td>
</tr>
<tr>
<td>R²</td>
<td>0.063</td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05
It was found that none of the predictors were significant in predicting the overall impression.

Another set of possible predictors was the quantifiable usage metrics. Firstly correlation between these metrics was calculated and results suggest that these measures are correlating significantly with each other (see Table 14 below).

Table 14: Correlations among Unique Page View, Number of Clicks and Time Spent on Website

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Page Views vs. Number of Clicks</td>
<td>.423 **</td>
</tr>
<tr>
<td>Unique Page Views vs. Time Spent</td>
<td>.618 **</td>
</tr>
<tr>
<td>Time Spent vs. Number of Clicks</td>
<td>.665 **</td>
</tr>
</tbody>
</table>

N = 29, *p<0.1; **p<0.05.

In order to see whether or not these metrics are part of one single underlying variable, factor analysis was performed. There is only one component appeared which is explaining the 72% of the variance. All of the measures were loaded highly on this factor. Even if the variables included in regression analysis correlate highly, it appeared that Unique Page Views is still a significant predictor of the Overall Impression (see Table 15).

Table 15: Regression analysis predicting Overall Impression by Unique Page View, Number of Clicks and Time Spent on Website

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.949</td>
<td>.303</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num. Of Clicks</td>
<td>-.002</td>
<td>.010</td>
</tr>
<tr>
<td>Time Spent</td>
<td>-.001</td>
<td>.002</td>
</tr>
<tr>
<td>Unique Page Views</td>
<td>.071**</td>
<td>.026</td>
</tr>
</tbody>
</table>

Model Summary

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>79</td>
</tr>
<tr>
<td>R²</td>
<td>0.101</td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05

Even if Time Spent and Number of Clicks were left out one by one from the regression analysis, the result did not change. Unique Page Views was still the only significant predictor among this set of variables.
**Research Question 4 (User Characteristics):** Can we find behavioural or characteristic patterns of users by making use of the Continuous Response Measurement in website context?

First sub-question in this context was about the categorization of the users based on their tendency to form their overall impression different continuous feedback measures. In order to start with, corresponding variables (First, Last, High Peak, Average and Weighted Average) were calculated for each participant regarding Task 1 (First website that they had seen). In the next step, these measures were compared to overall impression and each participant was categorized based on the closest value to the overall impression. For example, if the peak impression is the closest value to the overall impression of that participant, that participant was put in the category of peak impression. If there are two measures which are both closest to the overall impression, they are labelled together to create a new category (e.g. High Peak-First). After preliminary analysis, the categories appeared to be; First, Peak, Average, Weighted Average, Last, Both High Peak-First and High Peak-Last.

Using these seven categories, it appeared that only 2 out of 26 participants could be categorized successfully for Task 2 based on the their category on Task 1. Success rate was higher for Task 3; 6 out of 26 participants could be categorized for Task 3 based on their category on Task 1. During both of the attempts, it was seen that such a categorization was not successful. Taking the low success rates into consideration, a different strategy had been tried. It was hypothesized that failing in predicting the category may occur because of the closeness of the measures to each other. For instance, if average value is the closest value to overall impression but also first or the last impression are so close that there is not such a clear-cut category, this can be a reason for the low success rates. In order to test the phenomenon, only the cases where there is no other measure as close as 50 units to the closest measure were taken into account. Applying this logic left us with only two cases. Moreover, even these two cases failed in predicting Task 2 and Task 3 category.

Considering the unsuccessful results, the threshold distance was decreased to 20 units. This time, cases where there is no other measure as close as 20 units to the closest measure were taken into account. In this case, it appeared that there are 8 cases left and only one of them was correct in predicting Task 2 performance.

*Therefore, it can be claimed that such a categorization is not possible with the techniques that we have used so far.*

The next sub-question was whether or not people provide more or less feedback with respect to their like or dislike of websites. In order to answer this question, *Number of Slider Movements* was calculated for each participant regarding each website separately.
In this context, more slider movements mean that more feedback is provided by the participant. Using the Number of Slider Movements as a predictor variable where Overall Impression is the dependent variable, the following Regression Analysis was performed:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.452</td>
<td>.218</td>
</tr>
<tr>
<td>Measure Num. Of. Slider Mov.</td>
<td>-.017</td>
<td>.019</td>
</tr>
</tbody>
</table>

**Model Summary**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>78</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.010</td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05

As the results suggest, *Number of Slider Movements was not a significant predictor of the Overall Impression.*

The last sub-question for of this research question was “Do users follow a compensational behaviour by providing negative feedback after a positive one or a positive feedback after a negative one?” In order to see to what extent this question has an empirical basis, cross tabulation table was formed. This table makes it easier to see the relation between an increase/decrease and the next move. As the Table 17 below indicates, there is not such a loading in any cell of the matrix

**Note:** The cell Decrease x Increase means that a decrease occurred after a decrease and Increase x Decrease means that a decrease occurred after an increase.
Table 17: Cross tabulation of the current move and the previous move

<table>
<thead>
<tr>
<th>Previous Move</th>
<th>Decrease</th>
<th>Increase</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decrease</td>
<td>111</td>
<td>107</td>
<td>218</td>
</tr>
<tr>
<td>Current Move</td>
<td>218</td>
<td>216</td>
<td>434</td>
</tr>
<tr>
<td>Increase</td>
<td>107</td>
<td>109</td>
<td>216</td>
</tr>
</tbody>
</table>

Analysis continued with Logistic Regression where Previous Move is the predictor variable and the Current Move is the dependent variable. The results were as follows:

Table 18: Regression analysis predicting Current Move by Previous Slider Movement

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-.037</td>
<td>.192</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous move.</td>
<td>.055</td>
<td>.192</td>
</tr>
</tbody>
</table>

Model Summary

- N: 435
- Cox & Snell $R^2$: 0.000
- Nagelkerke $R^2$: 0.000

*<i>p<0.1; **p<0.05; Dependent Variable: Current Move</i>*

As the analysis suggests, no evidence could be found indicating that next slider move can be predicted with the current one.

**Research Question 5 (Methodological): Is the Continuous Response Measurement tool (namely Slider) obtrusive or distracting for the users?**

In order to see if the borders of the range (between 0 and 1000) were reached by the participants, all of the slider files were examined. Considering the 3 website that each participant visited, the results were as follows:

- Number of participants who reached the upper limit of the slider was 2.
- Number of participants who reached the lower limit of the slider was 8.
In addition to this investigation, answers of the participants during the post-task interviews were also taken into account and it appeared that only one participant bothered with this problem. The participant reported that he wanted to move the slider down to provide a negative impression but the slider was already at the lower limit. It can be claimed that even if the number of participants who reached the borders are high (especially the lower limit), participants who complain about it are relatively less.

The next possible distracting factor was that users might have been looking at the slider while they were interacting with it. In order to test this phenomenon, eye-fixation data was analysed to see if people are looking at the direction where slider is located. The data that eye-tracker provides includes coordinates of each eye-fixation on an X and Y coordinate system. In this plane, a negative X means that participant is looking at the left of the screen where the slider stands. In addition to that slider files were analysed to see how many times people changed the slider position. And finally it was investigated if participants were looking at the slider while they were interacting with it. Here are the results;

- Average number of slider movements while visiting a website was 9.5.
- Average number of times that participants look to the direction of the slider for one website visit was 1.5.
- Average number of times that looking at the direction of the slider and moving the slider at the same time was 0.7.

These calculations suggest that only in 0.7 out of 9.5 (7.36 %) cases; participants were looking at the slider direction while they were interacting with it. Therefore, it can be claimed that most of the times participants were not looking at the slider whenever they were changing their impressions. Thus, it would be reasonable to think that using the slider was not obtrusive for the participants.

**Research Question 6: Can we find any clue from eye-fixation behaviour indicating that participants are about to give a negative or a positive feedback about the website?**

Our first sub-question about this topic was as follows: “Does the average fixation duration differ from the average of 5, 10 or 15 seconds before a positive or negative feedback is provided by the user?” In order to answer this question, the following variables were calculated using the eye-tracking data;

- Average fixation duration.
- Average of the fixation durations 5 seconds before the slider movement.
- Average of the fixation durations 10 seconds before the slider movement.
• Average of the fixation durations 15 seconds before the slider movement.

Average fixation duration was calculated across each participant for each website (three measurements for each participant). Other variables were measured by calculating average fixation durations before slider movements (5, 10 or 15 seconds) and then averaging them once again across each other at the corresponding website.

Since we are comparing two sets of data which are coming from the same participant group, paired samples T-Test was applied individually for all three cases (Average Fixation Duration versus one of the other variable). Average of the fixation durations 5 sec before slider movements were significantly lower than the Average Fixation Duration (t (65) = 4.128 p<.05). There was no significance for 10 and 15 seconds. It can be concluded that just before participants provide a feedback about the website, their average eye-fixation duration decreases.

It should be noted that there were three samplings for each participant. Therefore, testing the interdependency of the cases was required to see if cases could be treated individually. Eventually, data failed to pass this test. We cannot make strong conclusions from these results but we can at least say that it is promising for further research. The same phenomenon holds for the next sub-question as well.

“Does the average fixation duration differ from the average of 5, 10 or 15 seconds before negative feedback provided is by the user? Does the average fixation duration differ from the average of 5, 10 or 15 seconds before positive feedback is provided by the user?”

As it is the case for the previous question, the following variables were calculated this time;

• Average of the fixation durations 5 seconds before a positive slider movement.
• Average of the fixation durations 10 seconds before a positive slider movement.
• Average of the fixation durations 15 seconds before a positive slider movement.
• Average of the fixation durations 5 seconds before a negative slider movement.
• Average of the fixation durations 10 seconds before a negative slider movement.
• Average of the fixation durations 15 seconds before a negative slider movement.

Once again, Paired Samples T-Test was applied where Average Fixation Duration is one of the pairs and one of the variables listed above is the other one. Average of the fixation durations 5 seconds before a negative slider movement was significantly lower than the Average Fixation Duration t (65) = 3.391 p<.05). For 10 sec, no significance was found. The results also indicated that average of the fixation durations 15 seconds before
negative slider movements was significantly lower than the Average Fixation Duration \( t(65) = 2.311 \ p<.05 \). Concerning the positive slider movement variables, none of them was significant. *It can be claimed that before participants give a negative feedback about a website their fixation durations decrease.*

**Research Question 7: Can we predict overall impression from quantitative eye-fixation metrics?**

Before starting the actual analysis, correlation among these variables was calculated before applying the regression analysis (see Table 19 below). It appeared that these quantitative metrics are correlating highly accept for the correlation between average fixation duration and the average distance travelled.

**Table 19: Correlations among Average Fixation Duration, Distance Travelled by Eyes and Average Distance Travelled by Eyes**

<table>
<thead>
<tr>
<th>Correlation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. Fix. Dur. vs. Dis. Traveled</td>
<td>.335 **</td>
</tr>
<tr>
<td>Av. Dis. Travelled vs. Dis. Traveled</td>
<td>.672 **</td>
</tr>
</tbody>
</table>

Factor analysis applied on the set of these variables also suggests that there is only one factor which explains 56% of the total variance. Because of the high correlation between *Distance Travelled* and *Average Distance Travelled* and considering the loadings of the measures from factor analysis, regression analysis was held with only *Average Fixation Duration* and *Average Distance Travelled* as predictor variables;

**Table 20: Regression analysis predicting Overall Impression by Average Fixation Duration and Average Distance Travelled**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.572</td>
<td>.487</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Fix. Duration</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td>Av. Dis. Travelled</td>
<td>-.001</td>
<td>.001</td>
</tr>
</tbody>
</table>

**Model Summary**

<table>
<thead>
<tr>
<th>N</th>
<th>69</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^2 )</td>
<td>0.15</td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05*
As it is seen, both of the variables failed to be a significant predictor of the Overall Impression.

Research Question 8: Is it possible that different predictors affect different dimensions of User Experience in different magnitudes?

Before starting the actual analysis for this question, it is helpful to remember that correlations between continuous response measurement variables (first impression, positive peak impression, negative peak impression, average, weighted average, last impression) were already calculated above (see Table 9). It has been seen that they are highly correlating with each other. Hence, regression analysis was held with First, Average and Last Impressions as predictor variables and Perceived Usefulness as the dependent variable. The results suggest that only Average Impression is significant in predicting Perceived Usefulness. In the next step, First Impression was also left out from the list of predictor variables and the results were as follows;

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.464</td>
<td>.333</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>.004*</td>
<td>.001</td>
</tr>
<tr>
<td>Last</td>
<td>.001</td>
<td>.001</td>
</tr>
</tbody>
</table>

Model Summary

| N   | 69     |
| R²  | 0.492  |

As it is seen, Average Impression is still the only significant predictor for Perceived Usefulness. Same procedure was applied to other dimensions (Perceived Visual Attractiveness, Perceived Ease of Use and Perceived Hedonic Quality). It was seen that only Last Impression was a significant predictor for Perceived Visual Attractiveness ($R^2 = 291$, Coeff. = -.002, SE = .001 $p < .05$). For other factors, there was no significant predictor at all.
3.3 All in one Regression Table

There are several variables used so far to predict the overall impression. Some of the variables are coming from the eye-tracking data; some are coming from questionnaire and some from slider. Until now, they were analyzed among each other together with the relevant predictor variables. At this point we will put them all together in one regression table to see if they neutralize the effect of each other or not. The criterion of inclusion of the predictor variables in this analysis is their correlation with the other similar variables and the results of the performed factor analyses. We do not include all the variables that are highly correlating but we are taking the ones to represent the group. Here are the selected variables; familiarity with the website, average of the impressions, last impression, unique page views, average fixation duration, average eye distance travelled number of slider movements and the order of presentation (namely, Ditzo, AllSecur and Unive as dummy variables). The results appeared as follows;

Table 22: Regression analysis predicting Overall Impression by all variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.012**</td>
<td>.486</td>
</tr>
<tr>
<td>Measure (Reference AllSecur)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>.033</td>
<td>.092</td>
</tr>
<tr>
<td>Average of impressions</td>
<td>.003**</td>
<td>.001</td>
</tr>
<tr>
<td>Last impression</td>
<td>.001*</td>
<td>.001</td>
</tr>
<tr>
<td>Unique page views</td>
<td>.012</td>
<td>.024</td>
</tr>
<tr>
<td>Average fixation duration</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td>Average eye distance travelled</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td>Num. of slider movements</td>
<td>-.015</td>
<td>.017</td>
</tr>
<tr>
<td>Average of absolute slopes</td>
<td>-.047</td>
<td>.289</td>
</tr>
<tr>
<td>Average of positive slopes</td>
<td>.140</td>
<td>.174</td>
</tr>
<tr>
<td>Average of negative slopes</td>
<td>-.009</td>
<td>.101</td>
</tr>
<tr>
<td>Ditzo</td>
<td>.283</td>
<td>.254</td>
</tr>
<tr>
<td>Unive</td>
<td>.130</td>
<td>.307</td>
</tr>
</tbody>
</table>

Model Summary

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.532</td>
<td></td>
</tr>
</tbody>
</table>

*p<0.1; **p<0.05;

As the results suggest, the only significant variables appear to be the average of the impressions and the last impression.
3.4 Post-task Interview

Even though most of the questions that have been answered in this study are based on quantitative data, we did not ignore the opportunity to collect qualitative data as well. It was believed that to understand the experience that participants are going through while they were participating in our experiments might provide us insights for future research. Therefore, understanding what people were thinking when they use the slider is one of our main interests. Our approach here was simply asking them what they have experienced. By doing small post-task interviews by means of a few questions like “In what occasions do people decide to change their impression about using the slider?” and “What makes them use the slider tool?” According to the post-task interviews, most commonly mentioned points by the users were as follows;

- Usability issues matter a lot for the users. When participants are frustrated about completing a task or when a function works better than expected, participants used to move the slider in a positive or negative direction respectively.
- Participants tend to use the slider when they see that the price of an insurance package is cheaper or more expensive compared to other websites with the same coverage. Moreover, coverage of the insurance package they get was also pronounced several times.
- Simplicity or being overloaded is a factor to move the slider.
- Being satisfied with the information found is a factor. When they are satisfied, they tend to move the slider up, when they are not, they do it the other way around.
- The language that the website uses mattered. Informal language, trying to show off by spoiling the other vendors leads to a negative slider move.
- Professional look is expected (e.g. not much flash banners)

4 Discussion and Conclusion

During this study, we attempted to explore the relatively untouched part of the literature where continuous response measurement has been used together with other techniques in order to understand user experience in websites. Since introduction of CRM systems have not been applied in the website context yet, there were a lot of questions raised throughout this research. Therefore, we were not trying to answer just one research question but trying to explore several issues appeared during the flow of this study. Research opportunities were multiplied as the data coming from
questionnaire and the eye-tracking data were also combined with continuous response measurement data. Hence, this study was relatively broad in its research focus and was aiming to create a background for the upcoming research. In the following paragraphs, we will discuss the findings which are relatively more valuable among the whole set of results. While applying CRM to the website domain for the first time, we found that average impression and last impression are significant predictors of users’ global assessment about websites. It was also found that average eye-fixation durations of the users are decreasing before they give a negative assessment about the website. We have seen that using the slider was not distracting users in the sense that they look at it while interacting. On the other hand it was found that considerable amount of users are reaching the borders of the range of the slider preventing them to provide feedback as they want. This study also showed that variables used for such a study can be reduced to a certain number instead of using various metrics measuring the same underlying concept. Moreover, it appeared that familiarity can create a difference among the websites but order of the presentation has no such effect.

Figuring out the components of the user experience in internet domain was one of our primary goals for this research. Based on the Technology Acceptance Model, several dimensions of the interaction between human and computer interfaces have been found (Davis, et. al., 1989). The existence of these components was further investigated also in internet domain (Heijden, H., 2003). Different studies performed different designs where one website or multiple websites have been used. Also the domain of the websites differed throughout these studies. For the purposes of our study, we believed that figuring out the dimensions that exist for our context of three Dutch car insurance websites is important for the rest of the analysis. Applying the combination of the dimensions mentioned in the literature (Perceived Usefulness, Perceived Visual Attractiveness, Perceived Hedonic Quality, Perceived Ease of Use, Intention to Use), we have seen that all the dimensions exist for our context except for the Intention to Use. The reason that Intention to Use did not appear as a factor in our analysis may be due to the number of participants used for this study. Another possible explanation can be that the websites did not have enough variance in terms of this dimension. After all, if the latter is the correct explanation, this result may suggest that it is a more convenient approach for further research to look for the “own” dimensions for each website rather than blindly using the dimensions mentioned in the literature.

On the other hand, it was seen that the factor loadings did not occur exactly as they were supposed to be. Most of the questions were loaded on the factors as expected but some of the questions switched the factor or appeared in two factors. This may be due to the fact that some questions belong more obviously to some factor than the others. For instance, the question “Using the website enhances my effectiveness in completing the task” loaded heavily on perceived usefulness factor which was expected. On the other
hand, the question “I find it easy to get the website to do what I want it to do” also appeared in both Perceived Usefulness and Perceived Ease of Use factors. This outcome might have occurred due to the meaning of the question which may infer both usefulness and ease of use. Another possible reason which caused the switch of questions among the factors may be due to the number of participants used. Including more participants may make the separation among the factors better specified.

One of the most important findings of this study was showing that the differences between websites regarding the overall judgements can be traced down to its components. It was seen that Unive was rated higher compared to AllSecur website in terms of overall ratings. When we compare these websites in terms of the dimensions of the user experience, it appeared that these websites significantly differ in two dimensions. Unive was rated higher in Perceived Usefulness dimension but on the other hand AllSecur outperformed Unive in Perceived Hedonic Quality. Since Perceived Usefulness was the strongest component in factor analysis, this can explain why Unive exceeded the overall ratings of AllSecur. Considering these results, it can be claimed that website owners are able to make more efficient improvements by investigating the performance of the website in different components of the user experience rather than taking the global assessment as the only criterion. If there is a lack of acceptance of website, the most problematic dimensions can be detected rather than changing everything about the website. Considering different dimensions separately may lead to better prioritization of the planned improvements about the website.

Before starting the experiment, we were hypothesizing that there may be an order effect of seeing websites on overall judgement the users. This would lead to an advantage for the websites which appear on top of a result set of a search engine. If the first website seen has an advantage regardless of the content of the website, there would be implications. Even if we counterbalanced and randomized the presentation of the websites, we could still check whether or not first website has an advantage in our context. It appeared that such an advantage does not exist. It may imply that it would not make any difference either your website appear on number one or number three in the result set as soon as users check all of them. But since it is not guaranteed that the third website will be visited, it is still a good idea to be on the top. Considering that we participants of our experiment could not choose the number of websites to visit, these results do not fully represent the real life conditions.

Another methodological question we have asked was the effect of familiarity of the website. The results suggest that Unive and AllSecur differ significantly in terms of being known by the participants of our experiment. It was an interesting result of the preliminary tests to see that Unive outperformed AllSecur in both overall impression and familiarity. Therefore, we included familiarity as a predictor variable and have seen
the difference in the ratings of Unive and AllSecur disappeared as soon as we controlled familiarity. Although it is a well-known fact that being known is favourable as a website, we confirmed that familiarity can create a difference among websites.

In line with the literature published by Baumgartner et al. (1997) and Hamberg & Ridder (1999), we have seen that some patterns of continuous affective feedbacks can lead to effect the overall judgement which can also be used to predict it. Baumgartner et al. (1997) was suggesting that “consumers’ global assessments of extended affective episodes elicited by advertisements are dominated by the peak emotional experience and the final moment of the series and also are correlated with the pace at which momentary affective reactions improve over time”. In addition to that, Hamberg et al. (1999) found the similar results in image quality domain. They suggest that “subjects rely predominantly on the worst events of a sequence when determining their overall quality judgment”. In our case, we found that last impression and the average impression of the users are significantly successful in predicting the overall judgements. This result makes sense because of the already proven phenomenon called “recency effect”. When participants are asked to recall the items that they have been presented, they tend to begin recalling with the end of the list rather than the middle of the list.

With respect to our setting, last impression of the participants was the most successful predictor rather than the average or the peak values. On the other hand, there was no evidence for the existence of the primacy effect since first impression was not successful in predicting the overall judgement. Our study confirmed that recency effect also holds for our context of continuous response measurement on websites. Also the results of factor analysis regarding the CRM measures might have some implications for the future research. We were dealing with different measures since we did not know which ones are relevant with our experimental setting. It appeared that all the metrics we have used (First, High Peak, Low Peak, Average, Weighted Average and Last impression) belong to one underlying factor. Future studies applying CRM on websites may consider not including all of them in their studies but using one or two variables among this list. Also calculating the slopes could not help us predicting the overall impression for the purposes of our study. These variables may also be reduced to two variables since factor analysis and correlations table suggest so. To sum up, we hope that this study was helpful in terms of refining the number of the measures that will be used for CRM systems in internet domain.

Same predictors were also used to predict the success of the user experience dimensions rather than the overall judgement. This time, it appeared that average of the impressions was successful in predicting Perceived Usefulness but on the other hand, last impression was the only significant predictor of Perceived Visual Attractiveness factor. These results suggest that different patterns of the time spent surfing a website may be good at predicting a dimension but not the other. Therefore these dimensions
should be examined differently rather than just using the overall judgement as the only dependent variable of the website user experience research.

From a practical perspective our research is relevant for those who are designing adaptive websites. Even though results are not robust enough to make hard conclusions, our findings about the eye-tracking data suggest that Average Fixation Duration decreases before participants use the slider to give a feedback. The further investigation also indicated that this effect is more severe from negative slider movements. One of the possible explanations of such a finding can be the fact that if eyes of the participants move more around the screen, the fixation durations would expected to be less. Therefore, frustration due to searching something but failing to find may result in giving a negative feedback for the website. In a future adaptive system, this hypothesis can be used in helping the frustrated website users. Assuming that the eye-trackers will be used more frequently by the users as it gets more affordable or webcams will perform better as eye-trackers, we can hypothesise that new generation adaptive websites can make use of it. In a possible scenario, website may pop up a help window as soon as it notices that eye-fixation durations are decreasing below the average of that particular user. These results were one of the most promising outcomes of this research. Further research may combine eye-fixation data with other quantitative measures like click behaviours to make better conclusions about the experiences that users are going through. This enhancement can help us designing a better user experience in real time.

While finding significant predictors of global assessments of users, we failed to find significant effects of some quantitative measures as well. It seems that some of the quantitative eye-fixation metrics (Average Fixation Duration, Average Distance Travelled and Total Distance Travelled), web usage metrics (Number of Clicks and Time Spent on Website) and Continuous Response Measurement metrics (Number of Slider Movements) failed to be a significant predictor of the Overall Impression. These results might have raised due to the fact that most of the metrics above can be interpreted in opposite ways. For instance, it can be claimed that longer time spent on the website may mean that users are enjoying it but on the other hand it can also be claimed that users are frustrated and it takes long to complete their tasks. The same may also hold for Number of Clicks. You might be clicking around just because you find the website interesting or because you could not find what you want. On the other hand, Number of Unique Page Views appears to be positively related to the overall judgement. People who like the website seems to visit more pages (or the other way around). In any case, it is a good idea for website owners to keep more traffic on their website and show more pages to their visitors.

One of the major findings of this exploratory study was seeing the extent how problematic it was using a slider in such a setting. The equipment used in the
experiments are supposed to be unobtrusive and do not distract participants while they are performing their tasks. Slider in that sense was a question mark for the environmental validity of this study and the followings. Looking at the slider each time interacting with it would be a possible distractor for the users. We provided an explicit test of the phenomenon and see that only in 7.36% of the cases this problem was observed. Therefore, it is reasonable to think that slider was not preventing the participants from performing their task efficiently which are also confirmed by the post-task interviews. One of the possible explanations of this finding can be the fact that most of the users were holding the device such a way that they can calibrate it using their fingers. First of all, it should be noted that all the participants were asked to keep their hand on the device during the experimental sessions to avoid them from forgetting the existence of it. It was noticed that the way they grasp the device was allowing them to calibrate themselves to see at what position the slider is regarding the scale (between 0 and 1000). This behaviour of the participants might have eliminated the need to look at the slider to be able to calibrate it. These results suggest that there is not much to worry about the distracting effect of slider in such experimental settings.

On the other hand, the possible problem due to the physical borders of the slider was more severe. It appeared that especially the negative lower limit of the slider had been reached by considerable number of participants. Even though only one of them complained about this problem, using other continuous response measurement tools such as buttons may increase the validity of the experiments. Another limitation that we faced was being able to use only one variable which is the overall impression in our case. By using the continuous response measurement tool, subjects were providing feedback about their overall impression about the website. Selection of the variable here clearly affects the results and their interpretation. We could have used frustration or confusion for instance as the selected variable. Being limited with only one dimension of measurement limited us from discovering more about the experience that users were going through. Future studies may consider overcoming this limitation by introducing new approaches.

Several research questions were answered regarding the integration of continuous response measurement into overall judgements about the websites. Furthermore this integration is enriched by using the eye-tracking data of the participants whilst surfing on internet. Our approach was investigating the components of the user experience as well as the overall impression of the users. The added value of the study was quantifying almost every research question and making use of the possibilities that arise by combining different techniques. We believe that the results of this study might open new research opportunities for the further studies in this domain. Throughout our investigation with CRM on websites, we found that average impression and last impression are significant predictors of users’ global assessment about websites. We
found considerable clues suggesting that average eye-fixation durations of the users are decreasing before they give a negative assessment about the website. Our methodological investigation showed that using the slider was not distracting users in the sense that they look at it while interacting. On the other hand it was found that considerable amount of users are suffering from the limits of the slider range. We hope that results of this study will provide a background for further research applying CRM on websites and combining it with different evaluation methods.
5 References


BS EN ISO 9241:1998. Ergonomic requirements for office work with visual display terminals (VDTs). Keyboard requirements.


6 Appendices

6.1 Appendix 1 – Pre-Task Questionnaire

Frequency of Internet Usage

On the average, how frequently do you use the internet?

1. Never/almost never
2. Less than once a month
3. A few times a month
4. A few times a week
5. About once a day
6. Several times a day

Daily Internet Usage

On the average working day, how much time do you spend on internet?

1. Never/almost never
2. Less than ½ hour
3. From ½ hour to 1 hour
4. 1-2 hours
5. 2-3 hours
6. More than 3 hours

Diversity of Internet Usage

Please indicate the extent to which you use the internet to perform the following tasks:

<table>
<thead>
<tr>
<th>Task</th>
<th>Not at all</th>
<th>To a great extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Get information</td>
<td>1 2</td>
<td>3 4 5</td>
</tr>
<tr>
<td>Get product support</td>
<td>1 2</td>
<td>3 4 5</td>
</tr>
</tbody>
</table>
3. Communicate with people 1 2 3 4 5
4. Get free resources 1 2 3 4 5
5. Purchasing/shopping 1 2 3 4 5
6. Apply for job 1 2 3 4 5

**Ability to Use Internet**

Please indicate the extent to which you agree with the following statements about World Wide Web usage.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Not at all</th>
<th>To a great extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I have mastered the use of the WWW</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>2. I do not perform WWW-related tasks as well as I would like</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>3. I am certain I can perform WWW-related tasks well</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>4. It is just not possible for me to solve WWW-related Tasks</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>5. I think my performance in solving WWW-related tasks could be improved substantially</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

**6.2 Appendix 2 – Post-Task Questionnaire**

**Perceived Usefulness**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Using the website makes it possible to complete the task.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>
7. I find the website to be useful completing the task.  

8. Using the website is helpful completing the task.  

9. Using the website enhances my effectiveness in completing the task.  

**Perceived Ease of Use**

10. I find it easy to get the website to do what I want it to do.  

11. I find the website to be easy to use.  

12. Interacting with the website requires a lot of mental effort.  

13. My interaction with the website is clear and understandable.  

**Perceived Hedonic Quality**

14. The website is standard.  

15. The website is innovative.  

16. The website is exciting.  

17. The website is original.  

18. The website is boring.  

19. The website is impressive.  

**Perceived Visual Attractiveness**

20. The colours that are used on the website are attractive.  

21. Overall, I find that the site looks attractive.
22. The design of the website is unattractive.  

23. The layout of the site is attractive.  

**Intention to Use**  

24. If possible I intend to use the website again.  

25. If possible I predict that I would use the website again.  

26. I would recommend to use the website to friends.  

27. In a similar situation I would use the website again.  

**6.3 Appendix 3 - Instructions**  

**Part 1**  

Welcome! Thank you for participating in this study.  

This study consists of reviewing 3 websites and filling questionnaires after each task. We will provide you a scenario and then ask you to respond to some questions where you can give us your opinion related to your experiences.  

The entire process should take about 30 minutes.  

Your name or personal information will NOT be used for any purpose. We will not record your face or voice, we will only have a screen capture for further analysis.  

We will not be testing your performance in this experiment. We want to have insights about how you are interacting with the websites.  

Have fun!
Part 2

Imagine that you just bought a new car and want to insure it online. You have 3 alternatives of companies which all have car insurance options:

- Ditzo
- All Secur
- Unive

Please review the websites to figure out the BEST option for you. You are free to click around once the website is opened automatically.

Please click DONE button standing on the bottom-right corner of the screen when you are finished with reviewing the website. But be aware that when you click DONE after reviewing, you cannot turn back to a website you already reviewed.

After completing the first questionnaire, you one of the websites will open up automatically and you can start with reviewing. After each website, you will be asked to complete a questionnaire.

IMPORTANT: While you are reviewing the websites, please use the slider to enter your impressions about the website. You can do that whenever you think that you want to update your impression about the website (in a positive or a negative way). Please click on the button at the top left corner to submit your impression.

Please don’t use the address bar to move to other websites.

6.4 Appendix 4 – Example Java Code

```java
import java.io.BufferedReader;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.io.PrintWriter;
import java.util.StringTokenizer;

public class timeSyncronizer {
    private static void doReadWriteTextFile() {
        try {
            // input/output file names
        }
    }
}
```
String inputFileName = "subject07.txt";
String outputFileName = "subject07_time_processed.txt";
int gap = 9;

// Create FileReader Object
FileReader inputFileReader = new FileReader(inputFileName);
FileWriter outputFileReader = new FileWriter(outputFileName);

BufferedReader inputStream = new BufferedReader(inputFileReader);
PrintWriter outputStream = new PrintWriter(outputFileReader);

String firstLine = null;
String inLine1 = null;

firstLine = inputStream.readLine();

StringTokenizer st1 = new StringTokenizer(firstLine, " ");
String column1 = st1.nextToken();
String column2 = st1.nextToken();
String column3 = st1.nextToken();

StringTokenizer st2 = new StringTokenizer(column2, ",");
String part1 = st2.nextToken();
String part2 = st2.nextToken();

int base = Integer.parseInt(part1);
int secsIn = gap;
int hours = secsIn / 3600;
int remainder = secsIn % 3600;
int minutes = remainder / 60;
int seconds = remainder % 60;

String time = (hours < 10 ? "0" : "") + hours
+ ":" + (minutes < 10 ? "0" : "") + minutes
+ ":" + (seconds < 10 ? "0" : "") + seconds;

outputStream.println("1 "+gap+" " + coulmn3+""+time);

int i = 2;

while ((inLine1 = inputStream.readLine()) != null) {
    st1 = new StringTokenizer(inLine1, " ");
    coulmn1 = st1.nextToken();
    coulmn2 = st1.nextToken();
    coulmn3 = st1.nextToken();
    st2 = new StringTokenizer(coulmn2, ",");
    part1 = st2.nextToken();
    part2 = st2.nextToken();
    secsIn = Integer.parseInt(part1) - base + gap;
    hours = secsIn / 3600;
    remainder = secsIn % 3600;
    minutes = remainder / 60;
    seconds = remainder % 60;
    time = (hours < 10 ? "0" : ") + hours
    + ":" + (minutes < 10 ? "0" : "") + minutes
    + ":" + (seconds < 10 ? "0" : ") + seconds;

    outputStream.println(i+" " + Integer.toString(Integer.parseInt(part1) - base + gap)+"" +
coulmn3+""+time);
    i++;
}
```java
outputStream.println();
outputStream.close();
inputStream.close();
} catch (IOException e) {
    System.out.println("IOException: ");
    e.printStackTrace();
}

/**
 * Sole entry point to the class and application.
 * @param args Array of String arguments.
 */
public static void main(String[] args) {
    doReadWriteTextFile();
}
}

6.5 Appendix 5 – Sample from raw slider file

Column1: Sampling index
Column2: Time stamp
Column3: Slider Position
000000331 40885,1501 518
000000332 40886,1501 518
000000333 40887,1501 518
000000334 40887,1501 518
```
6.6 Appendix 6 – Sample from fixation file

<table>
<thead>
<tr>
<th>FixationIndex</th>
<th>Timestamp</th>
<th>FixationDuration</th>
<th>MappedFixationPointX</th>
<th>MappedFixationPointY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1612</td>
<td>529018</td>
<td>250</td>
<td>320</td>
<td>274</td>
</tr>
<tr>
<td>1613</td>
<td>529268</td>
<td>142</td>
<td>391</td>
<td>269</td>
</tr>
<tr>
<td>1614</td>
<td>529410</td>
<td>216</td>
<td>358</td>
<td>245</td>
</tr>
<tr>
<td>1615</td>
<td>529626</td>
<td>266</td>
<td>611</td>
<td>404</td>
</tr>
<tr>
<td>1616</td>
<td>529892</td>
<td>108</td>
<td>647</td>
<td>437</td>
</tr>
<tr>
<td>1617</td>
<td>530001</td>
<td>175</td>
<td>875</td>
<td>423</td>
</tr>
<tr>
<td>1618</td>
<td>530176</td>
<td>691</td>
<td>842</td>
<td>282</td>
</tr>
<tr>
<td>1619</td>
<td>530867</td>
<td>250</td>
<td>855</td>
<td>352</td>
</tr>
<tr>
<td>1620</td>
<td>531117</td>
<td>166</td>
<td>368</td>
<td>664</td>
</tr>
<tr>
<td>1621</td>
<td>531283</td>
<td>350</td>
<td>270</td>
<td>695</td>
</tr>
<tr>
<td>1622</td>
<td>531633</td>
<td>366</td>
<td>281</td>
<td>742</td>
</tr>
</tbody>
</table>
6.7 Appendix 7 - Visualizations

In this visualisation, the X axis is the position of the slider and the Y axis is the time where one participant is evaluating one website. The diameter of the blue dots shows the eye fixation durations. The larger it gets, the longer the fixation duration. This visualization was used to see if there is any pattern about fixation duration can be detected before users are interacting with the slider.
In this visualization also, X axis is the position of the slider and the Y axis is the time. This one is also showing the data of one participant while interacting with one of the websites. The red circles show the left mouse clicks and the green ones are the keystrokes. The horizontal lines appeared are the period where users are filling in the questionnaire. This visualization was created in order to see the patterns of clicks or keystrokes if there is a change in participant's behaviour happens before interacting with the slider.