2017-3

An investigation of the correlation between Mental Workload and Web User's Interaction

Joaquim Filipe Romero
Technological University Dublin

An investigation of the correlation between Mental Workload and Web User's Interaction

Joaquim Filipe Braga Simões Romero

D14124201

A dissertation submitted in partial fulfilment of the requirements of Dublin Institute of Technology for the degree of
M.Sc. in Computing (Advanced Software Development)

2017

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Advance Software Development), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the test of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed:

Date: 05 March 2017

ABSTRACT

Mental Workload, a Psychology concept, was identified as being linked with task's and system's performance. In the context of Human-Computer Interaction, recent research has identified Mental Workload as an important measure in the designing and evaluation of web interfaces, and as an additional and supplemental insight to typical Usability evaluation methods. Simultaneously, web logs containing data related to web users' interaction (e.g. scrolling; mouse clicks) have been proved useful in evaluating the Usability of web sites by levering the data tracked for hundreds of users. In order to study if the potential of logs of user interaction can be applied in the study of Mental Workload in Web design, an online experiment with 145 participations was performed. Additionally, the experiment, composed of alternative interfaces, sought to assess the role of Mental Workload in the evaluation of interfaces using interactive Infographics, which were identified by literature as bringing new challenges and concerns in the field of Web Design.

The online experiment's results suggested that correlations between mental demands and users' interaction can only be observed when taking in consideration the web interface used or the profile of the users. Moreover, the used measurement methods for assessing Mental Workload were not capable of predicting task performance, as previous research suggested (in the context of other types of web interfaces).

Keywords: Mental Workload; User Interaction Tracking; Web Design Evaluation; Infographics

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to the supervisor of this project, Dr. Luca Long. Firstly, for helping me choosing this field of study, that combines Computing Science and Psychology. I've always had a strong interest in Psychology (even considered the field when applying to graduate institutions), and it was great to being able to study it in the context of Computing, my field of specialization. Secondly, for his incredible guidance, encouragement and availability through all the phases of the project, even though work and communication conditions were not always ideal.

Also, I would also like to thank all the family, friends and anonymous subjects that participated in the online experiment performed in the project. Without them, this study would not be concluded. Their availability was provided only in good will and in the name of science, without any further compensation for the time spent.

Finally, I would like to thank my family. My father and my brother who, no matter the circumstances, never cease to amaze me and make me smile. Susana, who is always by my side, in my success and failures, in my ups and downs. She has made me a better version of myself. Lastly, but not the least, I would like to express my tender gratitude to my mother, who everyday teaches me that, no matter what, we should always fight for ourselves and for those we love, and that we must never give up on our dreams and happiness.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1 INTRODUCTION

1.1 Background

In the last decades, the concept of Mental Workload (MWL) has gained importance in the field of Human-Computer Interaction (HCI). Although there is no universal or unique definition, Mental Workload can be described, in a simplified way, as the amount of mental effort needed by someone (operator) to execute a task over a given period of time (Xie & Salvendy, 2000). Therefore, "mental (or cognitive) workload is a way to measure the mental demands of complex systems" (Zhu & Hou, 2009, p. 387). Previous research linked mental overload and underload with human performance's degradation and, consequently, with negative impacts in a system's efficiency (Cao, Chintamani, Pandya, & Ellis, 2009; Xie & Salvendy, 2000).

In spite of being used to model and predict performance in several fields, Mental Workload has gained increasing importance in Human-Computer Interaction, due to the increasing use of the World Wide Web (Longo, 2011). Considering that, in the field of HCI, many tasks consist in complex activities and demand the use of users' cognitive abilities, it is only natural to assume that assessing Mental Workload in the context of HCI should be considered in the evaluation of HCI interfaces, allowing to optimize the Mental Workload imposed to users when interacting with those interfaces and, consequently, optimizing HCI systems' performance. Recent research addressed this matter and identified Mental Workload as a reliable and important dimension in the evaluation of HCI interfaces, more specifically in the HCI domain of Web Design (Albers, 2011; Gwizdka, 2010; Longo et al., 2012; Tracy & Albers, 2006; Wästlund, Norlander, & Archer, 2008; Zhu & Hou, 2009).

Additionally, the advent of digital technologies led to the increasing practise of using graphic and visual approaches in the representation of digital information and data (Womack, 2014). This type of representations are usually referred as "Infographics" or "Data Visualisations" (Ru & Ming, 2014). Since a significant portion of the digital data generated and consumed every day is characterized by being heterogeneous, complex, and large, Infographics and Data Visualisations domains have become a major focus of attention in the field of HCI, which lead to the advent of new sophisticated approaches and computing techniques capable of handling, be applied to, and be visually representative of large volumes of data (Womack, 2014).

1.2 Research problem

As mentioned, Mental Workload is an important concept in human-computer interfaces, and is related with performance and error's ratios (Xie & Salvendy, 2000). In order to succeed and in order to reach efficiency and productivity, recent research argued that a web application or service must respect the limits of human cognitive processing by seeking to achieve an optimum level of the Mental Workload imposed to users (Tracy & Albers, 2006). Hence, effective methods to monitor and assess Mental Workload are extremely useful, and should be taken in consideration, when designing web and systems' interfaces (Xie & Salvendy, 2000).

Methods for measuring Mental Workload are usually divided in three categories: Subjective methods (i.e. using surveys); Physiological methods (e.g. eye activity); and Performance methods (e.g. error rates) (Cain, 2007). Although previous research has identified the main advantages and disadvantages of using the several methods belonging to these three categories, it seems that the most used and reliable ones have the same problem in common: they are restrictively used in the context of laboratory experiments, which entails being limited in terms of number of subjects; types of tasks; and context. In other words, they do not scale well. This seems to be a huge disadvantage considering that, nowadays, web designers are continuously faced with the need of changing existing interfaces or creating new ones, since the field never ceases to evolve and change, as well as users' demands (Tracy & Albers, 2006; Van Orden, Limbert, Makeig, & Jung, 2001).

In other domains of Human-Computer Interaction, such as Usability, research has been exploiting logs of web users interaction (e.g. mouse pointer position; scrolling) in order to understand and study web users' behaviour (Agichtein, Brill, Dumais, & Ragno, 2006; Atterer, Wnuk, & Schmidt, 2006; Q Guo & Agichtein, 2012). This opens the possibility of also using user interaction logs in the study and assessment of Mental Workload in the context of Web Design, which would allow leveraging data from hundreds of users, and no longer be dependent on hard-to-scale Mental Workload measurement methods.

Additionally, although research that studied Mental Workload in the field of Web Design seems to be in concordance with the importance of the concept in the evaluation of web interfaces, it seems that there are lack of studies assessing Mental Workload in web interfaces that use Infographics for representing digital data, which is an approach that has been increasingly adopted by web sites and users (Azzam, Evergreen, Germuth, & Kistler, 2013).

1.3 Research Objectives

This project aims at contributing to the body of knowledge related to the research of the concept of Mental Workload in the field of Human-Computer Interaction, more specifically, in the domain of Web Design. Two research objectives were defined.

The main research objective consists in studying the possible correlation between the Mental Workload imposed by a web task and the interaction actions performed by the users that executed the web task. The study refers to User Interaction as the activity and actions that allow users to navigate and interact with a web page when using their desktop or laptop. This project considers the following User Interaction elements: keyboard usage; scrolling; mouse position and clicks. User Interaction logs have been analysed in Usability research with positive results, which raised the question of finding if levering those types of logs would also be beneficial in the assessment and/or prediction of Mental Workload in the context of Web Design.

The research question and respective hypotheses defined in order to accomplish this primary objective are the following:

Can objective indicators of users' interaction be used to approximately measure the subjective Mental Workload of a web-based task?

- Null hypothesis (H0): Objective indicators of users' interaction are linearly unrelated with the subjective Mental Workload required by a web task.
- Alternative hypothesis (H1): Objective indicators of users' interaction are linearly correlated with the subjective Mental Workload required by a web task.

The secondary research objective consists in studying if measuring Mental Workload is an advantage in the evaluation and comparison of web interfaces using different visual representations (i.e. Infographics) of the same dataset. Previous research suggested that Mental Workload is an important aspect to take in consideration in the process of designing web interfaces. This project aims to study if similar recommendations can be made in the particular case of interactive Infographics.

The research question defined in order to accomplish this secondary objective is the following:

Can subjective Mental Workload be used to evaluate the performance of interfaces using different Infographics representations?

- Null hypothesis (H2): Subjective Mental Workload cannot be linked to Infographics interfaces' performance.
- Alternative hypothesis (H3): Subjective Mental Workload can be linked to Infographics interfaces' performance.

1.4 Research Methodologies

This project started with a secondary research that allowed to further understand previous studies related with the topics addressed: Mental Workload; Data Visualisation; Web Design Evaluation; and Tracked User Interaction. This secondary research allowed building the necessary knowledge that allowed defining, based in the literature, what elements should be used in the primary research, regarding: Mental Workload measurement methods; types of Infographics; indicators of performance; and indicators of web users' activity.

The primary research consisted in performing an online experiment, coded from scratch, that allowed collecting the required data for performing quantitative and deductive research, which consisted in analysing and studying possible correlations between the data collected: Mental Workload (collected with surveys) and User Interaction (tracked with programming scripts). This correlation was assessed by using statistical correlation coefficients, capable of measuring the extent to which those variables tend to change together and, therefore, testing the hypotheses previously set for the project (empirical research).

1.5 Scope and Limitations

The focus and scope of the project were defined by its goals. The first goal, assessing a possible correlation between Mental Workload and User Interaction (e.g. Mouse Clicks), defined the need of performing a web experiment where the users had to use their laptop or desktop to perform a set of web tasks. The secondary objective, evaluating Infographics by measuring Mental Workload, defined the type of alternative interfaces to use in the online experiment. Moreover, secondary research led to defining fact-finding tasks as the type of tasks to use in the performed online experiment, since this type of tasks were the most common in the reviewed literature. Moreover, in order to reduce intrusiveness, subjective measurement methods of Mental Workload were used, also because they were identified as reliable and easy to implement in literature.

Regarding limitations, the number of reliable participations (145) surpassed the number of participations defined as the minimum sample needed (80). However, it is still a small sample, not statistically significant to support all the considerations discussed when analysing the final dataset. Especially because only 95 participations fully completed the experiment (the remaining were used for calculating performance indicators, such as Rate of Abandonment). Another limitation was the restrictions identified when designing the web tasks. Since the project aimed to reach a wide audience, in order to collect relevant amounts of data, the experiment was performed online rather than in a laboratory. Because of this, the difficulty of the web tasks could not be exaggerated, as well as the amount of time needed to finish the tasks (in order to avoid high ratios of abandonment). Since this design limitation could restrict the collection of high levels of Mental Workload, the web tasks were designed to be more laborious then complex or difficult, with the intention of collection a wide spectrum of Mental Workload demands. The fact that the experiment was performed online also limited the control over the experiment (e.g. standardize the devices used; identify cheating).

1.6 Document Outline

Chapter 2 - Literature Review: This chapter covers the relevant literature that studied the concept of Mental Workload, particularly in the field of Human-Computer Interaction. The chapter starts by covering domains of knowledge that need to be understood in the context of this project and related research: Human-Computer Interaction; Usability; Data Visualisation and Infographics. The following section covers the concept of Mental Workload, starting with its definition and relevant measurement methods, and finalising with its fields of application. The following section of this chapter aims to give the reader a perception of the type of research studies that have been performed in relation with the elements studied in this thesis, namely: Mental Workload Evaluation in Human-Computer Interaction; Evaluation of Data Visualisation Interfaces; and Tracking of User's Interaction. Finally the gaps in literature are identified and the resultant research questions and hypotheses of this project are described.

Chapter 3 – Design and Implementation: This chapter describes the design and the implementation of the online experiment that was developed and launched in order to address the main goals of this study. The chapter starts by describing the main components of the experiment and how they all fit together, followed by the main considerations taken when designing the experiment, which were grounded in the literature. Finally, the measures and indicators collected with the experiment are listed, as well as their calculation formulas and how they are supported by literature.

Chapter 4 – Results and Evaluation: This chapter shows the results obtained with the performed online experiment. The results are discussed in the light of what previous research had found. Moreover, the chapter address the results' relevance and corroboration with the hypotheses and research questions previously set for this study. The chapter starts by showing the results of each type of indicators: Mental Workload; Performance; User Interaction. Thereafter, the results of the correlations related with the study's primary objective - correlate Mental Workload and User Interaction - are discussed. Finally, the results related with the secondary objective - use Mental Workload to evaluate alterative Infographics - are addressed. The final section also discusses if the findings of this study are in line with previous research, which linked high and low levels of Mental Workload with low performance.

Chapter 5 – Conclusion: This final chapter summarizes the results of this study, in the light of the objectives defined. A reflection of what went well and what went bad is given, as well as what could have been done better. The contributions and impact related with this study's results are addressed, as well as suggestions for future work and studies.

2 LITERATURE REVIEW

This chapter covers the relevant literature related with the research of Mental Workload, particularly in the field of Human-Computer Interaction.

Since this project aims to study Mental Workload in the context of several domains belonging to the board field of Computer Science, it is important to overview the most relevant concepts, related with those domains, that are addresses and incorporated in the experiment performed by this project. Therefore, prior to the review of Mental Workload and similar work, the first section of this chapter - Relevant Concepts - aims to overview those domains, in order to put the reader in context and support the resultant design considerations and results' evaluation carried out in following chapters.

The second subchapter - Mental Workload - aims to give the reader a solid knowledge of the concept of Mental Workload. The subchapter starts with the history and definitions of the concept, as well what motivated its application in several fields. Then, the measures and methods used to assess and evaluate Mental Workload will be described, as well as their related issues and factors. Lastly, the reader will be provided with an overview of the fields of Mental Workload application, with focus on the Human-Computer Interaction field.

Finally, the third and last subchapter - Related work on Mental Workload and Human-Computer Interaction - aims to give the reader a perception of the type of research studies that have been performed in relation with the elements studied in this project. Due to the mixed elements that this study addresses, which are not usually combined in the reviewed literature, this chapter will encompass the following "distinct" domains:

1. Research related with the use and measure of Mental Workload in the field of Human-Computer Interaction;
2. Research that aimed to evaluate Data Visualisation interfaces. Since this project aims to present Mental Workload assessment as a new approach for evaluating Data Visualisation interfaces, it is important to review the current research and respective concerns that studied the topic, which is the goal of this section;
3. Research, in the scope of Human-Computer Interaction, which tracked user interaction (e.g. mouse clicks) with the purpose of extracting extra information that benefits the understanding of how HCI designers can improve users' experience and systems' performance. The research covered in this section was carefully selected in order to illustrate the usual approaches of dealing and

interpreting users' interaction. Several domains of HCI are covered, such as Usability, Data Visualisation, and Mental Workload.

Since literature holds several studies related with the aforementioned research, it would be impossible to cover all considerations, developments, and outcomes in the area. Instead, the literature reviewed in the "Related Work" subchapter purposes to overview some related studies, which serve as illustrations of the main approaches and conclusions that have been highlighted by scientific research.

2.1 Relevant Concepts

2.1.1 Human-Computer Interaction

Computer technology is widely available in people's life and is used everywhere: to work; to communicate; to shop; to seek out new information; and to entertain (Pantic, Nijholt, Pentland, & Huanag, 2008). In other words, computer technology is strongly accessible for everyone, anywhere, and anytime. Consequently, users of computer technologies have diversified and grown (Churchill, Bowser, & Preece, 2013). These circumstance lead to the scientific study of human-machine systems, which "refer to the system composed by humans and machines and to fulfil some functions through the interaction between humans and machines. The humans refer to the operators and managers of machines and the machines refer to machineries, equipments, tools and work environments. The whole human-machine systems consist of humans, machines, interfaces between humans and machines and environments in which the systems locate" (Chao, 2009, p. 230).

This research field is referred to as Human-Computer Interaction (HCI), and undertakes an approach that moves away from computer-centred designs towards human-centred designs. The designing process in this domain aims to ensure the overall system functionality and ease of use in the perspective of the user, aiming to optimize effectiveness usage and satisfying interactions, both from organizational and individuals' perspective. This purpose, of meeting both organizational and users' needs, entails acknowledging and taking in consideration all elements that form Human-Computer Interaction systems, such as: goals; users; roles; technology; tasks; and context (Kendall & Kendall, 2011).

2.1.2 Usability

According to Quesenbery (2001, p. 1), "the word 'usability' has become a catch-phrase for products that work better for their users, but it is difficult to pin down just what people mean by it". Nielsen (1993) briefly defined the concept as the quality attribute that evaluates the ease of use of HCI interfaces, and all system's elements that encompass and interfere with that quality. Further, Rosson and Carroll (2002) highlighted not only ease of use, but also ease of learning and user satisfaction. Furthermore, Bevan (2006) refers to the description provided by The International Standard ISO 9241-11: Guidance on Usability (1998), which defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". According to Welie, Veer, and Eliëns (1999), theses common descriptions of Usability are rather abstract, since cannot express Usability in one objective measure. However, they define the scope of the concept, which can entail conflicting goals with software product design, requiring sensible tradeoffs between the qualities of the system and the users' interface: functional but also simple and clear; guarantee both ease of use and ease of learning; with good performance and low cost but also intelligent and sophisticate; among others (Mayhew, 1999).

Budiu and Nielsen (2011) highlighted the importance of the concept, which they identified as one of the main drivers in the design of user interfaces of web-based applications. Personal computers are widely accessible for a broader diversity of users and these users are also using these computers for a large variety of different tasks. Hence, it is natural that users are familiarized with pleasant and approachable interfaces and they expect and demand to find these features on every interface. Consequently, a user-friendly and user-centred design has become an increasingly important feature of all modern systems, and the interface of a web-based application has become the key criterion for the utilization of a web-based application (Pearson, Pearson, & Green, 2007). As companies continually face competitiveness, they realize the benefits of developing products with user-oriented methods, instead of technology-oriented methods. As a result, usability has increased in popularity and importance (Bergaus, 2015).

Regarding evaluation of Usability, Nielsen (1993) defines these five dimensions as being the main areas of focus :

- **Learnability** – The system should be easy to learn and understand, regarding performing the basic tasks for the first time;
- **Efficiency** – After learning and understanding the system, the user should be able to efficiently use the system (i.e. relates to how quickly the user can complete a task);
- **Memorability** – If the user is re-using the system after a long period, he/she should easily remember how to proficiently use the system without the need to learn it again;
- **Errors** – The system should prevent errors from users when they are using it, and if the user commits some error while using the system, he/she can easily recover from the error (i.e. catastrophic errors must not occur);
- **Satisfaction** – The system should be pleasant and satisfying to use.

Fernandez, Insfran, and Abrahão (2011) studied the main Usability evaluation methods that have been employed in the evaluation of web applications, and denoted the wider use of the following five categories of evaluation, previously purposed by Ivory and Hearst (2001):

- **Testing** – This category includes all methods that consist in observing subjects interacting with a user interface, in order to identify Usability problems. An example of a method included in this category would be the "Think aloud protocol", where subjects verbally share their thoughts wile interacting with interfaces in order to perform a set of pre-defined tasks;
- **Inspection** – Refers to methods that involve an expert evaluator that assesses a user interface according with pre-defined Usability's criteria. An example would be the Heuristic Evaluation method, where evaluators judge a user interface according with well-establish Usability principles (the "heuristics");
- **Inquiry** – Methods that gather subjective input from subjects. For instance: interviews or questionnaires that allow gathering subject's preferences and feelings;
- **Analytical Modelling** – Evaluators and designers employ users' and interfaces' models in order to generate usability predictions. An example of this engineering approach would be the GOMS analysis, which consists in modelling the several elements of in the execution of the interface's tasks, allowing to identify and eliminate unnecessary users' actions, as well as comparing alternative interfaces;

- **Simulation** – Refer to methods that simulate user interaction, by using a simulation algorithm or by analysing usage data. For instance, the Information Scent Modelling methods, that mimics web navigation.

2.1.3 Data Visualisation and Infographics

Data Visualisation is a cross-cutting theme of several fields. In the field of Psychology, the phenomenon of visually perceiving data is studied, as well as the impacts that visual elements, such as colours and shapes, have on human perception. In the field of Computer Science, the concept is used in several manners (for instance, to represent data in the field of Bioinformatics). In Multimedia Design, the concept is used to build graphics and dashboards, which unfold in several elements, such as lines; bars; and coloured shapes (Aparicio & Costa, 2015).

Azzam, Evergreen, Germuth, and Kistler (2013, p. 9) define Data Visualisation as "a process that (a) is based on qualitative or quantitative data and (b) results in an image that is representative of the raw data, which is (c) readable by viewers and supports exploration, examination, and communication of the data". The representation of data or knowledge into an image, that should be interpreted here as a combination of visual elements composed of colours and shapes (e.g. interactive graphic), includes the careful selection of visual elements that better represent the data at issue, according with what is meant to be communicated and with the goals of the viewers and explorers that will have access to the data/dataset through the created and designed image. In other words, refers the process of designing the final representation of data, according with well-considered and thoughtful criteria (Azzam et al., 2013). Data Visualisation embraces the well-accepted and establish notion that the human brain is more capable of perceiving patterns and relationships if data is visually encoded, as opposed to words and numbers. For instance, a bar chart representing the revenue of several companies (i.e. the bars' length is proportional to the revenue's values) could better illustrate the proportion of the differences of revenue between companies than the numeric representation of those values alone (Siricharoen, 2013). Moreover, some studies refer to research that suggested that, for most people, visual memory is more persistence than auditory or verbal memory (Zinovyev, 2011). Data Visualisation's main objective is to provide an efficient and engaging graphical/visual display for the purposes of summarizing and reasoning about the represented data/information (Harrison, Reinecke, & Chang, 2015; Zinovyev, 2011).

The domain of Data Visualisation has growth with the use of Internet and new technologies. This growth, in significance and usage, has increased the complexity of Data Visualisation, especially because the technology and the quantity of data used in the field have also increased. This leads to the advent of new sophisticated approaches and computing techniques capable of handling and be applied to large volumes of data. Consequently, the field is now focus of study in different expert research fields, such as Advanced Graphic Design, High-performance Computing and research in cognitive perception of visual imagery (Womack, 2014). With the overflow of information that can be observed nowadays, visual representation of data provides a useful enhancement, helping Web users to understand and process information more clearly (Siricharoen, 2013). Consequently, many Websites provide Data Visualisation tools, which help users exploring the data on their own. Usually, Infographics are the form of Data Visualisation provided (Azzam et al., 2013). Womack (2014) sees modern Data Visualisation as the vast variety of interactive infographics used in the Web.

An Infographic is a "graphic visual representation of information, data or knowledge intended to clarify and integrate difficult information quickly and clearly (…) In Human and Computer Interaction, infographics can improve user cognition by utilizing graphics to enhance the human visual system's ability to see patterns and trends" (Siricharoen, 2013, p. 69). Other definitions do not exclusively highlight visual elements, and define Infographics as being a combination of text, images, and charts (Saleh, Dontcheva, Hertzmann, & Liu, 2015). An example of an Infographic could be a graphic about crime statistics, in a particular city, provided by a newspaper's Website. Some studies refer to Infographics, Data Visualisation and Visualisation as synonyms, which describe the mechanisms by which humans visually use, illustrate, recognize and communicate data, with the main goal of effectively and clearly communicate, using graphical means (Siricharoen, 2013). Other studies see Infographics as part of Data Visualisation, where Infographics refers to the representation of quantitative/statistical data with charts and diagrams, and do not represent other domains of Data Visualisation, such as advanced statistical methods applied for creating visualisations for high dimensional data containing qualitative and quantitative information, as well as connections and subjective associations between data elements (Zinovyev, 2011). In other words, Infographics refers to graphic and visual representations of information, whereas as Data Visualisation refers to the process of creating and designing those visual representations (Ru & Ming, 2014).

Within the field of HCI, Data Visualisation is used in several domains. For instance, news' websites; blog publishing statistics; Geographic Information Systems; Marketing (e.g. as a tool that allows companies to communicate their revenues); healthcare statistics; among others (Siricharoen, 2013). Based on its application, Siricharoen (2013) defined the following three types of infographics:

- **Statistical Based** – Charts; Diagrams; Tables; etc. For instance: pie charts and bar charts;
- **TimeLine Based** – Shows data according with a sequence of events of chronological relations. For instance: time-series charts or year-by-year paragraphs;
- **Location or Geography Based** – For mapping and geographical purposes. For instance: a digital city map localizing tourism spots.

2.2 Mental Workload

2.2.1 Concept

Mental workload (MWL) has been used in several contexts in order to predict, assess, and represent human's and systems' performance (Cain, 2007). Context and disciplines include, for instance: Human Factors; Engineering and Educational Psychology (Gwizdka, 2010). This multi-disciplinary use and scope of MWL results in the slender possibility of defining MWL in a universal way, even in same field of study (Gwizdka, 2010). With the recent increasing use of interactive computer systems, the concept gained importance in the field of Computer Science (Longo et al., 2012).

2.2.1.1 Supporting theories and definition

The definition of MWL, also known as Cognitive Load, is applied in several fields and depends on the context (Wästlund et al., 2008). For instance, the term "Cognitive Workload" is used in the context of the Multiple Resource Model of Attention defined by Wickens (1991) in the early nineties, whereas the term "Mental Workload" is more used in relation to the model of Limited Capacity for Attention define by Egeth and Kahneman (1975) in the seventies. In brief, the models differ particularly in their perspective towards attention, as one – the Limited Capacity Model - bases its premises in the notion of a single dimensional resource of finite capacity for processing information, whereas the other, Multiple Resource Model, bases its premises in the notion of a capacity composed by multiple-dimension resources (Xie & Salvendy, 2000). The later, Multiple Resources Model, prevails as the most relevant nowadays (Xie & Salvendy, 2000). Moreover, in the context of the Limited Capacity Model theory, the focus is currently more related with working memory than attention (Paas, Renkl, & Sweller, 2003).

Since these two theories are related with methods for measuring MWL (that will be addressed and covered later in this literature review chapter), it is important to describe both in more detail. Starting with the Limited Capacity Model theory, the premise is based on the notion that the human brain has a limited-capacity "processor" that can be shared, up to a limit, across multiple tasks. The general conviction of this theory is that task demand – i.e. the resources that the task requires in order to reach a certain level of performance - is not fixed. In other words, mental resources, that are limited, are allocated as required and in accordance with task's difficulty or the level of performance required. This theory suggests that the remaining recourses (i.e. not in use) can be allocated to other tasks, and that these resources have a certain degree of elasticity, which allows them to be allocated, dislocated and relocated according with the demand imposed by the task(s) (Egeth & Kahneman, 1975; Wickens, 2002). Therefore, and according with this theory, there is a linear relationship between the allocated capacity and the task's performance, in the sense that the allocate capacity is viewed as a single dimension resource (attention or memory) that is linearly related with task's performance and demands, and can be used concurrently, to a limit extent (Wästlund et al., 2008). This particularity links this theory with the dual task method, a Mental Workload measuring method (Wästlund et al., 2008), which uses secondary/additional simple tasks in order to verify if a operator's single dimension resource pool is fully allocated to a certain primary task (later, this method will be further explained).

On the other hand, the Multiple Recourse theory purposed by Wickens (1991) does not view human's mental capacity as a single resource pool, but rather as a pool of distinct resources that can be allocated currently and are used/distributed differently according with task's characteristics. In contrast with the Limited Capacity Model theory, this model defends that it is only possible to perfectly time-share resources between two tasks it the overlapping resources demanded between tasks are not of the same type. For instance, driving a car and writing a text message are two tasks that consume the same resource pool (i.e. visual demand) and, therefore, a decrease of performance of both tasks might occur when performed concurrently, which is not in line with the Limited

Capacity Model theory, that claims that with low capacity demands, performance of both tasks, or at least one task, can be optimized. Moreover, the theory defends that two tasks, demanding different cognitive resources, request the use of different recourse pools that can be allocated independently. For instance, walking and reading at the same time demand different pools of recourses that do not interfere or limit one another, regardless of the demand's levels imposed by each concurrent task. In more detail, Wickens' theory also claims that mental resources are limited, but adds that mental capacity uses different resources in nature, that can be somehow independently allocated or redirected by different types of attention according with limitations defined by four factor dimensions (Wickens, 1991), illustrated by the below figure.

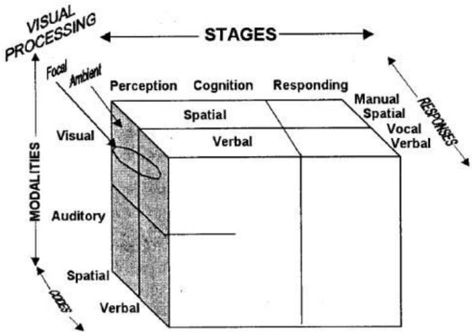


Fig. 2.1 Three-dimensional representation of the structure of multiple resources. The fourth dimension (visual processing) is nested within visual recourses" (Wickens, 2002, p. 163)

These two theories, Multiple Resource Model and Limited Capacity Model, support the main research theories that aimed to define and understand the psychological concept of MWL, also referred to as Cognitive Load. Although there is no universal or unique definition, MWL can be described, in a simplified way, as the amount of mental effort needed by someone (operator) to execute a task over a given period of time (Xie & Salvendy, 2000). Therefore, "mental (or cognitive) workload is a way to measure the mental demands of complex systems" (Zhu & Hou, 2009, p. 387). However, and in line with the aforementioned theories, this definition of the concept is very simple and limited, and more complex and thorough definitions exist across different fields (Cain, 2007).

Cain (2007), who reviewed literature that had studied the concept, stated that there is no formal definition for the concept, concluding that, in general, "Workload can be characterized as a mental construct that reflects the mental strain resulting from performing a task under specific environmental and operational conditions, coupled with the capability of the operator to respond to those demands. Operational definitions will likely continue to be proposed and tested, but unless an imperative need arises for a universal definition, each field and perhaps each investigator will continue with their culturally preferred definition of workload" (Cain, 2007, p. 4-3). Tsang and Vidulich (2006), who studied the concept in relation to situational awareness, highlighted that Mental Workload, in recent years, has been increasingly perceived as a multi dimensional concept, which comprises and includes factors related with the following three elements: the complexity and demands associated with the task; the ability/skill and mental state of the operator; and the situation and context enmeshed with the conditions that surround the task and the operator at the time the task is being executed. This intrinsic relation between task's characteristics, operator's traits, and conditions under which the task is executed is also highlighted by Cain (2007), who, as mentioned, reviewed the major works and conclusions of the research performed in the past decades around the matter.

In summary, MWL is, historically, a concept mostly used in the fields of Psychology and Ergonomics, and has been defined has a multidimensional complex construct related with the mental demand resulting from the combination of several factors: task's requirements and circumstances; operator's skills and behaviour (Longo, 2012).

2.2.1.2 Importance

Although the definition, factors, mechanisms and measurement methods of MWL are rather disparate, Cain (2007) highlighted that the importance of measuring MWL is consensual across different fields: succinctly, the capacity of enabling to predict the operator's and system's performance related with a certain task. In all fields of study, despite the different definitions used when defining MWL, the concept was linked to task performance, in the sense that there is a MWL threshold limit for which a certain operator, performing a certain task, will face a decline in terms of performance (Albers, 2011), as illustrated in Figure 2.2.

Although Figure 2.2 only refers to the decrease of performance when the phenomenon of overload occurs, it is also well-accepted that underload levels affect performance as well, and that there is an optimal range of MWL that should be achieved in order to maximize performance (Xie & Salvendy, 2000). In other words, "the workload level experienced by an operator can affect task performance. This effect can be caused by either excessive or reduce Mental Workload. Thus, estimating workload levels can help isolate sources that affect performance" (Cao et al., 2009, p. 113). The occurrence of overload can increase error rates and response times and, consequently, fewer tasks can be completed due to the decrease of the operator's mental capacity when dealing with other activities or tasks (Huey & Wickens, 1993). On the other hand, underload (i.e. low amount of MWL) can increase reaction times and lead to loss of attention (Cain, 2007). The common conviction, between MWL researchers in the different fields of application, is that although mental demand should be optimized, physical demand should always be kept to a minimum (Young & Stanton, 2002). The relation between performance and high and low levels of MWL is illustrated in Figure 2.3.

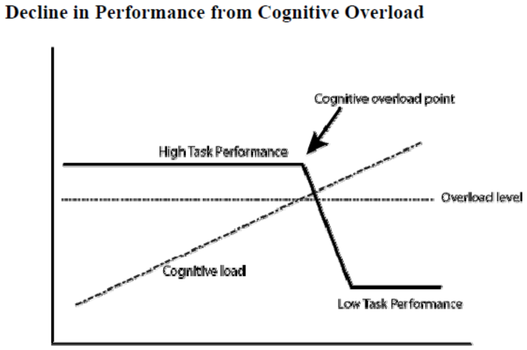


Fig. 2.2 "When the Cognitive Load level (increasing line) crosses the overload threshold (dotted line); the performance (heavy line) drops abruptly." (Albers, 2011, p. 27)

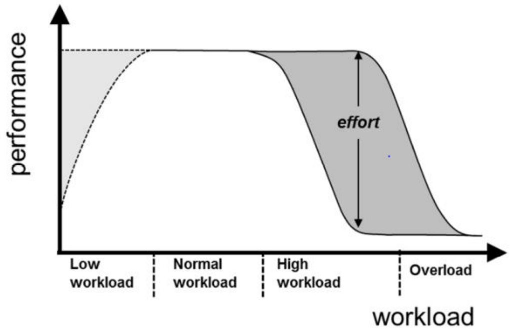


Fig. 2.3 "Relationship between performance and workload" (Chen et al., 2012, p. 5)

Furthermore, MWL is linked with frustration and stress (Zhu & Hou, 2009). And factors such as time pressure and lack of sleep can contribute to the decrease or limitation of the operator's resource pool and, consequently, can affect MWL (Tracy & Albers, 2006). Longo and Barrett (2010) reviewed literature regarding the factors that are linked with MWL and summarized the main factors with Figure 2.4.

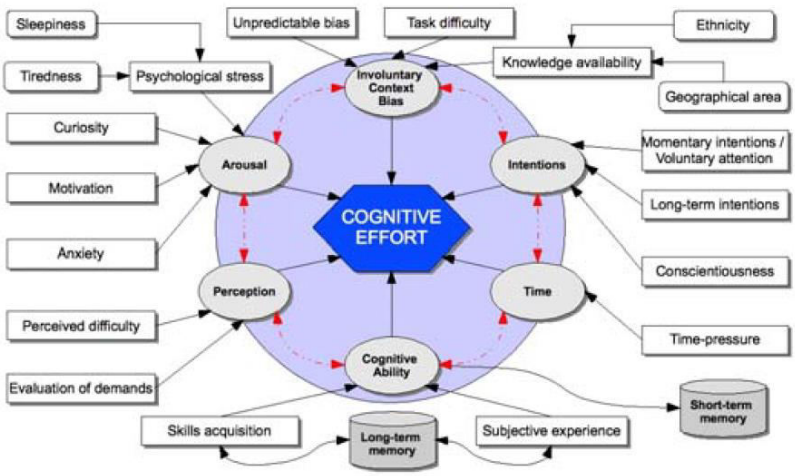


Fig. 2.4 Summary of the main Mental Workload's influencing factors (Longo & Barrett, 2010, p. 67)

In summary, the increasingly use of technology and HCI systems in the past years has boost the complexity and variability of graphical user interfaces. Hence, the likelihood of a system reaching or surpassing the limits of operators' capacity has also increased, reinforcing the need of studying and defining reliable methods of measuring MWL for different HCI interfaces' designs (Xie & Salvendy, 2000). In addition, Xie & Salvendy (2000) suggested that models for predicting MWL, in the context of HCI interfaces, should be applied in the early stages of the designing process, allowing to early standardize and quantify the mental demand levels and elements that should be specially taken in consideration.

2.2.2 Measures and Methods

Wierwille and Eggemeier (1993) divided MWL measurement methods in three main groups:

- **Subjective Measures** – Subjects auto-assess their MWL by rating a set of dimensions, within pre-defined scales, in relation with the execution of a task performed immediately before;
- **Physiological Measures** – Subjects have some physiological characteristics measured while performing a task. As, for instance, eye activity and heart rate;
- **Performance Measures** – The subjects' MWL is assessed according with the performance reached in a primary or for a secondary task (e.g. error rates; task completion time).

In the field of HCI, all these types, which will be individually described in more detail in this sub-chapter, have being used. For instance, subjective measures (e.g. NASA Task Load Index), which basically consist in asking the end-user, after completing a task, to self-assess the MWL imposed by rating a selected criteria of dimensions, were used in (Longo et al., 2012), a study that aimed to link Usability with MWL. Another example in the HCI field, using different measurement methods, was reported by Albers (2011), were MWL was assessed by using a performance based technique: dual/secondary task. Lastly, there are also studies using physiological measure methods. For instance, (Kamm & Gevins, 1995) used electroencephalography, which detects electrical activity in the human brain, to measure MWL demands in HCI tasks.
The following sub-chapters in this section describe, individually and in more detail, the three groups of MWL measuring methods.

2.2.2.1 Subjective Measures

This type of MWL measures basically includes methods that aim to measure MWL by asking operators to estimate, within a given scale, aspects of task difficulty after the exaction of a task or a set of tasks. Since this estimation/rating can be viewed as an afterthought (it is requested after performing the task), and since perception of the difficulty and the values of each ranked dimension depends on the individual (emotional state, motivation, attitude), this methods are seen as subjective (Cain, 2007; Wierwille & Eggemeier, 1993; Xie & Salvendy, 2000).

A subjective method usually assesses and considers several dimensions such as effort and performance, but there are methods that only assess one dimension (Wierwille & Eggemeier, 1993). Moreover, in order to benefit from a fresh recollection of the experiment, these methods are usually applied shortly after the operator performed the experiment, by requesting the operator to self rate a single or multiple dimensions/aspects of the imposed mental demand (Young & Stanton, 2002).

The following list describes the main methods that belong to this category (Cain, 2007):

- **NASA Task Load Index (NASA-TLX)** - This method was specifically designed in order to measure MWL for operators of human-machine equipment such as communication stations; aircraft cockpits and command control systems, and also to be used in laboratory tests (Tracy & Albers, 2006). The method consists in asking operators to self-assess themselves considering six rating scales: mental demand; physical demand; temporal demand; performance; effort; and frustration (Cao et al., 2009). These six dimensions were identified by a long research program that aimed to detect factors that directly interfere with MWL in the context of different tasks and fields (Hart, 1988; 2006). This technique is used in this study and its calculation and aggregation formula is described in the Design Chapter (page 59).
- **Subjective Workload Assessment Technique (SWAT)** – Like NASA-TLX, this is a multidimensional measurement method. Cain (2007) identifies this method as, probably, the most common subjective method reported in literature. This technique considers the following three dimensions: Time Load; Mental Effort; and Psychological Stress. Furthermore, this technique aims to minimize the subjective aspect of subjective methods, by previously asking subjects to give an ordinal ranking, from lowest to highest, to all the possible combinations of these three dimension’s values, where each dimension has three possible levels (with corresponding descriptions). This allows assessing MWL taking in consideration the relative importance that subjects individually give to the combined three dimensions (Cain, 2007).

- **Workload Profile** – This method is intrinsically related with Wickens' multiple resource theory and, consequently, its dimensions are directly linked with the dimensions that are purposed by the theory: Solving and deciding; Task and space; Verbal material; Auditory attention; Speech response; Response selection; Visual attention; Manual activity (Wickens, 2002). Tsang and Vidulich (2006) identified this method as being very reliable regarding the evaluation of different tasks' demands. This method is used in this study and its calculation and aggregation are described in the Design Chapter (page 59).
- **Rating Scale Mental Effort** – This is a single dimension method and, therefore, more related with the Limited Capacity Model theory, previously described in page 13. These method simply request subjects to self rate the required effort that the performed task(s) had demanded. The rating consists in a horizontal line, from 0 to 150, with the following labels distributed unequally across the line: "Absolutely no effort"; "Almost no effort"; "A little effort"; "Some effort"; "Rather much effort"; "Considerable effort"; "Great effort"; "Very great Effort"; "Extreme Effort" (Silva, 2014). Silva (2014) reviewed studies that identify this method as having a good degree of sensibility, despite its simplicity.

In summary, although this type of methods has been considered useful and capable of accurately detect and predict MWL – specially in low performance situations, i.e. underload and overload – they should also be viewed as a not 100% reliable afterthought since they are applied after the completion of the task(s) for which the operator's MWL is analysed. Moreover, these methods are considered subjective due to their self-assessment aspects (Tracy & Albers, 2006). Additionally, these methods have the limitation of not being able to assess MWL in real time (Wästlund et al., 2008), which is an important aspect in the field of HCI, specially for long tasks, since the subjective aspect of these measures is more suitable of compromising subjects' perceptions. Further, it could be useful to detect peaks of MWL when the user is interaction with a specific menu or element/step in a web user interface. Despite these limitations, these methods are considerable reliable, and easy to apply (Cain, 2007).

2.2.2.2 Physiological Measures

Physiological indicators have been known to be associated with mental activities such as emotion, cognitive load and attention. Hence, these indicators are expected to vary along with factors such as frustration and task difficulty. Consequently, physiological indicators would be, theoretical, suitable to use in the evaluation of users' reactions in the context of HCI (Ward & Marsden, 2003). The use of these indicators, such as brain electrical activity, in the context of HCI is normally used in the two following ways: detection of momentary changes in response to certain events; tracking physiological indicators over a period of time in order to make comparisons and associate patterns of physiological responses for distinct circumstances. Both approaches are based in the notion that physiological events observed in HCI are no different than any other physiological events resulting from other types of stimuli and, therefore, can be intuitively used to detect negative and positive emotions.

Ward and Marsden (2003) overviewed previous studies and suggested that it is not straightforward to use these indicators since, between individuals and occasions, readings are inconsistent and, due to the difficulty of interpreting and standardise signal analysis, it is hard to quantify, identify and correlate physiological responses related with MWL. Even hard to control factors (e.g. previous subjects activities; subjects' skin structure; room temperature) cannot be isolated or controlled, even though researchers know that these factors directly affect the measuring process that is taking place. Moreover, similar physiological readings (e.g. surprise; frustration; workload) can be associated with different mental events. In other words, it is not easy to determine the origin and relevance of the physiological signs measured during experiments (Ward & Marsden, 2003). Consequently, it appears to be impractical to use physiological measures as a reliable and methodical approach for evaluation user interfaces in the field of HCI. Further, these measurement methods are seen as being expensive, since require proper equipment and specialised training. Cain (2007) reviewed studies that suggest that, in the context of MWL measurement, and due to its complex nature, psychological measures should only be applied with solid knowledge of what physiological properties and responses can be taken in consideration, regarding particular situations and scenarios. Cain (2007) identified the below methods as the main physiological measures used and studied in the context of MWL evaluation:

- Electroencephalography;
- Eye Movement;
- Heart Rate;
- Respiration.

The main advantage of these methods is that they are capable of continuously measure the operators' state in an objective way (Wästlund et al., 2008). However, Cain (2007) reported studies where several physiological measures were used together with SWAT and NASA-TLX. Those studies concluded that the two subjective methods could be considered valid and reliable, in contrast with the physiological methods evaluated. Eye movement, eye blink, heart rate and blood pressure seemed to be insensitive to workload variance. Further, heartbeat, although reliable and valid, was found to be affected by other mental activities that could interfere with the MWL measurement purposes.

2.2.2.3 Performance Measures

Performance measures can be classified in two main categories (Cain, 2007):

- Primary task measures;
- Secondary task measures.

The former - primary task measures - directly focus on the performance of the task for which the imposed mental demand is being study. Measures such as speed, accuracy, response times and error rates are used. The later - secondary task measures - basically rely on the performance of a secondary task, simpler than the central/main task that is being study. The secondary task usually interferes concurrently with the main task and its performance is seen as insightful in the sense that allows detecting the operator's overload related with the main task (Cain, 2007).

The essential assumption and foundation of primary task measures is the notion that the increment of mental demand directly interferes, in a negative way, with task's performance. In other words, the more demanding and difficulty the task is, the less task's performance will be observed (Haga, Shinoda, & Kokubun, 2002). Tsang and Vidulich (2006) identified the following as the major issue of using these methods: they link performance almost exclusively with the MWL derived from the operator's traits, ignoring the interference associated with the task itself (e.g. a web interface poorly designed) and the conditions that surround the task and the operator at the time the task is being executed.

Regarding secondary task measures, they essential require that a subject or operator performs two tasks concurrently, one being the primary task and aim of the study, and the other being a secondary task that is used to assess the MWL imposed by the primary task. The idea is that if the operator has his full mental capacity allocated in the primary task, the performance of the secondary task will be affected, despite how simple the secondary task is (Cain, 2007). This approach seems to be directly related with the

Limited Capacity Model theory (i.e. does not take in consideration the multiple pools of capacity defended by the Multiple Resource theory, which claims that if different pools of recourses are used, performance of concurrent tasks will not be affected). This approach is related with the dual task paradigm (Wästlund et al., 2008) and is based on studies that suggested that the reaction time related with a secondary task is appropriate for measuring MWL, arguing that the amount of mental capacity requested in the primary task directly affects a secondary task, as simple as it may be, in the following way: the more MWL is demanded in the primary task the less reaction time will be observed in a secondary task (Verwey & Veltman, 1996). Examples of secondary tasks could be (Cain, 2007):

- Rhythmic tapping;
- Random Number Generation;
- Spatial reasoning;
- Probe reaction time.

In conclusion, performance measures essentially assume that MWL only earns importance when it affects system's performance. The main disadvantage of primary performance methods is the reported incapability of distinguishing performance of multiple parallel tasks. Further, secondary performance measures are considered intrusive and more indicated for short tasks and for measuring the operator's capability of dealing with additional tasks. The main advantage of secondary measures is their capacity of measuring MWL during task's execution (Longo, 2016).

2.2.2.4 Summary: Advantages and disadvantages of the different categories of MWL measurement methods

The following list intends to summarize the main concerns and benefits highlighted by the reviewed literature:

- Subjective measures are essentially viewed as easy to implement and use, having good subjects' acceptance and strong reliability. On the other hand, these measures are subjective since are performed after task execution and depend on the subjects' perception. Another disadvantage is their incapability of obtain MWL in real time, which could be useful in the field of HCI, since it would allow identifying critical design steps and elements. A significant advantage of these measurement methods, compared with the other categories, is the possibility of considering several cognitive dimensions linked with MWL.

- Physiological measures are capable of measuring MWL in real time. However, in order to be reliable they should be used in strongly controlled experiments, since they depend on external factors (e.g. room luminosity; subject's tiredness) and can be triggered by other mental activities (not related with MWL). Plus, these measures are expensive and difficult to implement.
- Primary performance measures are easy to implement but basically assume that performance mainly depends on MWL, ignoring the task's characteristics and the users' traits (e.g. age; skills) that affect performance.
- Secondary performance measures are capable of measuring MWL in real time. However, they can be intrusive and unable of assessing the multi-dimension nature of MWL.

2.2.3 Application

MWL has been used in several contexts in order to predict, assess, and represent human's and systems' performance (Cain, 2007) and is, historically, a concept mostly used in the fields of Psychology and Ergonomics, where it has its roots (Longo, 2012). Context and disciplines include, for instance: Engineering; Educational Psychology; and Human Factors (Gwizdka, 2010). Within specific contexts of use, several examples can be given as, for instance: designing control panels for industrial plants; and devising control panels for aviation (Albers, 2011). Cain (2007) highlighted the long application of Mental Workload in the fields of Psychology and Ergonomics, reporting that the concept has being studied for the last four decades and has been used for long in those fields, very often applied in the automobile and the aviation industries. In particular, Cain (2007) reported studies in the field of aviation, where pilots' MWL was assessed regarding the interaction with cockpit interfaces, not only to address and identify situations of pilots' overload, but also to detect situations of underload, which are related with a dangerous lack of attention promoted by the continuous automation of aircraft systems. Other studies are reported, with relevance to the field of car industry where, in the scope of automotive safety, drivers' MWL and performance was assessed in several situations. Another relevant field of MWL application is the field of Healthcare where, due to the emergent replacement of traditional paper-based patients' data with Electronic Health Record Systems, the concept gained importance due to its connection with performance, which is critical in the field, since practitioners are not expert users regarding new digital technologies (Longo & Kane, 2011).

Since that, in the field of HCI, there are many tasks that demand the use of complex cognitive activities, it was only natural to assume that it is important to study MWL in the context of HCI, in order to avoid overload demands and in order to increase performance when HCI interfaces are used. Naturally, this can only be handled with effective methods of measuring MWL in the context of HCI (Zhu & Hou, 2009). The awareness of MWL importance in this field is significantly supported by the notion that, with the increasing complexity and use of computer interfaces in modern life, it is not easy to design interactions as simple as possible or even effortlessly understand what an optimal user interface is. As a result, several studies concentrated their efforts on measuring MWL in the field of HCI (Tracy & Albers, 2006).

This notion of seeing MWL as a key element to take in consideration in HCI is reinforced by research performed by Wästlund et al. (2005). This research suggested that, regarding comprehension and mental demand, it is better to present information in a paper document than to display the same information in a computer screen. Yet, Noyes & Garland (2003) proposed that any findings that reached similar conclusions are related with improper digital interface layouts (i.e. which do not optimize human comprehension and perception of information), claiming that digital interfaces entail using, organizing and displaying information in a different way (than what is traditional used in the context of paper-related documents). Interested in analysing the uneven opinion of these and other similar studies, Wästlund et al. (2008) studied the effect of Web page's layout on MWL, concluding that performance increases and MWL decreases by optimising page layout for onscreen viewing. This study emphasised the importance and usefulness of MWL in this current era of increasing use of digital means for consuming and production of information, primarily because previous literature - for instance: Belmore (1985); Mayes, Sims, and Koonce (2001) – had claimed that using digital screens reduces cognitive recourses due to imposing navigation activities and, therefore, is it crucial to optimize digital interfaces as best as possible, focusing in designing "computer interfaces that optimize human information processing by minimizing MWL, rather than postulating that one medium is superior to the other" (Wästlund et al., 2008, p. 1230). Interestingly, MWL has not only been used to assess different HCI interfaces but also to assess the benefits of replacing paper based processes with computerizes systems. One example can be found in the study reported by Wastell and Newman (1996), who used physiological measures of MWL to evaluate the benefits of ambulance control operators using a computer bases system instead of the theretofore paper based system. This study, focused in a HCI critical system, emphasized the importance of assessing MWL in the field of HCI, especially for critical systems. The study of MWL in the context of HCI continues to be investigated in recent studies. For instance, Gwizdka (2010) set an experiment, composed of look up tasks, in order to access MWL during all steps and phases of a web based search. This type of study allows assessing and identifying particular elements of the designed interface instead of assessing the global MWL demanded by the task. The study of mental work load in the field of HCI will be discusses in more detail in the next section.

2.3 Related work on Mental Workload and Human-Computer Interaction

2.3.1 Mental Workload Evaluation in HCI

This section intends to give the reader a solid notion of the research that has been performed in recent years regarding the study of MWL in the field of HCI. The aim of the section is (rather them exhaustively present all literature related with the topic) to deliver a comprehensive and wide notion of the different approaches and concerns that have been reported in scientific literature. In other words, the research reported in this section will serve as an illustration of the main goals, measurement methods, findings, and assumptions that have being studied in the field.

First, this section will present some studies that consider the MWL's relevance and role in the field of HCI. Lastly, it will be covered research that argues that MWL should be used as an important indicator in the evaluation and design of HCI interfaces.

2.3.1.1 The role of Mental Workload in HCI

Wästlund et al. (2008), already referred in previous chapters, illustrates the importance that recent research has given to MWL in the context of HCI. This study used MWL as a tool for evaluating the quality of different page layouts for displaying certain information. The study based its premises on the notion that typical presentation modes (i.e. organization, layout) of displaying information in paper documents are not proper for use in digital presentations and, therefore, is essential to evaluate digital interfaces and layouts in order to determinate optimized displays for digital consumption and production of information. This study used a performance measure - dual task - as a MWL measure. The secondary task consisted in "popping up" a second browser's window while the subject was performing the primary task. The popped up window simply consisted in using a "close" button to close the window or answering "yes" or “no” regarding the comparison of two numbers (e.g. “Is the first number bigger than the second?”). The study concluded that optimizing page layout (accordingly with screen size) reduced the levels of MWL. The tasks focused essentially in assessing reading comprehension and working memory. Another study that illustrates the role of MWL in HCI, outlined by recent research, was performed by Gwizdka (2010), who aimed to further analyse the role of MWL in the field, by assessing the MWL during all the steps taken by the subjects while performing look-up tasks. This study’s results suggested that MWL variance can be detected for different stages of the loop-up search. Another experiment that aimed to study MWL variance across all the steps of a web based task was performed by Albers (2011), who used tapping as a way of measuring MWL during all task’s execution and, therefore, was capable of correlating MWL variances with the subjects’ interaction with different components of the designed interface. Regarding the other category of MWL measure methods capable of obtaining MWL in real time - physiological measures - one study that can be given as an example was performed by Wastell and Newman (1996), who used MWL measures in order to assess the Usability of computerized systems, aiming to evaluated an ambulance computer-based control system and its operators, on critical occasions.

The notion that MWL is an important indicator when assessing interaction with computer based devices is not delimited to coded interfaces and can also be applied in the evaluation of the effect that MWL has in different devices and users. For instance, Patrick Rau and Hsu (2005) performed an experiment where subjects of several ages where asked to perform several information browsing tasks. The same tasks were performed in three different computer based devices that encompassed different ways of interacting with technology: mouse and keyboard; touch screen and handwriting recognition; and voice control and voice input. The main goal of the study was to assess performance between older and younger subjects, taking in consideration subjects’ skills and expertise regarding the interaction with computer based designs. The goal was to understand what should be taken in consideration when designating and implementing computer interfaces, in respect of aged-related skills and cognitive abilities. The study took in consideration the following measures: performance, user errors; satisfaction and Mental Workload (assessed by using the NASA-TLX method). Although the results showed mixed conclusions between interfaces and age/expertise, they support the notion that high and low MWL levels are linked with bad performance.

Equally important, emphasizing the importance of MWL in the field of HCI, Heger, Putze, and Schultz (2010) purposed that a system capable of assessing MWL in real time would allow intelligent systems to adapt to the operator's mental state, which is something that the authors claimed that is not taken in consideration in most of the systems. This idea is also defended by Grimes, Tan, Hudson, Shenoy, and Rao (2008), who performed an experimental study that aimed to assess the accuracy of brain activity (obtained using Electroencephalograph) as a method for measuring MWL (the study's results supported this premise). Although the idea of both studies is interesting, the purposed method consists in physiological measures (brain activity through an electroencephalogram) and, therefore, the costs of implementing such mechanisms cannot be neglected.

In the context of Web Design, the importance of MWL in HCI was reinforced by Evans and Fendley (2017), who suggested that, even a "well-designed" interface (i.e. that comply with good design requirements an experts' opinion) are not capable of avoiding high peaks of imposed MWL, due to the daily increasing creation of new data that has to be handled by web interfaces. The authors suggest that MWL could be a key role in assessing interfaces regarding human cognitive capacity. This importance is also stressed by Van Orden et al. (2001), who referred that web interfaces need to be changed more often nowadays and that MWL plays a critical role in identifying the impact of those changes. Hence, a possible correlation between user interaction and MWL would be really useful for designers when evaluation and predicting the impacts of web design changes.

An interesting approach for assessing MWL in the field of HCI, which achieved practical results, was performed by Yamagishi et al. (2007), where MWL, along with other factors, was measured for operators of a Nursing Operational System in two moments: after initial training and after a month of using the interface. This allowed considering MWL in the evaluation of learning curves in the context of HCI systems. This tactic was also used by Felton, Williams, Vanderheiden, and Radwin (2012), who used NASA-TLX (MWL measurement subjective method) as the main procedure for evaluating training in a brain-computer interface. These researchers argue that it "is not well understood how people perceive the difficulty of performing brain–computer interface tasks" (p. 526) and identify MWL as a feasible measure to better understand and evaluate this type of interfaces.

2.3.1.2 Mental Workload as a measure for evaluating HCI interfaces

As mentioned, in the field of HCI, there are studies that link a web page design with the MWL needed to use the web page's functionalities and with the evaluation of the user interface's Usability (Longo, 2012; Longo et al., 2012). Longo (2012) argued that assessing and measuring MWL demands is extremely important for web designers, and should be taken in consideration in web interface designing processes as well as in the evaluation of web sites' Usability, adding that user behaviour and Usability can be usefully highlighted and understood by measuring the MWL imposed by web-tasks. In order to study this assumption, Longo (2012) performed an experiment using self-assessment techniques for both MWL and Usability. The experiment consisted in asking subjects to perform a set of tasks with different interfaces for Google's and Wikipedia's websites. By comparing the self-rated rankings filled by users, after completing the tasks, for MWL and Usability, the study's results did not show a good correlation between self-assessed MWL and rated Usability. These results corroborated the potential of using MWL in web-design, in line with the motives that prompted this study, since they showed that methods for measuring Usability hide high demands of workload, and vice versa. Further studies continued to identify MWL as a powerful aid in Usability measurement. Those studies consisted in performing experiments involving web tasks, alternative interfaces and rating methods for assessing MWL and Usability. The resulting outcomes suggested that the measured levels of Usability and MWL were not overlapping indicators of the interfaces' design quality, since a high score or variance in one does not also means a high score or variance in the other, which supports the idea that these methods assess different elements of web design and, consequently, they should be used together in the evaluation of web design (Longo & Dondio, 2015).

Further studies analysed the role that MWL and Usability play in the context of HCI. For instance, remarking that the use and complexity of web interfaces has been increasing in the past years, Albers (2011) advocated that users have less Cognitive Load capacity available to deal and use bad designed or unrecognized interfaces. Consequently, Albers (2011) purposed that measuring MWL during the use of web interfaces is a proper way to assess the Usability of that web site's design. In order to study this theory, Albers (2011) set an experiment that measured MWL using tapping as a secondary task, which allowed to measure MWL during all stages of the web task's performance. The study concluded that MWL cannot, by itself, be used to measure overall Usability. Adding that, however, that is a limitation of any Usability measure technique, and MWL measures should be used together with other Usability's tools and methods. Along with other studies, such as (Gwizdka, 2010) and (Xie & Salvendy, 2000), Albers (2011) also indicated that MWL cannot be considered without acknowledging the role of participants' traits in the levels of MWL measured. Adding that "Usability test designers must consider participant experiences when analysing the results" (Albers, 2011, p. 30). This conclusion is entailed by the experiment's results, which highlight that experience/technical users show less episodes of tapping variation (Table 2-1). In this study, tapping variation is seen as being linked to MWL increase.

Table 2-1 Number of users, per type of interaction, that showed noticeable tapping variation in the study conducted in (Albers, 2011)

	Link	Reading	Scrolling
Technical	5	6	4
Non-Technical	12	12	19
Total Instances	17	18	23

Similarly, Tracy and Albers (2006) also purposed using MWL as measure of web sites' Usability, stating that the design of HCI interfaces has to respect the limits of human cognitive capacities. The study, besides using the tapping measure as Albers (2011) did, also used the NASA-TLX method. In the end, the authors suggested that using MWL measuring techniques can be an additional way of evaluating Usability. Both MWL measuring techniques seemed to be coherent between each other's results. However, the tapping method allowed detecting and identifying peaks of high mental demands during the task(s). As a final illustration of research in the field, Gwizdka (2009), who studied the MWL in the context of web search tasks, also stressed that the dual task method allows measuring Mental Workload at discrete points of time, mentioning that this characteristic is an important factor when studying MWL in the context of web user interfaces and that, although physiological measures also allows to measure MWL in a real time, the cost and use of external devices end up to be impractical and expensive.

In conclusion, incorporating MWL as useful measure for evaluation human-computer interfaces has become an approved approach in the multi-disciplinary domains of HCI, such as Gaming (Evans & Fendley, 2017); Healthcare and Medicine (Longo, 2016; Longo & Kane, 2011); measuring learning curves for HCI systems (Yamagishi et al., 2007); and Pilot Simulators (Durantin, Gagnon, Tremblay, & Dehais, 2014).

2.3.1.3 Summary of Mental Workload importance in HCI

The studies reported before are given as illustrations of the major considerations and findings that can be found in literature. In order to provide clarity, the following list summarizes the main benefits identified by recent scientific research:

- MWL was identified as capable of being linked with task's performance;
- Measurement of MWL allowed to evaluate different web pages layouts (i.e. allowed optimizing interfaces' design);
- Real time measurement methods of MWL (physiological and dual task) allowed to identify issues in steps and elements of web interfaces;
- Besides allowing to evaluate the cognitive demands of different interfaces' designs MWL also proved to be useful to do the same in relation to the devices used and the operators' traits (such as age and skills);
- MWL was identified as a complementary technique to typical Usability evaluations;
- The argumentation that MWL can be used to evaluate the mental demands of interfaces has gained increasing importance since, in recent years, data that is presented in the same interface can often (e.g. daily) change in size, and nature. MWL assessment can be used to predict the impacts of those changes, complementing Usability assessment methods;
- MWL can be used to assess learning curves of users that started interacting with new software systems;
- Being able to track MWL in real time would be very important for critical systems and for the advent of adaptive systems.

2.3.2 Evaluation of Data Visualisation Interfaces

The experiment performed in this project included interfaces with different Visualisation techniques. The goal was to, in parallel with the main goal of finding a possible correlation between users' interaction and MWL, pursue a secondary objective: assess MWL as a measure for evaluating Data Visualisation interfaces. This section aims to go through, by way of example, some studies that aimed to evaluate different Visualisation interfaces. The performed literature review found out studies where MWL was used to assess web pages design - for instance, (Gwizdka, 2010) and Albers (2011) – but none that used MWL to assess web pages designs that used Data Visualisation techniques.

By way of example, a study that evaluated different visualization interfaces was performed by Kobsa (2001), who aimed to assess three different visualisation systems for multi-dimensional data. Two interfaces were table-based, but colorized and aggregated data differently. For instance, one used distinct colours to represent nominal and ordinal data. Other allowed the user, when clicking a value related with an attribute, to see the other records that had the same value for that same attribute. The third interface presented data through charts (the user could switch between scatter plots; histograms; pie charts; among others). The data used across these interfaces were based in three databases: anonymous data from an online web dating service; automobiles' technical data from the seventies; data about heavy metals concentration in Sweden. The 83 subjects, students and not experts in neither of the databases' fields, were asked to perform a set of tasks related with the three datasets. Each subject interacted with a unique visualisation system. This study used task completion time and correct answers ratio as its measurement indicators. Interestingly, the interfaces with bigger completion times were the ones with better correct answer ratios. The author suggested that this disparity was related with the fact that the differences in accuracy ratios were related only with 6 tasks, which suggest that the interface with less response times was suitable for quickly answering the tasks that had 100% correct answers across all interfaces, but failed to be suitable for the 6 tasks that gave the uneven accuracy ratios. The study concluded that the interface that used several possible graphics didn't show as good task accuracy as the table-based interfaces. This could be related with the inexperienced of users in using such interfaces (increased by having to choose which graphical representation would be the best). The two table-based interfaces seemed to be, alternately, better than the other one for specific tasks.

The effectiveness (task accuracy) and efficiency (response time) methods, used by Kobsa (2001) in order to evaluate different visualisation interfaces, are common methods in the field (Saraiya, North, & Duca, 2005). Another study that used these methods was performed by Cawthon and Moere (2007). In the study, online surveys and effectiveness/efficiency measures were used to evaluate eleven different visualisation techniques. Based on studies that linked aesthetics of Data Visualisation interfaces with task performance, the main goal of the study was to assess, using surveys, the aesthetics of the visualisation techniques in a quantifiable way, rather than coming from developers' and designers' personal judgments. The researchers suggested that aesthetics should be viewed as a major factor, in hand with task's performance, and not as a simple “nice to have” feature. Aesthetics were measured through a ranking and an “interfaces comparing” questionnaire. In general, the study showed a strong consistency between individual interface aesthetics rankings and comparison of aesthetics between interfaces, both given by the subjects. The correlation between performance and aesthetics was also significant, sporting the study’s initial assumptions.

However, these common methods of evaluating Data Visualisation interfaces are being questioned by other researchers. For instance, in the field of Bioinformatics Saraiya, North, and Duca (2005), aware that this field deals with very large datasets (that benefit from the use of Data Visualisation), purposed a methodology to evaluate bioinformatics visualizations different from the typical performance evaluation approach. The researchers of this study claimed that it is redundant to test a small set of variables in systems that consist in analysing thousands of variables. Adding that typical measures used in Visualization studies, such as task accuracy and task completion time, are only assessed through a limited set of predefined tasks and, consequently, the studies’ conclusions are unreliable when comparing different visualisation systems. Moreover, the researchers also suggest that these Visualisation studies are short-termed. The authors extend this opinion to other reported ways of evaluation interfaces, such as formative Usability testing (e.g. “think aloud” protocol) or expert evaluation. The authors highlighted that it is critical to use a methodology that allows practitioners in the Bioinformatics field to choose the best visualisation tools for their purposes. In the field, “eureka” situations and finding new insights in huge datasets (that are being analysed) are what add value to a Bioinformatics’ visualisation interface. The researchers purpose a new methodology that, in a simplified description, consists in comparing Visualisation systems by using the number of high-quality findings/insights (found by exploring the datasets with a specific Visualisation interface) as an improved evaluation method.

Scholtz (2006) is another study that addressed the need of creating new methods for assessing Visualisation interfaces. The study reported and gathered concerns of several authors and listed aspects that should be taken in consideration when developing new approaches for evaluation Visualisation systems. Besides mentioning aspects such as creativity and users’ situation awareness, the study identified two elements that are linked with the goals of the study reported by this document: cognitive complexity and extracting knowledge from user interaction. These concerns are also mentioned by Tory and Möller (2005), who questioned the usual approaches for evaluating the actual usefulness of visualisations, for real people performing real world tasks. The study questioned typical research studies, which are formal and performed in laboratory, claiming that those types of studies do not seem to be ideal for every situation. Above all, running such experiments consumes significant time and resources, and imposes restrict and controlled objectives in order to allow reaching well-structured conclusions. Plus, the authors added that those experiments are performed in early stages of design and, consequently, proper objectives and variables are yet to be defined. Moreover, the authors alluded to the simplicity of the user tasks typical defined for those experiments, which are low-level or simple cognitive tasks, rather than the high-level and complex cognitive tasks that would be performed in a real case scenario. Interestingly, this notion of "having" to design simple tasks for purposes of research was also observed in the experiment reported in this document (read the Design Chapter, page 50, for more information). As a superior evaluation alternative method, Tory and Möller (2005) defended that expert evaluation, within the scope of predefined heuristics, is a more reliable method for evaluation this type of interfaces. This claim is supported by research, listed by Tory and Möller (2005), where few Usability experts were able to find more Usability problems than a 50 subject's laboratory experiment. Nevertheless, the study still identified user studies/experiments as an important must-have evaluation approach, which allows identifying issues not detected by experts. However, the authors reinforce that that type of studies (laboratory experiments) should not have so much importance as experts' review methods.

A master thesis was found in the literature that considered Mental Workload as a method for evaluating different visualisation interfaces (Klingner, 2010). Mental workload is measured with eye physiological methods and the study is basically based on the perception that 80% of information is perceived by human brain through the eyes, especially in the context of Visualisation and HCI. The study conclude that MWL could be an useful measure, but that eye physiological measures are hard to measure due to external factors such as tiredness and room brightness. This idea is supported by Zinovyev (2011), who identified the over-complexification of graphical data and its representative visual elements (i.e. cognitive problems) as the major problems of designing Infographics. This study refers that the designing complexity as being related with the heterogeneity and the vast amount of the data used in this field, adding that the cognitive perception demanded when exploring Infographics is also related with user's traits (the paper exemplifies this by mentioning research that suggested that colour palettes are more accurately perceived by women than men).

In brief, the studies described in this section reported the following issues related with the evaluation of Data Visualisation interfaces:

- Typical Usability measurement methods, such as performance measures and surveys, are more limited in the context of Data Visualization, since this HCI domain entails multi-dimensional, multi-purpose and huge amounts of data, which hampers the possibility of defining real world tasks in the Design phases. These aspects minimize the value of the results performed in controlled laboratory experiments, where a small set of users and situations are assessed, and tasks are restricted, predefined and usually too much easy;
- Expert evaluation was mentioned as a better, but not sufficient, way of assessing this type of interfaces. However, this type of evaluation also lacks the full capability of predicting the nature and steps that will have to be taken to perform a task;
- New research presented insight measures as a more reliable way of examining and comparing different Visualisation techniques. However, standardize the quality of a good insight/finding is not straightforward;
- Thinking aloud methods are useful to understand users' behaviour, but do not scale well. Using tracked user interaction (e.g. mouse clicks) is seem as a more reliable way to study users' behaviour in the domain of Data Visualisation;
- Since Data Visualization representations of data can entail a more complex cognitive interaction than usual interfaces (for instance, making a purchase in Amazon), assessing MWL could be useful in the evaluation of this type of HCI interfaces.

2.3.3 Tracking of User's Interaction

Literature research, both in the scope of Usability (Woods, Velasco, Levitan, Wan, & Spence, 2015) and Psychology (van Steenbergen & Bocanegra, 2015), which aimed to review studies related with the evaluation of HCI interfaces in laboratory and in online scenarios, highlighted the usefulness and potential insight that can be obtained by tracking the interaction of hundreds of users. Those studies essential pointed out the benefits of collecting data from several users, in real case scenarios, compared with the costly and potentially thin sample of subjects that often are linked with laboratory experiments. The studies also mentioned that, naturally, it is harder to control some aspects of online experiments but, on the other hand, the reduced cost of involving a big number of subjects in an inexpensive way is very significant. Both studies stressed that it is imperative and important to develop methods of extracting knowledge from user tracked interaction in order to be able to compute, understand and take advantage of the benefits of online experiments (in terms of subjects reached, both in numbers and in heterogeneity).

The use of online interaction as a useful and insightful tool was already considered in studies that evaluated Usability (Ivory & Hearst, 2001). Several measures and indicators are mentioned, such as: pages per visitor; click streams; page-view durations; user's actions reproducing/tracking; and mouse position tracking. The potential of using this tracked data (i.e. users' interaction logs) in the evaluation of Usability can be expressed by the following advantages (Atterer et al., 2006; Ivory & Hearst, 2001):

- Reduced cost of Usability evaluation;
- Easily detect new users' behaviours that were not taken in account in the design phase;
- Automatically predict and aggregate errors and their impacts;
- Increased coverage of system's features;
- Easy comparison between different designs.

The aforesaid benefits, although seem to be based in behaviour analysis, could easily be translated into benefits that would help assessing Mental Workload in the field of HCI (based on the possible correlation between MWL and users' interaction, suggested in the study reported by this document). An example of a Usability study that used tracked user interaction was performed by Atterer et al. (2006), who, by consulting users' interaction logs (mouse clicks; scrolling; mouse position; and keyboard usage), were capable of understanding in what order and pace the elements of a web form were filled, identifying possible Usability issues. Using the same tracked user information, Guo and Agichtein (2012) studied user behaviour after obtaining results provided by web search engines.

In parallel, de Leeuw and Motz (2015) reviewed studies, in the field of Psychology applied in HCI, which used measures tracked by JavaScript (e.g. response times). The study claimed that "Although there has been some research on the accuracy and reliability of response time measurements collected using JavaScript, it remains unclear how well this method performs relative to standard laboratory software in psychologically relevant experimental manipulations" (de Leeuw & Motz, 2015, p. 1). The study mainly focused on the capacity of using JavaScript to record response times, which is a measure often used in the field. In order to assess JavaScript reliability the study performed an experiment with two different interfaces, one using a Web Browser environment (with JavaScript) and other using the Psychophysics Toolbox, which is a Matlab software package used in the field and for which several studies had ascertained its reliability. The subjects were asked to perform simple visual search tasks. In the end, although there were tiny differences between the timestamp of the response times recorded with JavaScript and with Psychophysics Toolbox (irrelevant milliseconds), the differences didn't affect the distribution of response times between subjects and, therefore, it was concluded that JavaScript is, probably, a reliable technology to use regarding this matter (i.e. response time tracking in the field of HCI applied Psychology).

Additionally, a research project that is very related with the goals purposed in the study reported by this document, was performed by Pimenta, Carneiro, Novais, and Neves, 2013, who studied the possibility of monitoring mental fatigue by analysing keyboard and mouse interaction. The research, supported by literature that linked fatigue with human error, highlighted how crucial it is to find a way to automatically detect mental fatigue in critical roles that depend on HCI activities. The study presented fatigue as a subjective concept, related with a group of combined symptoms such as attention loss; low performance due to skills limitation; mood; stress; sleepiness; among others. In addition, fatigue is also linked with the profile of on operator, which includes elements such as age; profession; and alcohol consumption. Methods of assessing fatigue, similarly to MWL (which one can argue that are intuitively related), include subjective assessment (questionnaires) and physiological measures (e.g. heat rate). The research, similar to the goals of the study reported in this document, also suggested that those methods are intrusive and costly and, therefore, it would be important to evaluate the possibility of using logs of user interaction to automatically detect operators' fatigue. A major difference between what is purposed by Pimenta et al. (2013) and what is purposed in the studied reported in this document, is that Pimenta et al. (2013) assumed since the beginning that user interaction can only be used to predict mental states by identifying patterns of each subject individually, whereas the study presented in this document tries to find a correlation between mental states and users' interaction, by "neglecting" the need to assess the data individually, despite some research found in literature that suggested that MWL depends on users' traits. In order to accomplish the goals purposed, Pimenta et al. (2013) asked 20 subjects to perform computer related tasks in two moments of the same day: the beginning of the day (where subjects were expected to be fully rested); and the end of the day (where subjects were expected to be more tired). Several indicators of user interaction were collected and a machine learning software layer was responsible of comparing the interaction behaviours between the two moments of the day, for each subject individually and, consequently, identifying patterns for detecting mental fatigue, for each subject in particular. The correlation between performance measures and the moment of the day seemed to corroborate this approach (in the assumption that fatigue is linked with performance) and later studies' results suggested that the algorithm computed by the machine learning layer was able to accurately predict mental fatigue for subjects, individually (Pimenta, Carneiro, Neves, & Novais, 2016).

In the context of MWL in HCI, a study that used tracked users' interaction was performed by Gwizdka (2009). This study used the dual-task method to measure the MWL imposed to the subjects while performing a set of fact-finding and information gatherings tasks. Subjects' mouse clicks and keyboard clicks were collected. However, this tracked activity was used to measure the reactions times related with secondary task's circumstances and not to detect patterns of interaction behaviour in the web task (for which MWL and performance were being studied), as it is purposed in the study reported in this document. Another study in the context of MWL and HCI was performed by Chen et al. (2012), which consisted in a five years long study that aimed to detect changes in MWL and user behaviour during system's interaction. This research used performance measures to quantify MWL and used eye-gaze tracking, gestures, and users' interaction logs (mouse position/clicking and keyboard usage) as behaviour measures. The use of mouse clicking and keyboard usage as behaviour measures was supported by studies that had linked user interaction with emotional state. However, in this study, user interaction is essentially used to understand users' actions in parallel with MWL, which is measured using eye tracking and speech patterns. This allows, for instance, understanding the action plan (top-down or bottom-up) of a user performing a fact-finding task or identity behaviour changes for a particular user (based on studies that linked behaviour changes with high workload demands). The study performed several experiments in real world scenarios (e.g. Emergency Communications Centre in North America; Contact Centre operator in Australia).

In summary, this section intended to present a brief overview of research that leveraged tracked user interaction as a potential tool that can bring major benefits in computing science studies. Essentially, the key notions extracted from the reviewed literature are the following:

- Tracking user interaction enables to study hundreds of users and interfaces, instead of being limited to a small sample obtained in a laboratory experiment;
- Potential benefits of using user interaction in the evaluation of HCI interfaces are: Reduced cost of Usability evaluation; Easily detect new user behaviours that were not taken in account in the design phase; Automatically predict and aggregate errors and their impacts; Increased coverage of system's features; Easy comparison between different designs;
- Tracked user interaction has proven to be a useful tool to analyse users' behaviour;
- Usability and MWL studies used tracked online information in order to automatically assess and aggregate response times, which leads to automation and reduces human error;
- Recent research used individual patterns of user interaction to predict mental fatigue;
- In the context of MWL and HCI, there are studies that used tracked user interaction in order to understand MWL peaks and changes.

2.4 Conclusion: Gaps and Limitations of Literature

2.4.1 Mental workload

Studies reported in this literature chapter suggest that the importance of measuring MWL in the field of HCI is well accepted. The benefits of MWL assessment are reported to be crucial and having an increasing importance due to the advent of more and more digital interfaces. Moreover, literature suggested that research performed in the combined fields of MWL and HCI are consistent in what is identified as the advantages and disadvantages of using the three types of MWL measurement methods. For instance, subjective measures are capable of assessing several dimensions but not real time MWL; physiological measures are costly to implement and do not scale well (Cegarra & Chevalier, 2008). To summarize, all studies that set experiments to assess MWL in the HCI field were "invasive" and required performing a controlled and planned experiment, with increasing costs when the number of subjects also increases.

In other words, the required effort and expenses in the collection and analysis of MWL' data is not flexible and simple. Cegarra and Chevalier (2008) suggested that combining the three type methods is the only current way of having a complete scope and notion of the MWL imposed by a given task in a given interface. In addition, new research has identified tracked user interaction (e.g. mouse moves) as a potential tool for enhancing the quality of methods used in the evaluation of web interfaces. Usability studies suggested that user interaction can be a mean to substitute typical Usability methods, which are hard to scale. This brings the idea of assessing the potential connection of users' interaction with MWL. If a correlation could be found, maybe the complexity of using all type of measures could be mitigated, since user interaction could, potentially, provide the same advantages of all methods combined: good scaling; reduced costs; real time workload; and taking in consideration several dimensions of MWL.

This idea was already purposed by Longo (2011), who suggested that it would be useful to study if online users' interaction data (e.g. mouse movement; scrolling) could be used to estimate the MWL needed to complete a web task. This idea is supported by the increase use of user interaction logs for evaluating Usability and understanding web users' behaviour (Agichtein et al., 2006; Atterer et al., 2006; Qi Guo & Agichtein, 2012). A recent study used machine learning to correlated a particular user's online activity data (i.e. individual behaviour) with his levels of mental fatigue (Pimenta et al., 2013). However, there seems to be lack of studies addressing if raw online activity data (i.e. without any information about the users) can be used to estimate the amount of MWL required in a web task. The study reported in this document aims to address this research gap. The assumptions taken by this project are supported by potentiality and plausibility. Regarding potentiality, Heger et al. (2010) accentuated the importance of using new methods for assessing MWL in the field of HCI, by suggesting that the possibility of assessing MWL in real time would allow systems to adapt and react to operator's mental state. Regarding plausibility, a study that suggested a possible connection between MWL and users' interaction was reported by Chen et al. (2012), that linked emotional state with mouse and keyboard usage, supporting the possibility of inferring MWL using indicators user interaction.

In conclusion, although MWL has been accepted in recent research as a key measure/indicator to take in consideration in HCI, its measurement is not trivial or scalable and could benefit from the potential that logs of user interaction can bring. These considerations led to the primary goal of this study, which aims to study if it is possible to correlate web users' interaction (e.g. scrolling; mouse clicks) with the MWL imposed by a web task.

2.4.2 Data Visualisation and Infographics

Regarding Data Visualisation, Saraiya et al. (2005) claimed that typical evaluation methods for interfaces (such as task accuracy, Usability "think aloud" protocol) are insufficient for evaluating visualisations representing huge and multi-dimensional sets of data, since they are based in a limited set of tasks, incapable of representing the Usability of exploring large and multi-dimensional datasets, which are typically used in the field. Plus, Saraiya, North, Lam, & Duca (2006) suggested that Visualisation studies are too much short-termed to be relevant. Taking in consideration that the methods used in those studies (performances ratios and surveys) are restricted by limitations in terms of number of tasks and users, there is a potential benefit of assessing the possibility of using Mental Workload as a new and reliable method for evaluating Visualisation interfaces. Especially if it is found a relation between users' interaction and MWL (primary goal of this project), which would allow to evaluate an interface using data provided by the interaction of hundreds of users. Equally significant, Scholtz (2006), who collected concerns of several studies aiming to create new ways of evaluating interfaces using Data Visualisation, highlighted two components that are the major focus of the study reported in this document. One is the necessity of assessing the cognitive complexity of visualisations (i.e. MWL). Other is the importance of tracking user interaction in order to identify patterns of user behaviour, allowing to extract knowledge form the users' interaction with the interface. These concerns are also supported by Tory and Möller (2005), who suggested that usual Usability tests, performed in the scope of Data Visualization, do not address high cognitive demanded tasks that are usually performed in real world scenarios. The authors claimed that expert reviews give a more reliable evaluation than laboratory user studies, but are insufficient to address and detect all Usability issues. This considerations support the notion of discovering new complementary methods for evaluating Visualisation interfaces, which is one of the goals of the study reported by this document. This conclusion is also supported by Bolchini, Finkelstein, Perrone, and Nagl (2009) that, in the field of Bioinformatics, evaluated the Usability of different interfaces, concluding that the results obtained couldn't be used with confidence since there is a limited number of significant samples (tasks and users) used in the performed experiments (due to their laboratory set-up), which strengthens the potential that using users' interaction to evaluate Visualisation interfaces could bring.

Additionally, the complexity of designing interfaces and Infographics in HCI, caused by the heterogeneity and huge amount of the typical high dimensional data used, leads to possible high cognitive demands in the exploration and perception of over-complex designed Infographics and Visualization representations of data and knowledge (Zinovyev, 2011).

In conclusion, the concerns reported by reviewed literature regarding the issues of developing a reliable methods for assessing Data Visualisation interfaces spawned a secondary goal for this study, which consists in assessing if MWL, that literature reported to be useful in the evaluation of typical HCI interfaces, can also be used to evaluate the more specific and singular interfaces that are used in the field of Data Visualisation.

2.5 Research Question

Since this study covers several fields and themes within the wide field of HCI, it is useful to review the train of thought, in line with the reviewed literature, that led to the main research question. This description of the "train of thought" also helps remembering the key ideas documented in recent research and described in the several sections of this literature review chapter:

- MWL, a Psychology concept, is linked with task performance;
- Research studying MWL in the field of HCI recommend assessing MWL as a measure for evaluating interfaces and measuring the cognitive demands imposed by HCI tasks;
- Research in the HCI field suggested that measuring MWL is complementary of Usability studies, allowing to assess and evaluate more aspects of HCI interfaces;
- The use of digital platforms and devices has been increasing in the past years. Therefore, the pace of needing to create new platforms or new futures for existing HCI platforms has been, continuously, more and more demanding and complex. Consequently, standard methods of assessing and ensuring the design quality of HCI interfaces have been challenged, due to short periods of development and the complexity of predicting users' interaction behaviour or how new data (daily generated) will be used;

- Recently, some studies (e.g. Usability studies; mental fatigue studies; web user's behaviour studies) used logs of user interaction (e.g. mouse clicks) in order to evaluate HCI interfaces. These logs allow obtaining data from hundreds of users in an inexpensive and easy to implement way, which can potentially allow quickly assessing the impacts of new interfaces or interfaces' changes. This contrasts with the limited capability, regarding scalability and cost, of standard design evaluation methods that are usually performed in laboratory (e.g. surveys; expert reviews; performance assessment for pre-defined tasks);
- Current state of art measurement methods of MWL, although having their well establish advantages and disadvantages, suffer from the same drawbacks of standard Usability measurement methods: imply laboratory setup and small samples of subjects; are hard to scale; among others. Moreover, the most easy to implement techniques (subjective measures) do not assess mental load in real time, which prevents from assessing MWL demand related with web task's steps or elements;
- Being acknowledged by research that MWL is a key aspect to take in account when designing, developing and evaluating HCI interfaces, leads to study and question the possibility of leveraging the potential of users' interaction logs in order to predict or measure the MWL imposed by a HCI task.

As a result and consequence of the aforesaid train of thought, the research question that this study purposes start answering is:

Can objective indicators of users' interaction be used to approximately measure the subjective Mental Workload of a web-based task?

Linked to this main research question, the following hypotheses and objectives were defined:

- **Null hypothesis (H0):** Objective indicators of users' interaction are linearly unrelated with the subjective Mental Workload required by a web task.
- **Alternative hypothesis (H1):** Objective indicators of users' interaction are linearly correlated with the subjective Mental Workload required by a web task
- **Objective:** Correlate variables using statistical methods after gathering data using Web experiments (with tracking tools) and surveys (after task self-report measurement).

Definitions of terms used in the hypotheses' description:

- **Objective indicators of users' interaction**: tangible online interaction activity of users in the browser (mouse movement, scrolling, clicking, focus).
- **Subjective Mental Workload**: collected using post task MWL self assessment methods.

Moreover, additional knowledge founded in literature led to the creation of a second objective for this study. The summary of the knowledge and suggestions, provided by recent research, which raised this secondary objective, is listed below:

- Within the field of Data Visualisation, recent research has questioned the suitability of standard Usability measurement methods for assessing the particular interfaces of Data Visualisation systems, claiming that Data Visualisation interfaces are, in nature, significantly different from typical HCI interfaces. For instance, a commercial web site has predefined tasks (e.g. purchase order) that are easy to set and define in laboratory. On the other hand, Visualisation interfaces are more complex in nature and handle heterogeneous data and tasks. Consequently, it is hard to predict, structure and simulate the tasks that will take place over those interfaces.
- Recent research suggested "Expert Opinion" or "Insights' Evaluation" as more reliable ways of evaluating Data Visualisation interfaces. However, these methods are still not sufficient for standalone evaluations and there is the need of finding and developing new approaches of evaluation;
- In the field of Bioinformatics, there are studies that highlighted the cognitive complexity demanded by Data Visualisation interfaces as a cause of the unsuitability of applying other evaluation methods that obtain reliable results in other types of HCI interfaces;
- The nature of the data used in this field, characterized by its heterogeneous and high volume, lead to the designing of over-complex Infographics, which impose high cognitive demands on users, in order to perceive the data and the knowledge visually represented;
- Acknowledging these literature considerations, which looked for new approaches for evaluating Visualisation interfaces, and linked high cognitive demands with Infographics' interaction, the idea of assessing if MWL would be a suitable approach for measuring this type of HCI interfaces arises. Especially

since another studies identified MWL as a valuable measure for assessing the design of other types of HCI interfaces.

The above considerations led to a secondary goal for this study, illustrated by the following research question:

Can subjective Mental Workload be used to evaluate the performance of interfaces using different Infographics representations?

The following hypotheses and objectives were set regarding this secondary goal:

- **Null hypothesis (H2):** Subjective Mental Workload cannot be linked to Infographics interfaces' performance.
- **Alternative hypothesis 2 (H3):** Subjective Mental Workload can be linked to Infographics interfaces' performance.
- **Objective:** Evaluate if the conclusions supported by literature, that linked high and low levels of Mental Workload with performance's degradation, are also verified in Data Visualisation/Infographics interfaces. Supporting recent research that indicated that it is useful to measure Mental Workload in the evaluation of digital interfaces.

Based on the gaps found in literature, and according with the purposes of the study, the next chapter documents the design and implementation of the online experiment that was performed in order to accomplish the study's goals.

3 DESIGN AND IMPLEMENTATION

This chapter describes the design and the implementation of the online experiment that was developed and launched in order to address the main objectives of this study. The primary objective intended to study the possible correlation between MWL and the interaction of web users (mouse clicks; mouse movement; etc). Consequently, the experiment used methods that allowed collecting these elements. Moreover, the secondary objective of this study was to study MWL as measure for evaluating web interfaces representing data with different types of interactive Infographics. Therefore, the web interfaces used in the experiment were defined by the use of visual representations of data (i.e. Infographics).

The chapter starts by describing the main components of the experiment and how they all fit together, followed by the main considerations taken when designing the experiment, which were grounded in the literature.

Thereafter, the measures and methods used to collect the essential data for accomplishing the objectives of this study are described and explained, highlighting their purpose and nature. Additionally, this chapter provides the formulas used to calculate and obtain, from the data collected, important indicators of MWL, Performance, and User Interaction. The coefficients' techniques used in order to verify the correlation between all types of indicators are also addressed.

Finally, the chapter summarizes the strengths and limitations of the solution taken, with respect to the objectives pursued.

3.1 Experiment Overview

In brief, in order to address the proposed research question, the student developed and performed an online experiment aiming to reach, at least, 80 participations. The experiment was set online rather than in a laboratory, which didn't limit the diversity of subjects. This allowed obtaining the most statistical relevant sample possible, in order to better support the conclusions and considerations related with the resultant data. The experiment lasted a month and subjects were gathered by sending an email with the experiment's link to several mailing lists (e.g. university), co-workers; family; and friends. Moreover, the link was shared in some scientific forums.

The experiment was designed in order to request each subject to complete a specific fact-finding task. The student defined two tasks to use in this experiment. Each task is related with a dataset and has two different interfaces that represent a different visualisation technique (Infographic). During the execution of each task, a piece of software, developed by the student, was responsible for collecting the subjects' online activity/interaction. Moreover, each subject was asked to complete a self-assessment questionnaire after completing each task.

The below list describes, in more detail, the online experiment's main elements:

- **Tasks**: Two tasks were defined. Both were fact-finding tasks. Since the key idea of this experiment was to collect a wide spectrum of different MWL demand levels, tasks were specified with multiple goals instead of one, in order to allow collecting different levels of Mental Workload.
- **Visual Encodings**: Different user interfaces were defined in order to complete the same task. Since literature identifies length, area and other attributes as playing a different range of significance in human perception (Cleveland & McGill, 1985), different Infographics were used to display the data for each task. There were two different visual representations of data for each task. Each Infographic/interface was randomly applied when a user started a task on the online experiment. Task 1 had an interface with ranking bar charts and other interface with a two-dimensional Map. Task 2 had an interface with time series charts and another interface displaying the data using tables. These different interfaces played the role of aiming to establish a varying range of levels of difficulty between tasks, in order to meet the goals of this study. Firstly, by aiming to collect a spectrum of different levels of Mental Workload among the subjects, which allowed correlating these different levels of Mental Workload with the subjects' online interaction data. Secondly, by allowing meeting the second goal of this study: use Mental Workload as an evaluation measure of web and interactive Infographics.
- **Subjective MWL collection**: In order to collect the MWL demanded/imposed, each subject was requested to answer a survey after finishing a task. The survey contained rating questions related with the NASA Task Load Index and Workload Profile methods, which are subjective MWL measurement methods.
- **User interaction tracking**: During each task the subjects' online activity was collected (e.g. mouse movement; scrolling). This was achieved by using the JavaScript programming language. The student developed a software package capable of recording the desired users' activity data.

The diagram below highlights the main steps of the designed online experiment. In other words, presents the main flow that was followed and took place in each participation:

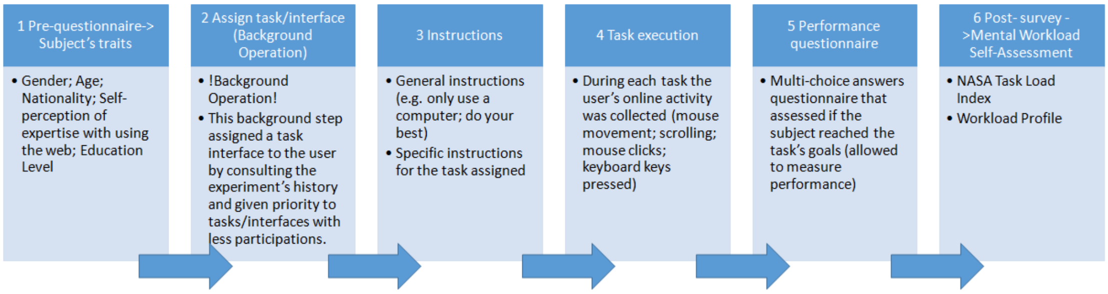


Fig. 3.1 – Online Experiment Flow and Steps

3.2 Experiment considerations

3.2.1 Data

Statistics related to country's data, due to its high dimensional features, are regularly used in projects and studies that aim to represent data/knowledge by using interactive visual representations or Infographics (Friendly, 2008), and can be used and represented taking in consideration several visual encoding characteristics such as line length and area, or different chart types (Few, 2004; Friendly, 2008). Consequently, statistics data for all around the world provided by World Data Bank [1] seemed to fit the needs of the online experiment performed in this project.

Since it was expected that the majority of subjects participating in the experiment would be European, data from the territories of the Euro zone were collected. Two major topics were selected, due to their relevance and interest (according to the student's opinion): Population Growth (8 attributes) and Unemployment (18 attributes). All the data, related to these topics, were collect from the World Data Bank website, containing 47 territories and 48 years (between 1967 and 2015). The complete list of attributes included in the dataset is presented in Appendix A, page 97.

Additionally, one of the experiment interfaces consisted in a two-dimensional map. The data (polygons) for building the two-dimensional map were obtained from Natural Earth[2], which provides free vector and raster data related with public domain maps.

[1] http://www.worldbank.org/
[2] www.naturalearthdata.com

Natural Earth’s contributors give permission to modify the content and structure of any downloaded data.

3.2.2 Task type

Athukorala, Gowacka, Jacucci, Oulasvirta, and Vreeken (2016) and Navarro-prieto, Scaife, and Rogers (1999) defined two types of task: fact-finding and explanatory. In brief, a fact-finding task is assumed to have precise search goals, whereas an explanatory task is assumed to “center around the acquisition of new knowledge” (Athukorala et al., 2016, p. 2635). Naturally, a certain task can be perceived as being a mix of both fact-finding and explanatory task types (Marchionini, 2006). Due to the online nature of the experiment and the anonymity of most participants, it seemed to be wiser to user fact-finding tasks in the experiment, since they are related with more precise goals and also allow better quantifying and computing performance (i.e. rate subjects’ answers, related with post survey questionnaires, more accurately).
Being established that the tasks would be, in their essence, fact-finding tasks, it was defined that each task would have two goals, later described in this chapter, designed to keep the subject occupied for a while, which would allow tracking enough user online activity (i.e. interaction data such as scrolling). Although the main nature of both tasks was related with look-up activities (fact-finding), there were elements of the defined goals that included actions related with learning (comparison) and investigation (exclusion/negation), which are characteristics more associated with exploratory tasks, as can be observed in Figure 3.2.

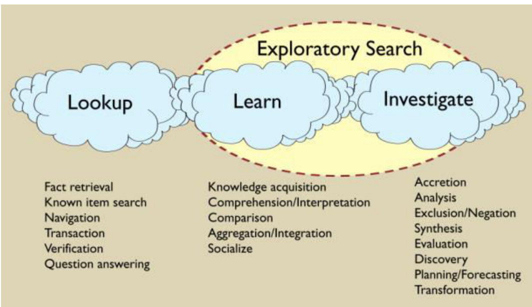


Fig. 3.2 – Search activities (Marchionini, 2006, p. 42)

The use of fact-finding tasks was also based on research found in literature, where a variety of studies concerned with the role of MWL in HCI reported using fact-finding tasks as a way of accomplishing their research goals (Gwizdka, 2009, 2010; Ward & Marsden, 2003). Moreover, the use of this type of tasks was also found in experiments that evaluated interfaces with different visual representation techniques (Kobsa, 2001).

3.2.3 Task complexity and duration

Since the experiment performed by this study consisted in an online website were anonymous subjects performed web tasks, it was really important that the designed tasks could engage and motivate subjects to complete each task until the end, especially because MWL was assessed via surveys at the end of each participation (and collecting MWL demands was crucial for accomplishing the project's goals). In order to not overwhelm subjects with a wide set of tasks, it was defined since the beginning that the experiment would consist in two tasks, which together would fit and allow reaching this project's purpose and main goals. Moreover, also due to the concern of having subjects fully completing each task, this study had the limitation of having to design tasks that are short (around 15 minutes) and friendly (avoiding that the user lefts the task uncompleted). This is was a major limitation especially because, at the same time, the study needed to track a wide range and variety of levels of MWL in order to address the main goal of this research study: assess if users' online activity can be correlated with the MWL imposed by a web task. Therefore, it was essential to create tasks that demanded a high workload so user interactivity could be associated with it. In order to accomplish this essential aspect of the study, but keep the tasks engaging and short as possible, the goals set for this experiment were defined in order to be more tedious and laborious, rather than complex or hard to achieve. Plus, each goal was composed by a set of two or three sub-goals. This approach of using multiple goals and having multiple activities to do was based on the idea that human brain has a limited capacity of processing more than five items of information (Mandler, 1967), supported by new findings that suggested that the capacity is even lower. For instance, Cowan (2010) suggested four items and Oberauer (2002) suggested one item. A MWL study in the context of single and multi-task HCI environments also corroborated this approach by suggesting that "multi-task performance normally requires more demands than single-task performance and may cause overload more easily" (Xie & Salvendy, 2000, p. 75). Hence, this approach of using sub-goals, that implied exploring the same values within the dataset, aimed to capture a wide range of different levels of MWL.

Nevertheless, this concern imposed a major limitation comparing with similar studies, which context encompasses laboratory or on-site experiments, allowing setting other types of tasks, as well as longer tasks. For instance, Gwizdka (2010) reported a MWL experiment in the field of HCI, which consisted in tasks that lasted more than an hour.

3.2.4 Visual encodings

Friendly (2008) suggested that displaying the same data with different graphic types (e.g. bar chart; time-series) leads to different assimilation and comprehension of the data, since humans perceive and rank information differently according with the visual encodings used to represent data, such as length and area (Cleveland & McGill, 1985). The study reported in this document based its premises on the aforesaid studies and created two interfaces for each task. The first interface of Task 1 is based on bar charts, whereas the alternative interface consists in a two-dimensional map with different colourized and sized circles. The first interface of Task 2 is based on time-series charts, whereas its second interface displays data using tables. The choice of each interface is explained later in this chapter, where task 1 and task 2 are described in more detail, as long as their interfaces.

3.2.5 Technology

Atterer, Wnuk, and Schmidt (2006) suggested that one of the major requirements for user online tracking is "Transparent Operation". In other words, the subject's browsing experience should not be affected or altered in any way, despite tracking actions are being executed. This was a critical requirement and consideration in the coding of the tracking JavaScript functionality, since performance could be affected by user interactivity tracking and, as a result, the subjects' browser experience would be affected (e.g. slower transition interacting within the different interfaces). The technologies used in order to develop the web application for this experiment are listed and described in page 97 (Appendix A).

3.3 Task 1

As previously mentioned, all tasks were composed of objectives typically associated with fact-finding tasks. The task 1's goals were the following:

- Please find out if the following sentence is true or false according to the data you are going to access: "Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's Death rate, crude (per 1,000 people) decreases, whereas Life expectancy at birth, total (years) increases". Plus, can you find in which turn of the decade (e.g. 1969 to 1970) Ireland had the biggest Death rate, crude (per 1,000 people) variation (absolute value of the difference between the Death rate in each year, for instance between 1969 and 1970)?

- Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's Death rate, crude (per 1,000 people) decreases, whereas Population growth (annual %) increases. According to the data you are going to access, which years contradict this theory?

As can be seen by these task's objectives, the data used for this task is related with Population Growth. The subject was able to select the year for which he wanted to visualize the data.

Two alternative interfaces were created for this task. The first being an ordered bar chart and the other being a two-dimensional map with circles centred in each territory. Both visual encodings, bar and circles, had their aspects (length and area respectively) linked with the percentage of the annual's population growth variation. A tooltip box, with extra information, would pop up when the user placed the mouse pointer hover a territory visual encoding (bar or circle). Since interface 1 ordered countries by the value of Population Growth for a respective year, users could not rely in previous positions, when navigating through different years, for locating a country. This increased the difficulty of this interface comparing with interface 2 (where each country was always placed on the same location, since it was a map). The intention was to increase the difficulty of performing this task with interface 1, since previous studies report that unorganized data can be related with higher cognitive demands (Ward & Marsden, 2003).

Figures 3.3 and 3.4 illustrate the visual aspects of each interface of Task 1.

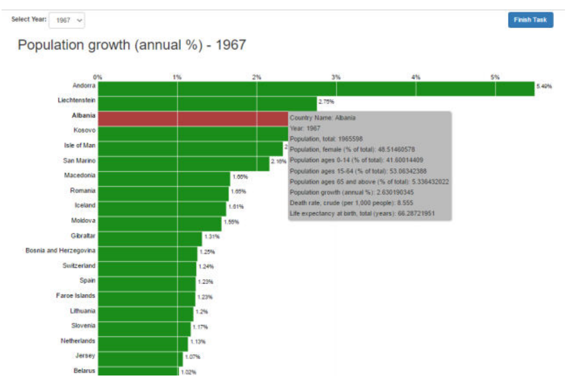


Fig. 3.3 Task 1 Interface 1 - Bar Chart

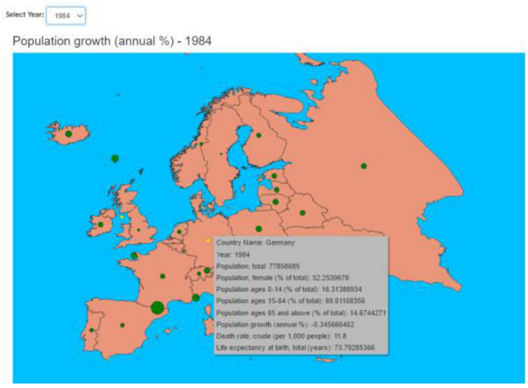


Fig. 3.4 Task 1 Interface 2 D Map with circles

3.4 Task 2

Task 2's dataset was related to Unemployment. Similarly with task 1, there are two possible interfaces for this task (only one interface was assigned when a subject started a task). Instead of organizing the data identically as it was done in task 1 (where the user would select a year to visualize each subset of the dataset) this task had data organized differently between interfaces. Ward and Marsden (2003) studied the effect of different web page designs in MWL and, in order to provide different interfaces to perform the study, the experiment grouped data differently. This approach was used in this task in the following way: Interface 1 (time series charts) allowed subjects to navigate through different subsets of data that were grouped by country, whereas Interface 2 (table) allowed subjects to navigate through different subsets grouped by year. A dropdown list was provided to subjects, enabling them to choose the dataset's subset to explore.

The goals for this task were the following:

- Knowing that Scandinavia contains the following 5 territories: Denmark; Norway; Sweden; Finland; and Iceland. Please find out which of these territories had a larger Long-term unemployment (% of total unemployment) variation between 2012 and 2013 and how much was that variation (absolute value).
- Consider the following Southern Europe countries: Portugal; Spain; Italy; and Greece. Which one had the biggest Unemployment, female (% of female labour force) in 2014? For that country, the Unemployment, youth total (% of total labour force ages 15-24) increased or decreased between 2012 and 2013?

The first interface for this task used time-series charts (figure 3.5), which are commonly used in Visualisation studies as, for instance, in the study performed by Purvi Saraiya, North, and Duca (2005), where different Visualisation techniques were tested in the field of Bioinformatics. The time-series chart linked years with the total unemployment of each country. If the user placed the mouse pointer above a year's occurrence, extra unemployment related information would appear (inside a pop-up box).

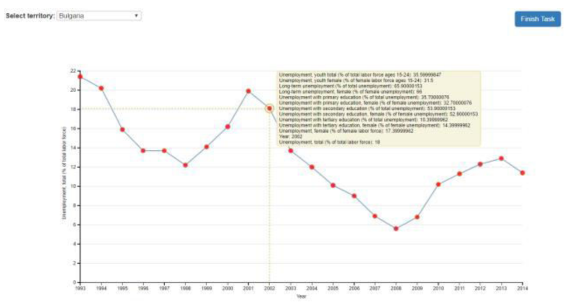


Fig. 3.5 Task 2 Interface 1 – Time Series Chart

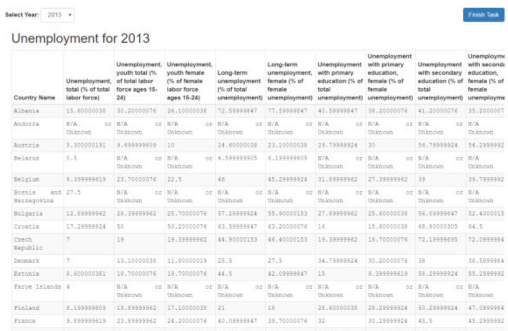


Fig. 3.6 Task 2 Interface 2 – Ill-Designed Table

The second interface of this task (figure 3.6) was based on Kobsa (2001), who reported an empirical comparison between interfaces that use different Data Visualisation paradigms. One of the interfaces was a common approach for representing data: a “table-like” with rows being the objects and columns the dimensions (i.e., the attributes of objects), and the other alternative interface had data represented with visual techniques (i.e. Infographics). Moreover this second interface was based in the “ill-design” approach followed by Ward and Marsden (2003), that performed an web design experiment that aimed to use Mental Workload as a Usability measure for different web page designs. Likewise, this experiment takes advantage of this “ill-designed” concept, that was also used in research related to technology within the context of education (Hartley, 1994, 2004). Consequently, in interface 2 the tables representing the dataset were “badly” designed on purpose, by not showing all columns (within the browser window’s width); not fixing the top line row (that contains the column names); and by requiring subjects to use a hard to manipulate side scrolling bar in order to explore the dataset. The intention of this second interface was to impose higher workload levels on subjects, in order to collect significant and diverse levels of Mental Workload.

3.5 Measures and Indicators

3.5.1 Users’ Interaction

3.5.1.1 Elements and considerations

The primary goal of this study is to assess if user online activity can be correlated with the self-perceived MWL needed in order to complete a web task. Thus, user activity/interaction measures and indicators assume a key and crucial role in the study.

Atterer et al. (2006) stated that every measure should be a time-based measure, in the sense that any action (e.g. scrolling; mouse click) should be associated with a timestamp value in order to provide meaning and context during the post experiment results’ analysis. Hence, every online activity measure collected in this experiment was associated with a timestamp value (relative to the time when the user starts the experiment). The below list identifies the user interaction elements that were tracked during the experiment:

- **Mouse position** – This interaction element was mentioned in literature as being extremely insightful for web designers and developers (Arroyo, Selker, & Wei, 2006; Atterer et al., 2006), especially because it allows to understand if the user is focusing on the correct/appropriate elements of the webpage, which can reflect the Usability of the web interface (Arroyo et al., 2006). Moreover, studies reported that mouse position can be linked to eye focus within a 84% to 88% correlation (M. Chen, 2001), which is supported by other studies that found similarities between mouse movements and the users' eye movement, regarding rhythms and focus (Cooke, 2006). In this experiment, while a subject was performing a task, the x and y positions (in pixies) of the mouse pointer were collected, at least every second.

- **Mouse clicks** – Atterer et. al. (2006) mentioned that, in the context of web tracking, "the interaction should be tracked at the widget level, i.e. the mouse coordinates are mapped to elements like buttons, links etc." (Atterer et. al., 2006, p. 203). With this in mind, mouse clicks events were tracked not only by recording mouse coordinates and respective timestamp, but also by collecting the HTML/SVG element where the user clicked. This extra info is expected to provide additional insight or meaning, for this project but also for further studies (that can analyse the dataset collected by this study's experiment).
- **Scrolling** – Another online activity action tracked. Besides collecting the timestamp of this event, the distance to the top of the page (in pixies) was also collected.
- **Keyboard activity** – Although keyboard was never needed for completing any of the experiment's tasks, it was decided to track these events because the use of keyboard could be somehow relevant to this Mental Workload study (e.g. could be related with reactions of frustration). Moreover, tracking this type of iteration could also help identifying subjects using browser's short keys (e.g. looking for a word in the web page's content).

The next section explains which indicators were computed and aggregated related to the tracked users' interaction data.

3.5.1.2 Indicators of User Interaction

The aforesaid data tracked related with the users' actions (Mouse Cursor Position; Mouse Clicks; Keyboard Use; Scrolling) allowed calculating several indicators of User Interaction. The majority related with subjects' activity levels (i.e. rhythm or pace) and based on the Usability study performed by Guo and Agichtein (2012), which focused on analysing a dataset containing user interaction logs related with the handling of web searches, with the goal of assessing if the web pages' listings resultant from the user's search activity are relevant or not for the task's goals.

<u>Mouse Clicks Indicators</u>

Arroyo et al. (2006) mentioned studies that found out that users often use the mouse pointer as a helper tool for reading text in web sites (e.g. clicking in order to highlight text that they are reading). This is an interesting behaviour and it was decided that using an indicator that counts the number of void clicks would be interesting. "Void clicks" are clicks that are not associated with a button or a link (i.e. the user is not navigating through the web page or its components by using that clicking action). Plus, it was expected that this measure can also be linked with reactions like frustration or focus. As a result, the following indicators were computed based in the logs containing mouse clicks occurrences:

- Number of Clicks
- Number of Clicks per Minute
- Number of void clicks
- Number of void clicks per minute
- Average Time between Clicks (seconds)
- Average Time between Void Clicks (seconds)
- First Click (seconds)
- First Void Click (seconds)

<u>Keyboard Use Indicators</u>

The following indicators were computed by using the experiment's logs containing the tracked keyboard use actions:

- Number of Keys Pressed
- Number of Keys Pressed Per Minute
- First Time Keyboard Was Used (seconds)
- Average Time between Keyboard Use (seconds)

<u>Scrolling Indicators</u>

The following indicators were computed based in the logs storing scrolling activity:

- Number of Scrolling Events - Note: if three or more seconds separate scrolling occurrences it is considered that there are two distinct scrolling events.
- Number of Scrolling Events Per Minute
- First Time Scrolling Was Used (seconds)
- Average Duration of Scrolling (seconds)
- Average Time between Scroll Events (seconds)

Mouse Pointer Position Indicators

The indicators calculated from mouse pointer position logs are considered belonging to two categories: activity and attention/focus.

Regarding activity, which is the main category from the other users' interaction indicators (mouse clicks; scrolling; keys pressed), the mouse position indicators used in this study are based on research documented in Pimenta et al. (2013), which studied mental fatigue and identified mouse velocity (i.e. velocity at which the cursor travels) as interaction indicators linked with mental state. The activity indicators were the following:

- Distance travelled with mouse (pixels)
- Distance travelled with mouse per Minute (pixels)

Regarding the attention/focus category, the indicators created are linked with several studies. Arroyo et al. (2006) mentioned studies that suggested that eye tracking is a useful tool for assessing user's attention and studies that linked mouse pointer position with eye focus. Those studies indentified tracking of mouse cursor's position as a cheaper and effortless, yet still reliable, approach of assessing users' visual focus of attention as an alternative to eye tracking methods, which are hard and expensive to implement. Moreover, the study also mentioned research that suggested that users use the mouse cursor has a reading aid, which is supported by parallel studies. For instance, Atterer et al. (2006, p. 211), a study about web usability, reported that "during the user study, it also became clear that the mouse pointer is often used as a reading aid when scanning through menus on the web page". Moreover, a study where eye tracking was used to assess Mental Workload during web tasks reported the following:

- "The most common type of eye tracking event refers to a focused state when the eye remains still over a period of time. This event is called fixation and lasts from 200-300 milliseconds to up to several seconds. It is a voluntary movement.

The number of fixations indicates the number of times that a user looked to a certain area of interest“ (Zagermann, Pfeil, & Reiterer, 2016, p. 79).

Based on the aforesaid statement and in the already mentioned research that linked eye focus with mouse cursor’s position, the study reported in this document created several indicators, listed below, that are related with mouse position and are expected to be linked with the subjects’ (visual) focus and attention. The term “fixation” is used as it in purposed by Zagermann et al. (2016) and is related to the occurrence of periods of time where the mouse cursor, interpreted as mimicking the eye attention, remains still from 1 to several seconds.

- Average time in the same area (seconds) – Note: In order to calculate this indicator, the web browser’s window was divided in 16 rectangles (by diving the browser window’s width and height for equidistant lines). Then, it was calculated during how much time the mouse cursor was placed at each rectangle (e.g. area of focus of attention) before the user moved the cursor to another rectangle/area. Figure 3.7 exemplifies how the browser’s window was divided. Note that this indicator is also based in some suggestions made by Zagermann et al. (2016), who interpret quick or frequent shifts between locations (eye focus) as being probably correlated with high levels of workload.
- Number of fixations
- Number of fixations per minute (seconds)
- Average Fixation Time (seconds)

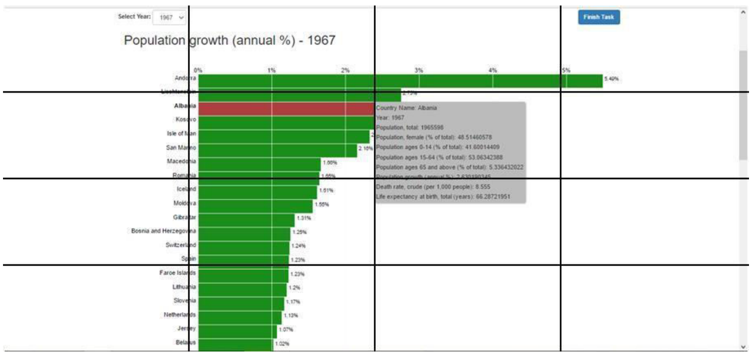


Fig. 3.7 – Illustration of how a subject’s window was divided in order to determine rectangles/areas of focus/attention

3.5.2 Mental Workload

3.5.2.1 Chosen Methods

MWL was collected by using post survey questionnaires (i.e. subjective methods). Namely: NASA Task-Load Index and Workload Profile. The list of questions used in the post questionnaire can be found in Appendix A, page 96. Each subject rated each method's dimension/question according to a provided ranking (from 0 to 100, with "multiples of five" ticks).

This type of methods (subjective methods) was chosen because the experiment was performed online and, therefore, dual task and physiological methods could not be used since they imply performing a controlled experiment in a laboratory. Moreover, dual task and physiological methods would be intrusive and could interfere with the user interaction patterns. In addition, it was decided to use multidimensional subjective methods, since different mental dimensions can be related with different indicators of user interaction. The three most used and tested multi-dimensional subjective methods, Workload Profile, NASA-TLX and SWAT, were identified by previous studies as being robust and reliable (Rubio, Díaz, Martín, & Puente, 2004). Since comparisons between these methods suggest that all are reliable – with Workload Profile having an outstanding sensitivity for task's manipulations (Rubio et al., 2004) - Workload Profile and NASA-TLX were selected as the methods to use in this study. SWAT was left out because it would imply a large pre-questionnaire (comparing levels and dimensions), which would complicate the experiment and discouraged anonymous subjects to participate.

Moreover, since Workload Profile and NASA-TLX use a somehow disparate assess of mental dimensions, it was perceived that it would be useful to use both since together they would allow to consider several cognitive aspects for evaluating possible correlations with indicators of User Interaction. Moreover, both Workload Profile and NASA-TLX methods are reported to have high acceptability from subjects, as well as ease of use and implement (Longo et al., 2012), which are ideal features for inclusion in an online experiment with anonymous users.

3.5.2.2 Calculation of Mental Workload Indicators

The original NASA-TLX calculation formula included the use of different weights between dimensions. The weights' values would be obtained by asking subjects, before completing a task, to compare all possible pairs of dimensions by stating which dimension imposed a bigger demand. However, recent studies eliminated this weighting process and estimated the overall workload rating by simply averaging the ratings obtained for each dimension. This method has been referred to as Raw TLX (Hart, 2006). The same approach was used in this study since, in order to minimize the extent of the experiment's post-questionnaires, subjects were not asked to compare dimensions (only rate each one). Moreover, since the dimension "Physical Demand" was not expected to play a role in subjects' participation (since tasks only required mental and cognitive skills) the ratings obtained for that dimension were not included in the formula used in this project. Naturally, some subjects, due to some physical constraints could have some limitations using the mouse for instance and, consequently, report high levels of physical demand. However, that was note the case according with the obtained data.

Consequently, the indicator of MWL for the NASA-TLX dimensions was the following:

$$NASA\ TLX_{HMW} = \frac{\sum_{x_1} NTQ_x}{5} = \frac{NTQ_{MD} + NTQ_{TD} + NTQ_E + NTQ_P + NTQ_F}{5}$$

where:
$NTQ_x: [0 \ldots 100] \in \aleph$
$x_1: \{MD|TD|E|P|F\}$

MD – Mental Demand
TD – Temporal Demand
E – Effort
P – Performance
F - Frustration

The standard formula for calculating the Workload Profile indicator of MWL is merely the sum of the ratings of all dimensions (Longo et al., 2012). Since it was not found in literature a redefinition of the formula for handling dimensions that are not related with the task's characteristics, this study used the standard formula, which is the following:

$$WP_{HMW} = \frac{\sum_{i=1}^{8}(WP_i)}{100} = \frac{WP_{SD} + WP_{TS} + WP_{VM} + WP_{AA} + WP_{SR} + WP_{RS} + WP_{VA} + WP_{MA}}{100}$$

where:
$WP_{HMW}: [0 \ldots 8] \in \mathbb{R}$
$WP_x: [0 \ldots 100] \in \mathbb{R}$
$i: \{SD|TS|VM|AA|SR|RS|VA|MA\}$

SD – Solving and deciding
TS – Task and space
VM – Verbal material
AA – Auditory attention
SR – Speech response
RS – Response selection
VA – Visual attention
MA – Manual activity

3.5.3 Performance

Assessing MWL as a measure for evaluating web interfaces that use different Visualisation techniques was the second goal of this study. Since previous literature linked high and low demands of MWL with performance's degradation, some performance measures were considered in this experiment. These performance measures allowed evaluating the relationship between performance and MWL, and understand if the same aforesaid linkage mentioned in literature was also observed in this study, which used web interfaces with interactive Infographics. Moreover, indicators of task's performance are often used as indicators of Usability in Data Visualisation studies (Cawthon & Moere, 2007; Freitas et al., 2002; Kobsa, 2001; Koua, Maceachren, &

Kraak, 2006; Saraiya et al., 2005, 2006; Scholtz, 2006) and as indicators of MWL (Gwizdka, 2010; Hart & Staveland, 1988; Tsang & Vidulich, 2006; Xie & Salvendy, 2000), although not entirely reliable in the case of MWL (when speaking of primary task measurement methods). Moreover, due to its documented relation with performance, there are studies studying MWL in the Web Design field that consider performance measures in their experiments (Hart & Staveland, 1988; Ward & Marsden, 2003; Wästlund et al., 2008).

In order to track performance, subjects answered to a multiple choice questionnaire after finishing a task. There was a question for each one of the two task's goals, and each question had five possible answers: one 100% correct answer; one partially correct answer; two wrong answers; and a "couldn't find out" answer/option. The questionnaire for each task is stated in Appendix A, page 96.

Based in the aforesaid studies the following indicators of performance, as well as respective formulas, are described below:

Task accuracy (%)

Percentage of correct answers for each task/interface combination results. Related with effectiveness. Calculated with the following formula:

$$Task\ accuracy\ (\%) = (p_1 + p_2) \times 100$$

where:

$p_1, p_2: \{0|0.25|0.5\}$

Calculation of points:
0.5 points per 100% correct answer
0.25 points per 50% correct answer
0 points per incorrect answer

p_1 is the points obtained on question 1
p_2 is the points obtained on question 2

Correct response time (minutes)

Average time needed to reach a 100% correct answer. Related with Efficiency. Calculated by using the following formula:

$$Avg\ Correct\ response\ time = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{t_1 + t_2 + \cdots + t_n}{n}$$

where:
$n: [1, \infty[\ \in \mathbb{R}, t: [0, \infty[\ \in \mathbb{R}$

n is total of participations
t belongs to the set of elapsed times of all participations that scored 100% task accuracy

Rate of Abandonment (%)

Considers participations that did not fully complete a task. A task is considered completed only when subjects fill all the post questionnaires. Strongly linked with interface's usability. Calculated with the following formula:

$$Rate\ of\ Abandonment\ (\%) = \frac{T_{NC}}{T_{NC} + T_C} \times 100$$

where:
$T: [0, \infty[\ \in \mathbb{R}$

T is the total of participations (NC – Not Completed and C –Completed)

<u>Average Time of Abandonment (minutes)</u>

Considers all participations where subjects didn't fully completed a task. Strongly linked with interface's usability. Calculate by the given formula:

$$Avg\ Time\ of\ Abandonment = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{t_1 + t_2 + \cdots + t_n}{n}$$

where:
$n: [1, \infty[\ \in \mathbb{R}, t: [0, \infty[\ \in \mathbb{R}$
$t = T_{Abandonment} - T_{Start\ Time}$
$T: [0, \infty[\ \in \mathbb{R}$

n is total of participations where users abandoned the task before completed
t is the elapsed time in minutes from the starting time until abandon the task

<u>Task completion time (minutes)</u>

Only considers participations that fully completed the experiment. Calculated by the following formula:

$$Avg\ Task\ Completion\ Time = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{t_1 + t_2 + \cdots + t_n}{n}$$

where:
$n: [1, \infty[\ \in \mathbb{R}, t: [0, \infty[\ \in \mathbb{R}$

n is total of participations where users completed the task
t belongs to the set of elapsed times of all participations completed

Equally important, self performance assessment (i.e. subject self rating themselves regarding performance) is listed and used as an important indicator of performance (Wästlund et al., 2008). This measure is already contemplated within the post-survey Mental Workload self-assessment method, NASA-TLX, described previously. In addition, stress and fatigue, also self assessed with NASA-TLX, were also identified as performance factors in previous studies (Cao et al., 2009).

3.5.4 User Traits

Although the aim of this study was to assess if user online activity by it-self (i.e. without knowing nothing more about the users performing the tasks), can be used to assess the Mental Workload needed to complete a web task, previous studies indicated that personal traits can strongly influence the amount of MWL demanded during task's execution (Gwizdka, 2010; Hart & Staveland, 1988; Xie & Salvendy, 2000). Moreover, research identified age, gender and education level as factors linked with the perception of visual elements (e.g. colour) of Infographics (Harrison et al., 2015). Therefore, since the beginning of this study, it was understood that, the correlation between MWL and User Interaction could possibly be dependent of the subjects' personal traits. In order to handle this possibility, before starting a task, the subjects were asked to give information related with following personal traits: Gender; Age; Nationality; Self evaluation of expertise as web users; Education Level.

This knowledge about the subjects allowed creating subjects' profiles with the purpose of aggregating indicators in the light of the aforesaid considerations mentioned by literature. Subjects' profiles were created using the following criteria: a profile is a combination of values of two personal attributes (e.g. Education Level = "Bachelor" / Nationality = "Irish") found at least in 20 participations, and at least in 5 participations per interface.

3.6 Correlations between Indicators

In order to assess the level of correlation between indicators of MWL and indicators of User Interaction the following correlation coefficients were used:

- Pearson coefficient
- Spearman coefficient

Both coefficients measure the extent to which two variables tend to change together. The calculated coefficient indicates the direction of relationship and the strength of the relationship. If the coefficient is negative it means that one variable increases while the other decreases. If the coefficient is negative it means that one variable increases while the other also increases. The strength of the relationship is given by the value of the calculated coefficient, which varies between -1 and 1. If the absolute value of the coefficient is bigger than 0.7 it means that a strong correlation was found between variables. If it is bigger than 0.5 the variables can be said to be moderately correlated. Smaller values then 0.5 indicate weak or inexistent correlations.

The Pearson coefficient evaluates the linear relationship between variables (i.e. a change in one variable is linked with a proportional change in the other variable), whereas the Spearman coefficient evaluates the monotonic relationship between variables (i.e. a change in one variable is linked with a change in the other variable, but not necessarily at a proportional or constant rate). The studied defined that only one of these coefficients would be used to correlate indicators of Mental Workload and indicators of User Interaction. The criteria for selecting the coefficient were the following: some preliminary comparison between some indicators was performed using scatter plots. If, despite the type of correlation, there are significant outliers, the

Spearman coefficient should be used (due to Pearson coefficient's high and misleading sensitivity to outliers). In the opposite case, the Pearson coefficient should be used.

Note: When an indicator of User Interaction didn't have any value for all the participations, then those participations with missing values were not considered in the coefficient's calculation. For instance, not all participations used the keyboard and coefficient's calculation involving the indicator "First Time Keyboard Was Used" only included participations with values for that indicator. The indicators that can have inexistent values are: Average Time between Void Clicks; First Void Click; First Time Keyboard Was Used; Average Time between Keyboard Use; First Time Scrolling Was Used; Average Duration of Scrolling; Average Time between Scroll Events;

The selected correlation coefficient formula was applied between:

- All indicators of User Interaction and the aggregate values calculated for the NASA-TLX method;
- All indicators of User Interaction and the aggregate values calculated for the Workload Profile method;
- All indicators of User Interaction and individual values of each dimension of the NASA-TLX method;
- All indicators of User Interaction and individual values of each dimension of the Workload Profile method.

Moreover, these correlations were applied within the scope of the following datasets:

- The entire data set;
- Each interface's dataset;
- Each subject profile's dataset.

3.7 Conclusion: Strengths and Limitations of the Designed Solution

This chapter covered the design of the online experiment set to accomplish the main goal of this study: correlate MWL with user's web interaction. Two multiple goals tasks were designed, each one having two distinguish interfaces. The online activity data collected were: mouse position; mouse clicks; scrolling; and keyboard use. The methods for measuring MWL, both subjective, were NASA-TLX and Workload Profile. Furthermore, subjects' traits such as age were also collected. Appendix "Online experiment" shows extra print screens of the online experiment, as well as the instructions and the consent form; among others.

The design options taken in this project entailed choosing conflicting strengths and limitations. On one hand the designed solution has the following strengths:

- The experiment was performed online, which presumably allowed reaching a wider audience. But, more significantly, a diverse audience. This theoretical would allow obtaining a more significant sample, statistical speaking, regarding the strength of the collected dataset. Especially because the aim of this project is related with general web user interaction;
- The chosen methods for measuring MWL were identified, by literature, as being reliable and having good subjects' acceptance;
- The two chosen methods for measuring MWL consider different mental dimensions between them, bringing the total of 14 dimensions, which can be important in the study of a possible correlation between aspects of MWL and different types and indicators of User Interaction;
- The chosen indicators of User Interaction and Performance were strongly supported by literature;
- A significant number of indicators of User Interaction was defined, bringing the total of 23 indicators, allowing to correlate those indicators with the 14 dimensions of MWL collected;
- The alternative interfaces for each task used interactive Infographics supported by literature, which gave strength to the secondary goal of this project: assessing the reliability of MWL in the evaluation of interactive Infographics;
- The collected user interaction was related to the majority of the possibilities of interaction with web sites using a desktop or laptop;
- The tasks were designed in order to take approvability 15 minutes to complete, which allowed tracking a significant amount of subjects' interaction and actions;
- The JavaScript code responsible of tracking user interaction did not affect the execution of the tasks;
- Subjects' traits were obtained, which could be useful to understand the experiment's results, in the light of studies that identified users' traits as factors related with different cognitive demands and ways of processing visual representations of data;

- The users were instructed to only use a desktop or a laptop to perform the task, which limits the number of factors that could interfere with the experiments' results.

On the other hand, the following limitations are inherent to the designed solution:

- Since the experiment was performed online it was not possible to fully control everything (e.g. know if subjects had breaks or were distracted while executing a task; identify cheating; standardize the hardware and devices used). These concerns are supported by literature that identified hardware used (e.g. screen size; screen resolution) (Wästlund et al., 2008); room temperature (Ward & Marsden, 2003); or subjects' mental rest (Pimenta et al., 2013) as factors that influence mental demands;
- In order to avoid discouraging subjects and facing high rates of abandonment, the tasks could not be too much difficult or long, which could limit the range and type of the MWL levels collected. In order to mitigate this, the tasks were designed as being laborious and composed of several goals. However, this could possible also had entailed that the same MWL dimensions were imposed (e.g. frustration due to the laboriousness of the task);
- The methods of measuring MWL were subjective. In other words, the values obtained were dependant on the subjective perception that subjects had on scales and dimensions;
- The methods of measuring MWL didn't measure MWL in real time, which could be useful when studying the correlation between interaction patterns and mental demands;
- The coefficients used for correlating indicators only assess relations were variables change together in a positive or negative direction (i.e. other possible relations were not assessed, such as quadratic relation). Moreover, the coefficients used only assessed two variables, living out possible correlations between one variable and the combination of other two or more variables;
- Pimenta et al. (2013) reported studies that suggested that people often hide truths or lie when answering questionnaires, besides being reluctant and afraid of giving correct answers. These considerations question the reliability of the MWL levels collected in this experiment.

The next chapter will describe the results of the performed experiment, calculated using the formulas and correlations approaches described in this chapter.

4 RESULTS AND EVALUATION

This chapter shows the results obtained with the performed online experiment. The results are discussed in the light of what previous research had found and in line with the hypotheses and research questions previously set for this study. The chapter starts by showing the data related with participations: MWL and Performance. Thereafter, the results of the correlations related with the primary study's objective - correlate Mental Workload and User Interaction - are discussed. Finally, the results related with the secondary objective - use Mental Workload to evaluate alterative Infographics - are addressed. The final section discusses the strength and limitations of the findings.

4.1 Participations

4.1.1 Criteria for Participation Exclusion

Since the experiment was performed online it was not possible to control all aspects of attendance and, consequently, several participations proved to be not useful for the purposes of this study. Consequently, some data records were completely removed from the final dataset according with the following criteria:

- Participations that rated every dimension with the same value; or clicked through all steps of the experiment just to know how the experiment is (e.g. doing the task in less than 2 minutes); or filled post-surveys nonsensical (for instance; indicating that the task demanded total attention regarding auditory attention);
- Participations that only proceeded until the instructions' screen. In this study task abandonment is only considered if users give up after starting the task. In other words, a participation is only valid it the subject at least started a task;
- Participations with 100% correct answers that performed the tasks in less than 2 minutes or didn't consult enough data to get the answers right, which indicated cheating.

4.1.2 Distribution

Table 4.1 presents the nature of participations regarding subjects' traits. The table also shows the total number of participations (145) and the total number of complete participations (95). Complete participations refer to participations that completed the task as well as all surveys. Since surveys related to Mental Workload self assessment were the last step of the experiment, it was only possible to obtain Mental Workload data for complete participations.

Table 4-1 Experiment's Participations

Participations	**Total**	**145**	**Nationality**	Portuguese	37.20%
	T1/I1	33		American	22.10%
	T1/I2	35		Irish	7.60%
	T2/I1	39		Others (27)	33.10%
	T2/I2	38	**Gender**	Male	64.80%
Complete Participations (task and post surveys completed)	**Total**	**95**		Female	35.20%
	T1/I1	20	**Web Expertise**	Amateur	3.40%
	T1/I2	29		Average	34.50%
	T2/I1	30		Expert	62.10%
	T2/I2	16	**Education**	Primary	2.10%
Age	Between 18 And 25	27.60%		Secondary	9.70%
	Between 26 And 35	46.20%		Bachelor	36.60%
	Between 36 And 50	15.20%		Master	33.80%
	Bigger Than 50	11.00%		Doctoral	17.90%

4.1.3 Subjects' Profiles

The following subjects' profiles were identified from the dataset containing all participations that fully completed the experiment:

- Profile A (32 participations) – Age between 26 and 35 / Expert web user;
- Profile B (24 participations) – Age between 26 and 35/ Portuguese nationality;
- Profile C (22 participations) – Master Education / Male;
- Profile D (40 participations) – Male / Expert web user;
- Profile E (33 participations) – Age between 26 and 35 / Male.

Each profile is verified in at least 20 participations, and at least 5 participations per task/interface. Two different profiles can contain a subset of the same participations.

4.2 Mental Workload Indicators

The outcomes of each Mental Workload measurement method are presented in figures 4.1 (NASA-TLX) and figure 4.2 (Workload Profile).

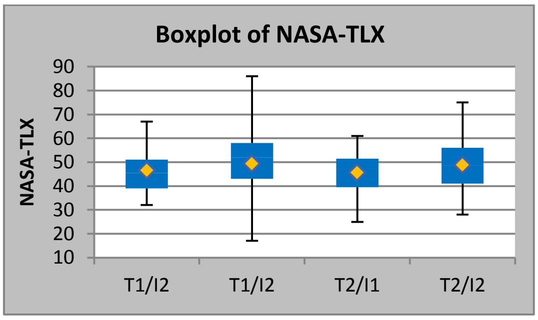


Fig. 4.1 Boxplot of NASA-TLX grouped by interface (scale of 0 to 100).

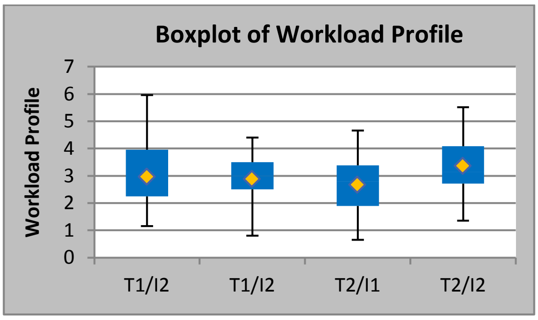


Fig. 4.2 Boxplot of Workload Profile grouped by interface (scale of 0 to 8).

These figures show that the designed experiment was capable of imposing a variety of Mental Workload levels, despite previous mentioned design concerns and restrictions about task difficulty and complexity. Interestingly, Workload Profile (figure 4.2) shows that most of the results were placed above the scale's midpoint, whereas NASA-TLX (figure 4.1) doesn't. This probably is related with the use of Raw TLX in the calculation of NASA-TLX indicators, which allowed removing dimensions that were not relevant for the experiment (physical demand), whereas in the calculation of Workload Profile indicators of all dimensions were considered (even though dimensions such as auditory attention and speech response didn't play a relevant role). Moreover, the expected variance between values in interfaces of task 2 was verified, since the ill-designed Interface 2 (T2/I2) shows bigger levels of mental workload for both methods. Interestingly, expected variance between values in interfaces of task 1 (T1I1 and T1I2) differs between each MWL measurement method. Interface 2 (T1I2) was expected to be easier since the data to look for was always on the same page's location, which is only verified for Workload Profile indicators (figure 4.2). The unexpected values presented for NASA-TLX (figure 4.1) can, possibly, be related with visual length encodings (T1I1) which had been reported, in previous literature, as being better perceived by human's brain than visual area encodings (T1I2). Nevertheless, these results support the choice of using these two methods, by suggesting that they assess different aspects and elements of the imposed MWL.

4.3 Performance Indicators

Figure 4.3 shows the results calculated for task accuracy between interfaces. Again, it is interesting to realize that is not verified the assumption that task 1 / interface 2 (T1I2) would be easier than interface 1 (T1I1). This can suggest that, despite the disorder imposed in the bar charts of interface 1, bar charts are more proper for the task in question than maps (interface 2). However, rates of abandonment (presented in figure 4.5) show the expected bigger rates of abandonment for the foreseen more difficult interfaces in each task: disordered information in interface 1 of task 1 (T1I1) and "ill-designed" tables of interface 2 of task 2 (T2I2). Interestingly, Correct Response Time (figure 4.4) shows better scores for the ill-designed version of task 2 (T2I2), compared with the expected better version (T2I1 – time series chart). In addition, the score obtained for Task Accuracy (figure 4.3) between interfaces of task 2 are very similar. Possibly, these results are related with the subjects being probably more used to access information in tables than in time-series charts, which is also suggested by the Visualisation study performed by Kobsa (2001). Furthermore, some studies identify tables as superior representations of information for tasks that consist in the comparison of few variables (Few, 2012), as it was asked by the experiment's tasks.

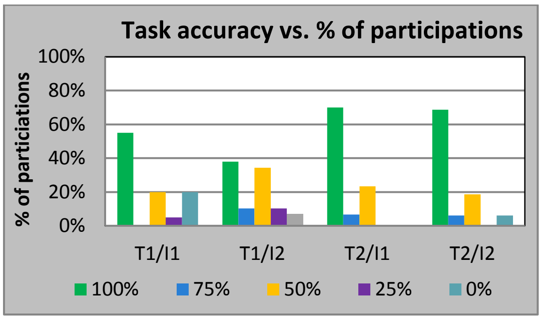


Fig. 4.3 Task accuracy vs. Percentage of participations

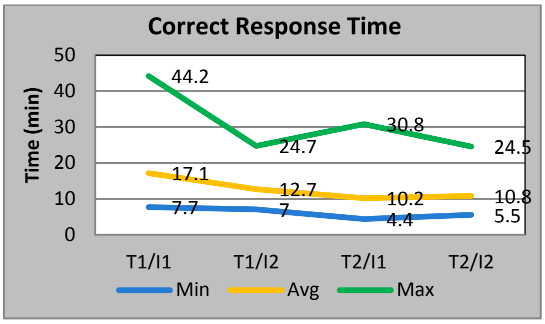


Fig. 4.4 Correct Response Time

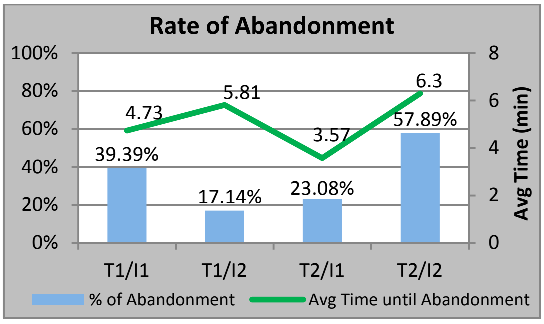


Fig. 4.5 Rate and Average Time of Abandonment

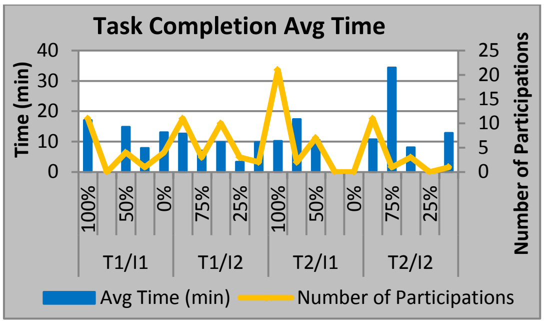


Fig. 4.6 Average Task Completion Time

4.4 First hypotheses: Correlation between Mental Workload and User Interaction

This section addresses the main objective of this study, expressed by the research question: Can objective indicators of users' interaction be used to approximately measure the subjective MWL of a web-based task?

All the values obtained for the indicators of User Interaction are shown in Appendix B, page 98, as well as all correlation Spearman coefficients, page 99. In this section, only at least moderate correlations are shown (absolute value > 0.5). Most significantly, no correlation under this criterion was found when trying to correlate indicators by considering all the dataset. In other words, only in subsets of the dataset (e.g. task 1 / interface 1; profile A) were found concrete correlations. Naturally, coefficient correlations near zero (i.e. absolute value < 0.2) are as important as strong correlations values. However, it would be impossible to comment all coefficient values calculated for the combination of the 23 indicators of User Interaction and the 14 dimensions/indicators of Mental Workload. Consequently, this chapter only addresses combinations which are correlated, according with the results obtained in this experiment.

Since scatter plots showed that, in cases of linear or monotonic relationships, outliers would affect the values of the Pearson coefficient, only the Spearman coefficient was used.

Note: All figures/charts shown in this section use bar's length in order to represent the absolute value of each correlation. Correlations related with the increase of one variable when the other decreases (downhill relationship) are labelled with "(-)".

4.4.1 Mouse Clicks

Figures 4.7 and 4.8 show the at least moderate correlations coefficients obtained after dividing the final dataset into profiles' and interfaces' subsets, respectively.

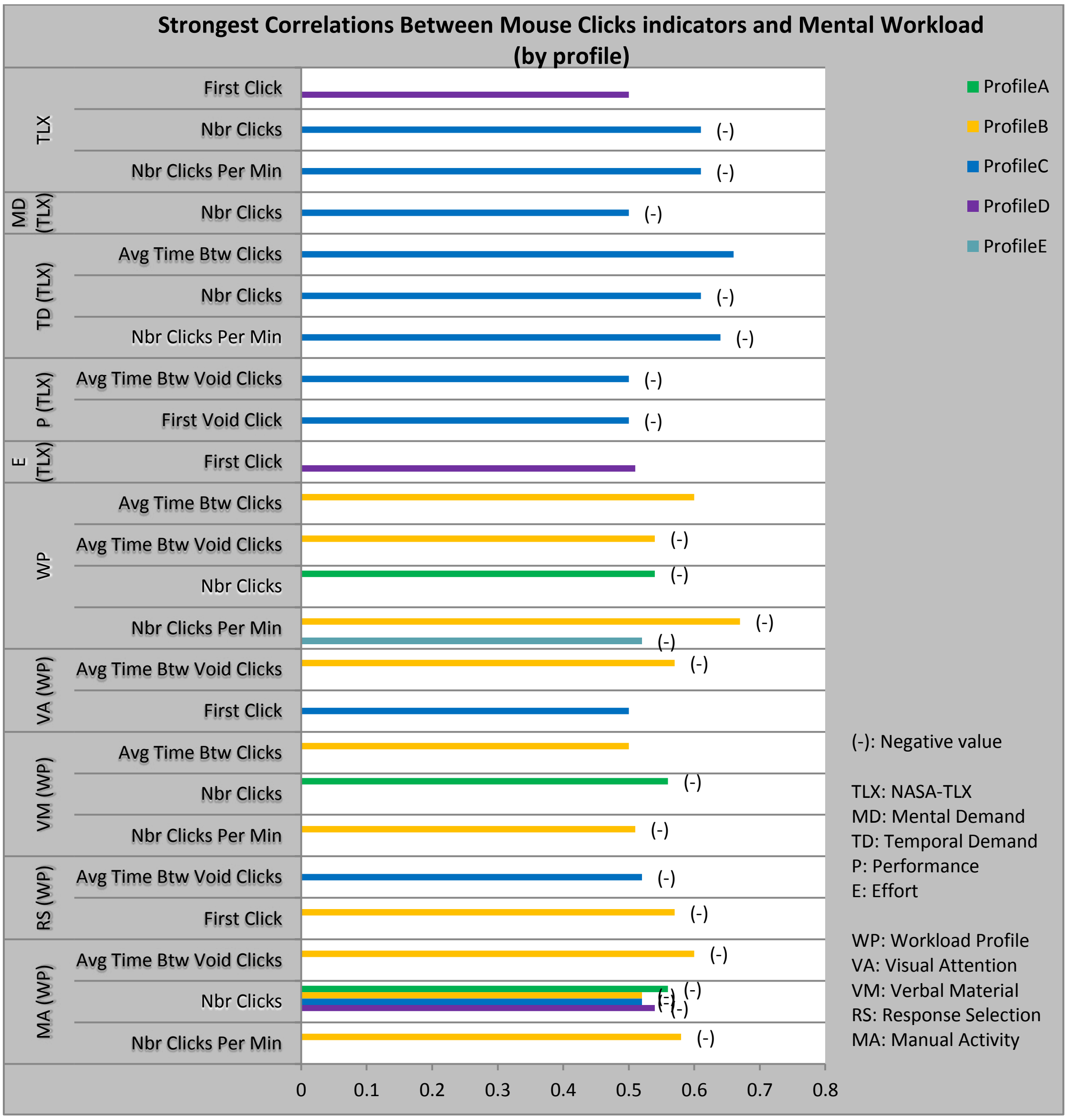


Fig. 4.7 Strongest correlations between Mouse Clicks indicators and Mental Workload (by profile)

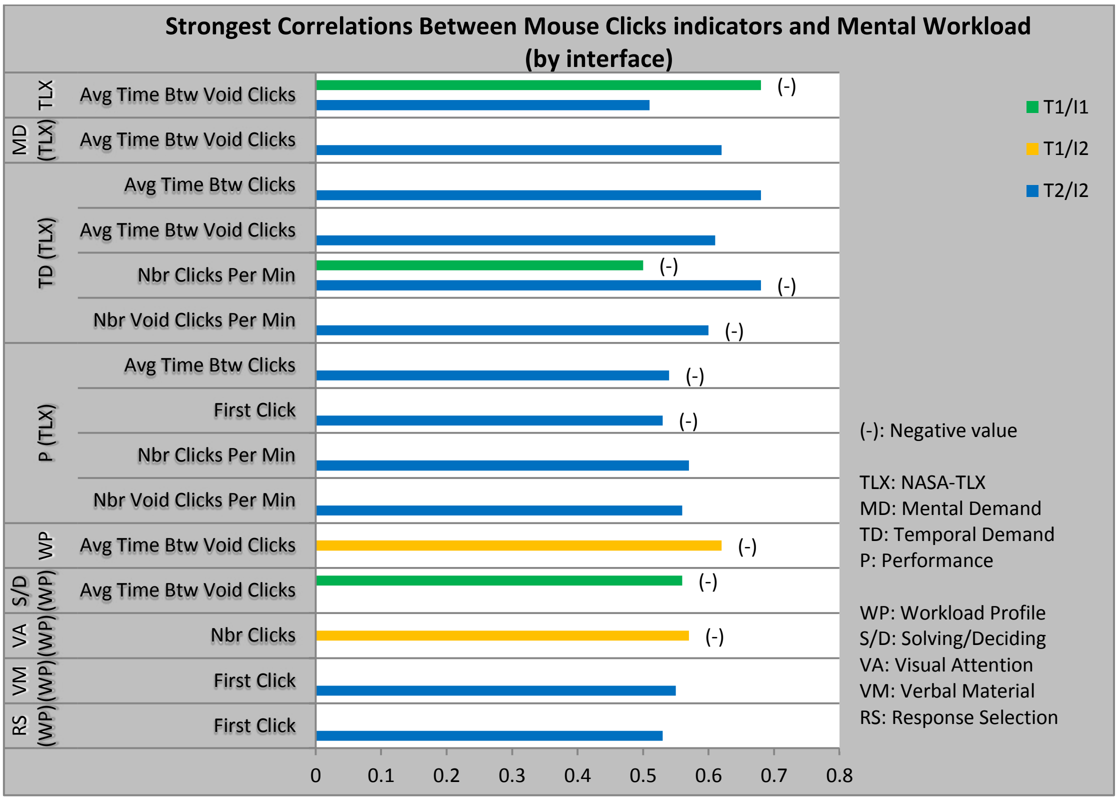


Fig. 4.8 Strongest correlations between Mouse Clicks indicators and Mental Workload (by interface)

Considering the correlations found when analysing subsets of each interface (figure 4.8), it is important to note that the most listed interaction indicator (Average Time between Void Clicks) is calculated by only considering participations that used void clicks. Moreover, interface 2 of task 2 (I2T2), which is also significantly listed, also refers to a statistical smaller dataset, since there were noticeable task abandonment rates for this interface. Nevertheless, some interesting correlations are shown. For instance, interface 2 of task 2 (T2I2) shows a strong uphill relationship (0.68) between temporal demand and the Average Time between Clicks and a strong downhill relationship (-0.68) between temporal demand and Number of clicks Per Minute, which indicates that users clicking often are less pressured by time, and feel that the task pace is slower, despite how intuitively contradictory it seems. Regarding the same interface, the most active subjects regarding mouse clicks seem to feel more confident when rating the performance they achieve (First Click / Number of Clicks (per minute) vs. Performance). This can indicate that less active users are having more trouble when interacting with this type of interfaces. Finally, it is interesting to notice that, for all interfaces, measures related with void clicks' activity, which can be related with the action of using the mouse's cursor as a reading aid, seem to be negatively (i.e. downhill) related to indicators of Mental Workload demand. These results suggest that subjects using this type of action are having more difficulties performing the task. Similar considerations can be made by analysing the results obtained for subsets defined by subjects' profile (figure 4.7). The substantial variety of indicators between profiles, shown by these results, suggests that correlations between these types of indicators are also dependent on subjects' traits.

4.4.2 Keyboard

Figures 4.9 and 4.10 show the correlations coefficients related with keyboard indicators obtained after dividing the final dataset into interfaces' and profiles' subsets, respectively.

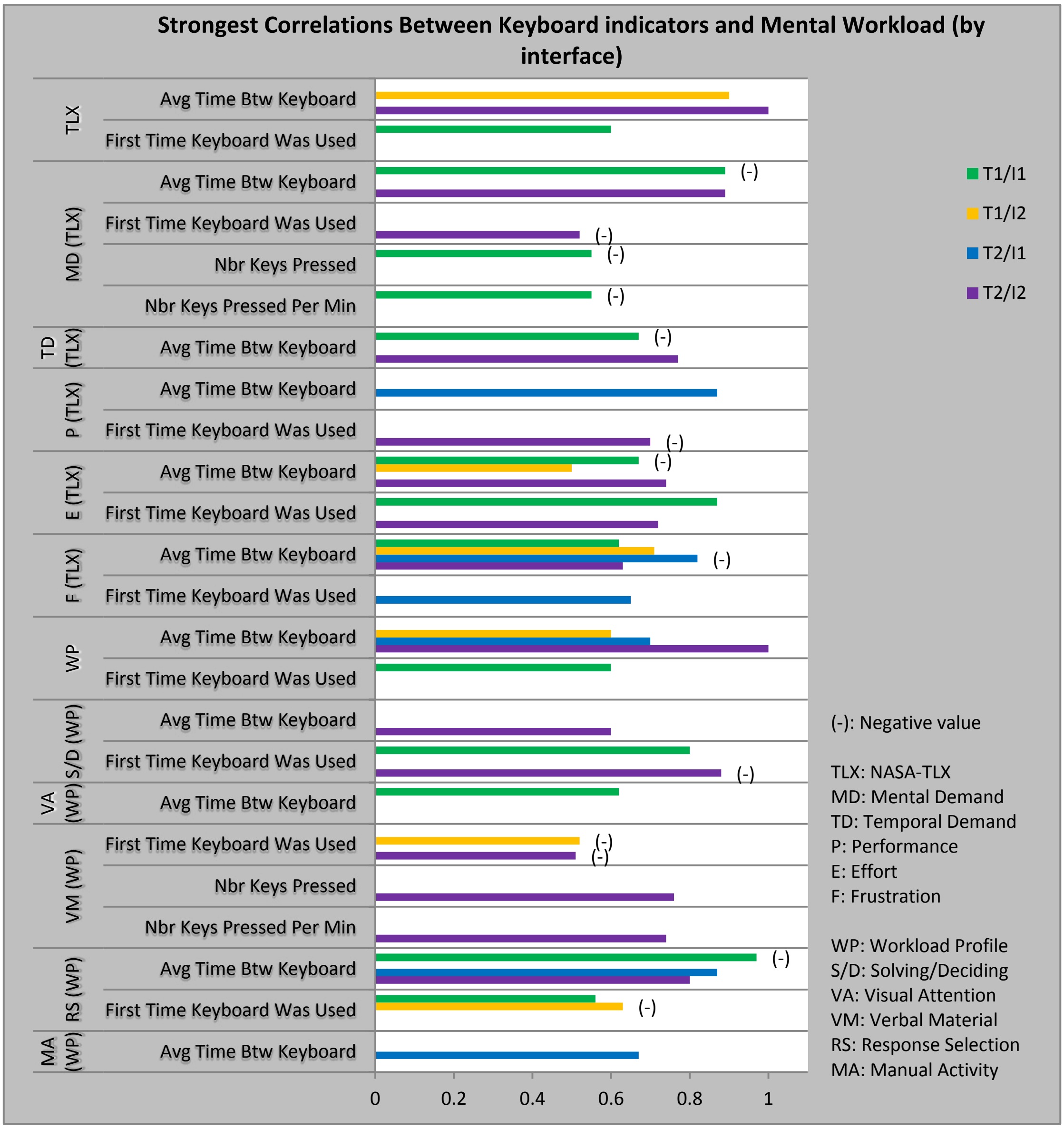


Fig. 4.9 Strongest correlations between Keyboard indicators and Mental Workload (by interface)

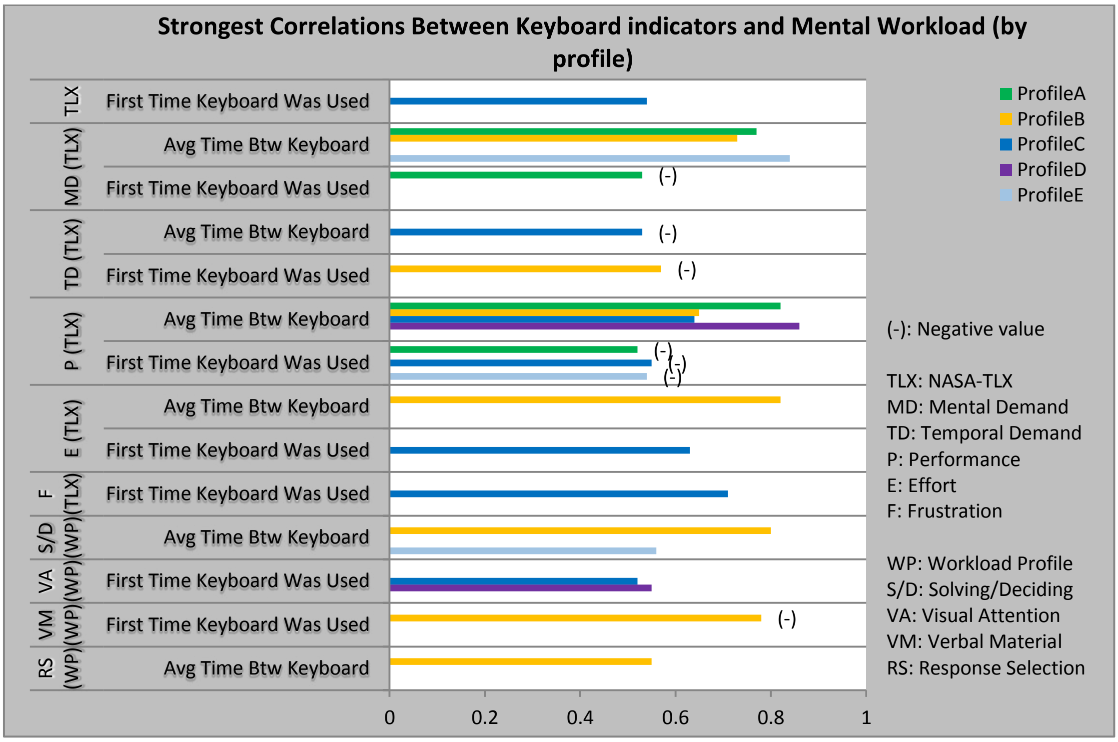


Fig. 4.10 Strongest correlations between Keyboard indicators and Mental Workload (by profile)

The results related with keyboard indicators are the strongest correlations with Mental Workload indicators found in this project. However, they have to be carefully analysed because none of the experiment's interfaces requested the use of the keyboard. Since it seems that there is a downhill/negative relation between MWL and participations that frequently use keyboard, it is possible that this keyboard activity is related with the use of browsers' short keys options (e.g. find word; go to the bottom of the page; navigate up or down), an less related with reactions of frustration (i.e. using the keyboard randomly as a way of dealing with frustration). This is supported by noticing that the majority of users classified themselves as expert web users.

Further research needs to be carried out in order to study several aspects related with keyboard indicators, such as: verify if the same correlations are found in tasks that specifically request the use of the keyboard; and distinguish between expert and random utilization of keyboard (not straightforward in online experiments since browsers' short keys can be customized).

4.4.3 Scrolling

Figures 4.11 and 4.12 show the at least moderate correlations (absolute value > 0.5) found between MWL and indicators of scrolling.

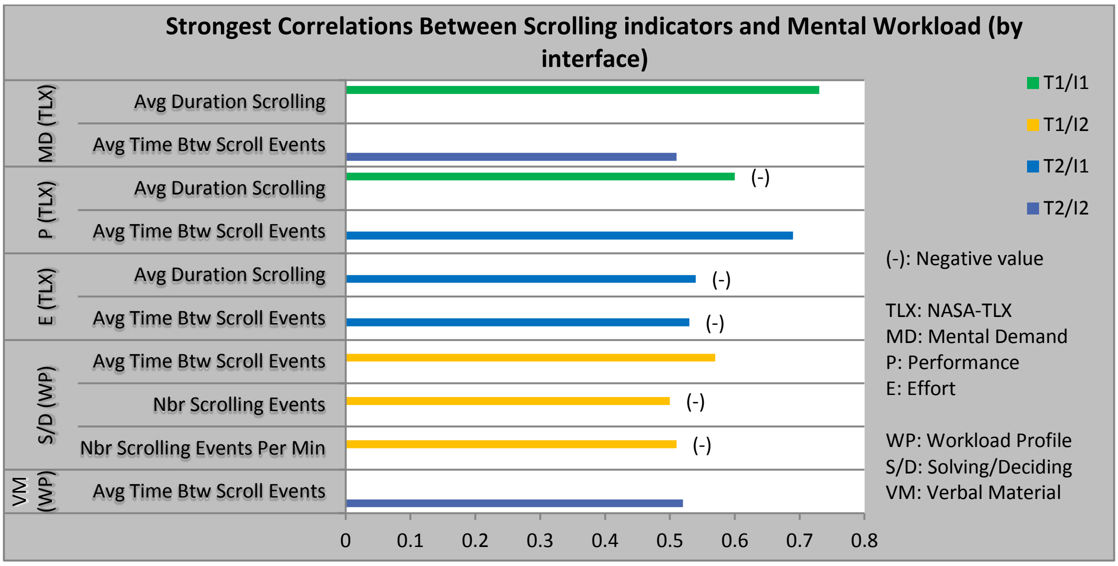


Fig. 4.11 Strongest correlations between Scrolling indicators and Mental Workload (by interface)

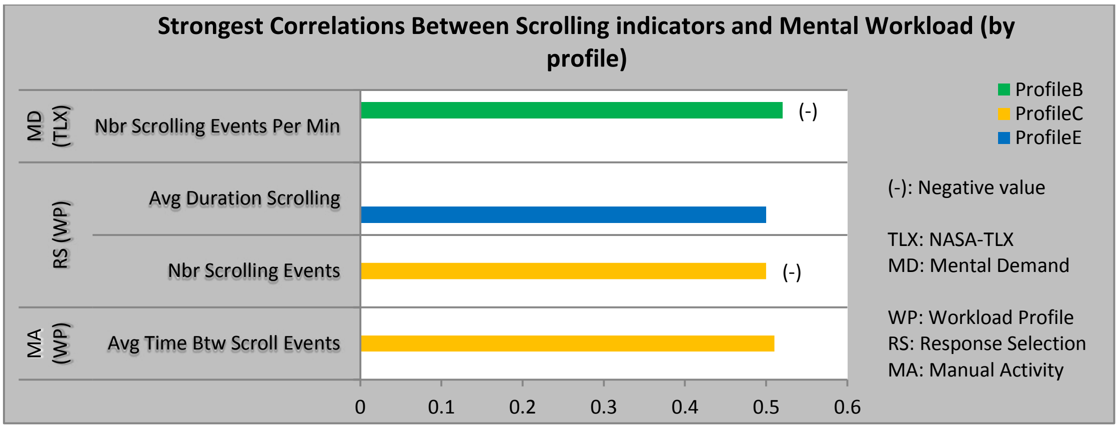


Fig. 4.12 Strongest correlations between Scrolling indicators and Mental Workload (by profile)

In the first place, knowing that only interface 1 of task 1 (T1I1) and interface 2 of task 2 (T2I2) demanded scrolling in order to explore the dataset, it is interesting to notice that, in interface 2 of task 1 (T1I2 – 2D map), the results show opposite (i.e. downhill) relationships between high scrolling activity and mental processes related with Solving and Deciding. A similar relationship is suggested in interface 1 of task 2 (T2I1, which also didn't require scrolling) but, this time, between high scrolling activity and effort. Moreover, self-performance is identified as being high for low scrolling activity. These results suggest that users performing scroll actions when they are not needed are having more difficulties. Similar considerations can be made for interfaces that demanded scrolling usage (interface 1 / task 1- I1T1; interface 2 / task 2 – I2T2), although for different MWL dimensions.

Similarly to previous analysed indicators of interaction, the diversity of correlations found within profiles' datasets (figure 4.12) suggests that users' traits possibly play an important role when assessing certain types of interaction indicators and mental demand dimensions.

4.4.4 Mouse Cursor

Tables 4.13 and 4.14 show the at least moderate correlations (absolute value > 0.5) found between MWL and mouse cursor's indicators.

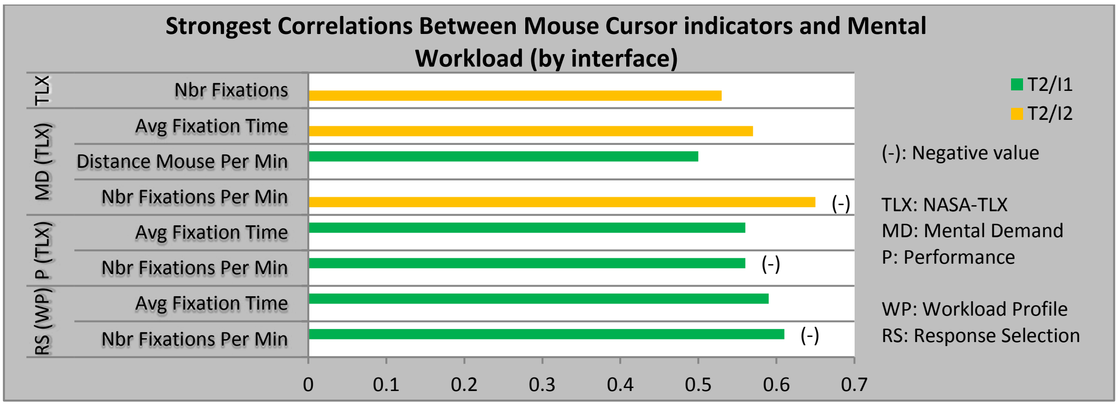


Fig. 4.13 Strongest correlations between Mouse Cursor indicators and Mental Workload (by interface)

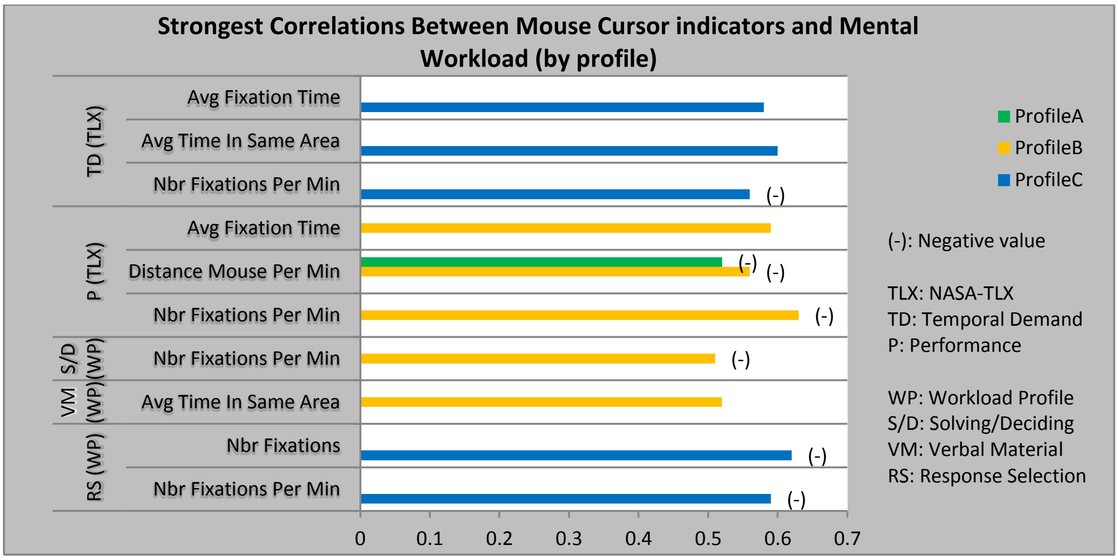


Fig. 4.14 Strongest correlations between Mouse Cursor indicators and Mental Workload (by profile)

These indicators were supported by studies that linked eye focus with mouse cursor's position and suggested that a high number of fixations is related with high mental demands. Regarding interfaces' datasets (figure 4.13), relationships are only found for task 2 (T2I1 and T2I2), which suggests that these correlations depend on the interface used and, possibly, in the type of task performed (although task 2 is similar to task 1, regarding the type of actions required, despite being shorter in time). In both interfaces of task 2 (T2I1 and T2I2), higher numbers of fixations (per minute) seem to not support literature that linked the frequency of these events with high mental demands. However, the moderate relationship between the number of total fixations, in interface 2 (T2I2), with the aggregated indicator of NASA-TLX points to other conclusions (although less relevant than the number of fixations per minute indicator).

Similar considerations regarding fixations can be found in profiles' subsets (figure 4.14), where bigger average times in fixation actions or in the same web window's area are linked with some mental dimensions' demands. Furthermore, profiles A and B show an inverse correlation between performance and the distance travelled with the mouse cursor per minute.

4.4.5 Final remark

The results shown in this section suggest that there is a set of indicators of User Interaction that present a moderate monotonic relationship with Mental Workload, which leads to support the hypotheses that this type of indicators can be used in order to predict the MWL demanded by a web task, validating hypothesis H1:

- ✓ Alternative hypothesis (H1): Objective indicators of users' interaction are linearly correlated with the subjective Mental Workload required by a web task.

However, the results suggest that this correlation is only found when taking in consideration the type or nature of the interface used or the profile of the user (in line with research of user interaction's patterns that suggested that mental fatigue prediction depends on the subject). This observation weakens the strength of the results because it implies that these positive correlations are supported by smaller samples which, therefore, are less statistical relevant. Moreover, these findings are restricted to the type of interfaces used, mainly related with interactive Infographics.

Significantly, it was found more weak correlations (i.e. absolute value of coefficient < 0.5) than strong correlations. Naturally, it is also important to notice the results that suggest that there is no correlation between certain interaction's indicators and certain mental workload resources (for instance, the indicators "total number of void clicks"; "first time scrolling was used" and "total distance travelled with mouse's cursor" didn't show any significant correlation with mental workload). On the whole, results suggest that the study of these relationships can only be made by considering the web interface in use. Additionally, most found correlations are related with single dimensions of the MWL measurement methods used.

Further studies need to be carried out in order to support these findings and highlight the most effective research path for correlating users' interaction and mental workload.

4.5 Second hypotheses: Mental Workload as a measure for evaluating interactive Infographics

Figure 4.15 and Figure 4.16 show the Scatter Plots for the values of performance (task accuracy) and MWL (NASA-TLW and Workload Profile, respectively) obtained.

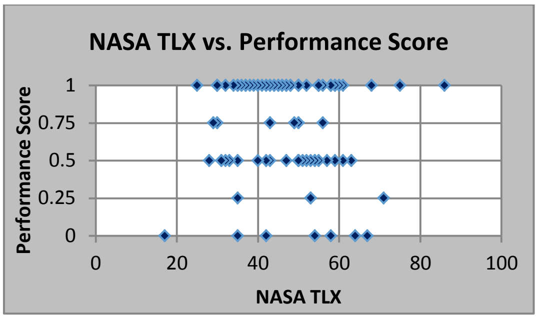


Fig. 4.15 Scatter Plot – NASA-TLX vs. Performance

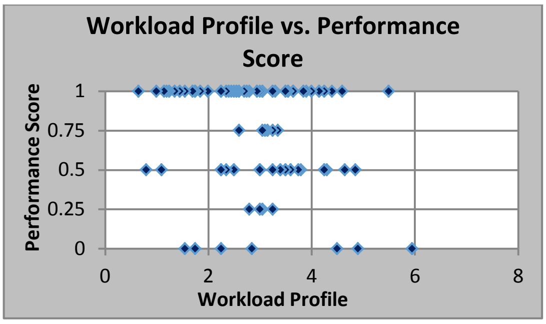


Fig. 4.16 Scatter Plot – Workload Profile vs. Performance

These scatter plots represent the data that directly aim to answer the second objective of this project: Can subjective Mental Workload be used to evaluate the performance of interfaces using different Infographics representations? The assumption defined was that it would be possible to link performance with the Mental Workload imposed by a web task. The assumption was based in literature that studied Mental Workload in the context of HCI, which highlighted Mental Workload as an important measure element in HCI design, and verified the same relation with performance that was found in other domains: low and high mental demands degrades performance (i.e. there is a range of optimum mental workload that optimizes performance). However, as shown by Figures 4.15 and 4.16, this is not verified for any of the two methods applied in the experiment to measure Mental Workload. Further, no relation is evident between the two indicators. These results meet the null hypothesis (H2) defined for this objective, and contradict hypothesis H3:

- ✓ Null hypothesis (H2): Subjective Mental Workload cannot be linked to Infographics interfaces' performance.

The results indicate that Infographics might require new approaches and considerations in order to understand the role of Mental Workload in the performance of tasks using this type of interfaces. This conclusion is in line with similar conclusions, expressed by studies in the domain of Usability, which also identified Infographics as having a different nature than other HCI interfaces and which argued the necessity of developing new methods of measuring Usability for this particular type of interfaces.

Nevertheless, the results also reinforce that Mental Workload is an aspect that should not be neglected in the evaluation of Infographics, since more directly extracted measures, such as performance, seemed to not be capable of expressing the mental demand imposed to users. Further investigation needs to be carried out in order to reinforce the results obtained in this experiment and in order to discover other indicators that can help estimate and understand performance and Mental Workload in interactive Infographics.

In order to verify if these conclusions are also verified within profiles and interfaces, other scatter plots were made, presented in Appendix B, page 107. Similarly, no relationship was identified between performance and Mental Workload.

4.6 Conclusion: Strength and Limitations of Findings

To summarize, the main findings related with the objectives of this study are the following:

- There are indicators of users' interaction that show moderate to strong correlations with dimensions of MWL. However, these correlations are only found when diving the main dataset into subsets of interfaces or subjects' profiles;
- In tasks that consist in exploring interactive Infographics, there is no evident correlation between performance and the imposed MWL.

Naturally, these conclusions are restricted to the major type of data representation used: Infographics. Furthermore, the fact that relevant correlations were found only when considering interfaces or profiles, implies that correlations coefficients were calculated with smaller samples and, therefore, are less statistical relevant. Additionally, the findings can also be categorized by a certain lack of diversity between participants (for instance, the majority were expert web users). On the other hand, this lack of diversity also strengthens findings, since it implies that there were less factors and variables interfering with the final results.

Moreover, the main strengths and limitations of the findings are directly linked with the strengths and limitations of the designed solution listed in the Design chapter, page 64. Of course, these strengths and limitations are mainly characterized by design tradeoffs. For instance, on one hand it was beneficial for the purposes of the study that the experiment was set online. But, on the other hand, it also implied that external factors such as devices' screen resolution and participants' rest levels were not controlled or standardized.

Additionally, regarding limitations, it should be noted that recent research questioned the suitability of using NASA-TLX in the context of web-based tasks, arguing that the method has its background in other domains (aircraft cockpits, supervisory and process control environments). This research adds that web based tasks are, in nature, associated with a wide spectrum of context of use, and operator's capabilities, such as motivation and skill, suggesting that new methods, based on argumentation theory, could be more suitable for covering the context's awareness and users' characteristics that are inherent to the MWL imposed by web tasks (Longo, 2014, 2015). Moreover, the multiple and somehow distinct definition of MWL as a concept has being addressed by recent studies which aimed to deploy new methods of inferring MWL by using artificial intelligence in the form of expert rule-based systems (Rizzo, Dondio, Delany, & Longo, 2016). Additionally, other studies identified task familiarity as a major factor in mental demands (Gwizdka, 2010), which was not controlled or assessed in this study.

The next chapter – Conclusion - overviews the project's objectives and findings. Moreover, suggestions and recommendations for future studies are made.

5 CONCLUSION

This section summarizes the main structure and findings of this project. The contributions for the scientific body of knowledge are also addressed, as well as suggestions and recommendations for future work.

5.1 Research Overview and Problem Definition

This project aimed to further study the role of MWL in the context of HCI. Previous studies indentified MWL as an important measure in the evaluation and designing of web interfaces. However, common Mental Workload measurement methods can usually only be applied in experiments and/or to a small sample of subjects and tasks. Since other domains of HCI have been leveraging the potential of logs of user interaction (e.g. mouse clicks; scrolling) in order to assess Usability or performance on a bigger scale (i.e. data from hundreds of users), it was suggested the possibility of using those logs in order to predict or measure the amounts of MWL demanded by web tasks.

Additionally, the advent of interactive Infographics as representations of data in the web, raised the question of assessing if these type of interfaces verify the same relationship between performance and MWL reported by previous studies: low and high mental demands cause the decrease of performance.

In order to address these objectives, an online experiment was set with four different interfaces and two different main tasks.

5.2 Experimentation, Evaluation & Results

The designed online experiment had 145 participations and collected MWL and performance with post-task surveys, whereas user interaction was collected with a JavaScript tracking API. The 23 indicators of user interaction defined were correlated with the 14 questions assessed by two subjective MWL measurement methods. The findings obtained suggest that correlations can only be verified when considering the type of interface used or the profile of the user that performed the task, and are more likely to be able to predict dimensions/resources of mental demand (i.e. effort; visual attention) than the overall Mental Workload demanded. Moreover, the results identify/suggest indicators that seem to be correlated (e.g. averaged duration of scrolling and mental demand in interactive Infographics using bar charts) as well as indicators that don't seem to be correlated (e.g. distance travelled with mouse and frustration in 2D interactive Maps). Additionally, the Mental Workload imposed by the interaction with web Infographics was not linked with task's performance, neither for the quadratic relationship reported by literature nor for any other type of evident relationship.

5.3 Contributions and impact

Regarding the possible correlation between indicators of User Interaction and Mental Workload, this project unveils and sets the tone for future and similar studies. The main premises suggested by the results propose that, in the context of interactive Infographics, correlations should be studied taking in consideration the type of interfaces used as well as users' profiles. On the account of having found strong correlations only within subsets of the main experiment's dataset (i.e. interface subset; profile subset), these findings are less statistical significant, which also implies the need of setting additional instances of the performed experiment in order to better support the findings achieved by this project. Moreover, the findings identify which indicators of interaction are likely to help predicting MWL (e.g. users with low frequency of clicks were found to have higher temporal demands).

In brief, the findings support the defined alternative hypothesis (H1) for the main research question of this project: Can objective indicators of users' interaction be used to approximately measure the subjective MWL of a web-based task?

- Alternative hypothesis (H1): Objective indicators of users' interaction are linearly correlated with the subjective MWL required by a web task.

Regarding the secondary objective of studying the impact of MWL in the performance of web tasks that consist in exploring interactive Infographics, the findings suggest that this type of interfaces are very different in nature than other typical web interfaces. This is in line with considerations made by other studies in different domains of HCI. In short, results didn't present any evident relationship between MWL and task's performance, which leads to the likely support of the null hypothesis (H2) defined for the secondary research question: Can subjective Mental Workload be used to evaluate the performance of interfaces using different Infographics representations?

- Null hypothesis (H2): Subjective Mental Workload cannot be linked to Infographics interfaces' performance.

These findings suggest that further studies should aim to discover other indicators that can help estimate and understand performance and MWL in the context of interactive Infographics, which seem to be different in nature than more common and more studied web interfaces.

5.4 Future Work & recommendations

Regarding future research, the main consideration that can be made in relation with this project's findings is that further experiments should be carried out in order to support the results found, which would allow strengthening the statistical relevance of the findings and planning future research accordingly. One the one hand, it would be useful to repeat or continue the experiment in order to reach a wider audience, which would allow obtaining a more significant and reliable sample. On the other hand, new experiments could include alternative design elements that would allow studying additional factors, for instance:

- Define other types of task or interfaces in order to verify if the same findings are observed;
- Design a wider variety of tasks' duration and complexity (also in order to verify if the same findings are observed);
- Further study the findings found in relation to the usage of keyboard and scrolling when the task does not request their use. Particularly, in the case of keyboard's usage, it would be interesting to distinguish between expert (i.e. browser's short keys) and random utilization of keyboard. In the sane line of

though, it would be interesting to identify void clicks that are linked with the action of using the mouse cursor as a reading aid;

- Use Mental Workload measurement methods capable of assessing mental demands in real time (e.g. dual tasking; eye activity). This would allow correlating patterns of user interaction with peaks of Mental Workload and/or correlating users' interaction and Mental Workload with specific elements of web visualisation, such as dropdown lists and radio buttons;
- Define other indicators of User Interaction. For instance, "total distance scroll distance (pixies)" and "maximum scroll top (pixies)" used by Guo and Agichtein (2012), who performed an Usability study. Another interesting example would be detecting mouse cursor actions characterised by quick shifts between locations. This type of indicators would be based on the concept of "Saccades", which are described, by a study that assessed Mental Workload in HCI with eye tracking measures, as "a shift between two locations. When the eye performs a voluntary movement from one fixation to another (...) and typically takes from 30-80 milliseconds to complete (...) we propose an influence of Cognitive Load on saccade length (the higher the load, the longer the saccades) and saccade velocity (the higher the load the higher the saccade velocity)" (Zagermann et al., 2016, p. 80). Based on previous studies that linked eye movement with the mouse cursor's position, saccades could be extracted from mouse cursor logs. It would be also interesting to identify switches of attention between HTML elements: for instance from a menu to a graph;
- Track the web interaction of the same subject for a period of time across several experiments. By continuously assessing the subjective Mental Workload of individual subjects, it could be possible to develop an algorithm to predict Mental Workload, based on user's interactivity, for that particular individual. This would be in line with recent studies that link individual interaction patterns with mental fatigue (Pimenta et al., 2013).

Additionally, the dataset obtained by the online experiment performed can be a source of valuable and insightful data analysis for other studies. For instance:

- Knowledge extraction techniques can be applied (e.g. machine learning), which is in line with studies that predict the mental fatigue of a particular user by using patterns identified by machine learning techniques;

- Analyse users' behaviour (e.g. sequence of steps and actions), as it was done in previous Usability (Qi Guo & Agichtein, 2012) and Mental Workload (F. Chen et al., 2012) studies. For instance, linking the curiosity or engagement produced by an interface with the detection of subjects exploring data and information that were not related with the task's goals, or identifying users' search strategy and actions, which was studied and correlated with Mental Workload by Gwizdka (2010) and Navarro-prieto et al. (1999). Another example of a possible behavioural study would be to detect different patterns of interaction between the alternative interfaces;
- Study if there are other types of relationships between Mental Workload and User Interaction indicators (e.g. quadratic relationships). Or study if the value of one indicator is related with the value of two or more indicators (i.e. multiple regression analysis). This study exclusively focused on relationships characterized by correlations where two variables tend to change together (in a positive or negative direction);
- Generate mouse position graphics, which allows identifying the main areas of subjects' focus. This approach is often used in Usability studies (Qi Guo & Agichtein, 2012);
- Assess the correlation between indicators of user interaction and task abandonment.

6 BIBLIOGRAPHY

Agichtein, E., Brill, E., Dumais, S., & Ragno, R. (2006). Learning user interaction models for predicting web search result preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 3–10). New York, New York, USA: ACM Press. https://doi.org/10.1145/1148170.1148175

Albers, M. J. (2011). Tapping as a measure of cognitive load and website usability. In *Proceedings of the 29th ACM international conference on Design of communication - SIGDOC '11* (p. 25). https://doi.org/10.1145/2038476.2038481

Aparicio, M., & Costa, C. J. (2015). Data visualization. *Communication Design Quarterly Review*, *3*(1), 7–11. https://doi.org/10.1145/2721882.2721883

Arroyo, E., Selker, T., & Wei, W. (2006). Usability tool for analysis of web designs using mouse tracks. *CHI 06 Extended Abstracts on Human Factors in Computing Systems CHI 06*, *Montr\éal,*(3), 484. https://doi.org/10.1145/1125451.1125557

Athukorala, K., Gowacka, D., Jacucci, G., Oulasvirta, A., & Vreeken, J. (2016). Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, *67*(11), 2635–2651. https://doi.org/10.1002/asi.23617

Atterer, R., Wnuk, M., & Schmidt, A. (2006). Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. *Proceedings of the 15th Int. Conf. on World Wide Web*, 203–212. https://doi.org/10.1145/1135777.1135811

Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation*, *2013*(139), 7–32. https://doi.org/10.1002/ev.20065

Belmore, S. M. (1985). Reading computer-presented text. *Bulletin of the Psychonomic Society*, *23*(1), 12–14. https://doi.org/10.3758/BF03329765

Bergaus, M. (2015). *Design Issues for Service Delivery Platforms: Incorporate User Experience: A Grounded Theory Study of Individual User Needs.* Retrieved from http://dx.doi.org/10.1007/978-3-658-10541-9

Bevan, N. (2006). International Standards for HCI. *Encyclopedia of Human Computer Interaction*, *55*(May), 1–15. https://doi.org/10.1006/ijhc.2001.0483

Bolchini, D., Finkelstein, A., Perrone, V., & Nagl, S. (2009). Better bioinformatics through usability analysis. *Bioinformatics*, *25*(3), 406–412. https://doi.org/10.1093/bioinformatics/btn633

Budiu, B. R., & Nielsen, J. (2011). Usability of iPad Apps and Websites. *Small*, (2nd Edition), 1–129. Retrieved from http://www.nngroup.com/reports/mobile/ipad/

Cain, B. (2007). A Review of the Mental Workload Literature. *Defence Research and Development Toronto (Canada)*, (1998), 4-1-4–34. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA474193

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: software for assessing subjective mental workload. *Behav Res Methods*, *41*(1), 113–117. https://doi.org/10.3758/brm.41.1.113

Cawthon, N., & Moere, A. Vande. (2007). The effect of aesthetic on the usability of data visualization. *Proceedings of the International Conference on Information Visualisation*, 637–645. https://doi.org/10.1109/IV.2007.147

Cegarra, J., & Chevalier, A. (2008). The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological

improvements. *Behavior Research Methods*, *40*(4), 988–1000. https://doi.org/10.3758/brm.40.4.988
Chao, G. (2009). Human-computer interaction: Process and principles of human-computer interface design. *2009 International Conference on Computer and Automation Engineering, ICCAE 2009, March 8, 2009 - March 10, 2009*, 230–233. https://doi.org/10.1109/ICCAE.2009.23
Chen, F., Ruiz, N., Choi, E., Epps, J., Khawaja, M. A., Taib, R., … Wang, Y. (2012). Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems*, *2*(4), 1–36. https://doi.org/10.1145/2395123.2395127
Chen, M. (2001). What can a mouse cursor tell us more? Correlation of eye / mouse movements on web browsing. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 281–282.
Churchill, E. F., Bowser, A., & Preece, J. (2013). Teaching and learning human-computer interaction. *Interactions*, *20*(2), 44. https://doi.org/10.1145/2427076.2427086
Cleveland, W., & McGill, R. (1985). Graphical Perception and Graphical Methods for Analyzing Scientific Data. *American Association for the Advancement of Science*. https://doi.org/10.1126/science.229.4716.828
Cooke, L. (2006). Is the Mouse a“ Poor Man's Eye Tracker”? *Annual Conference-Society for Technical Communication*, *53*, 252–255. Retrieved from http://www.stcsig.org/sn/PDF/Cooke_mouse_eye_tracker.pdf
Cowan, N. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277
de Leeuw, J. R., & Motz, B. a. (2015). Psychophysics in a web browser? Comparing response times collected with JavaScript and psychophysics toolbox in a visual search task. *Behavior Research Methods*, (2014), 1–12. https://doi.org/10.3758/s13428-015-0567-2
Durantin, G., Gagnon, J. F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, *259*, 16–23. https://doi.org/10.1016/j.bbr.2013.10.042
Egeth, H., & Kahneman, D. (1975). Attention and Effort. *The American Journal of Psychology*, *88*(2), 339. https://doi.org/10.2307/1421603
Evans, D. C., & Fendley, M. (2017). A multi-measure approach for connecting cognitive workload and automation. *International Journal of Human Computer Studies*, *97*, 182–189. https://doi.org/10.1016/j.ijhcs.2016.05.008
Felton, E. A., Williams, J. C., Vanderheiden, G. C., & Radwin, R. G. (2012). Mental workload during brain-computer interface training. *Ergonomics*, *55*(5), 526–37. https://doi.org/10.1080/00140139.2012.662526
Fernandez, A., Insfran, E., & Abrahão, S. (2011). Usability evaluation methods for the web: A systematic mapping study. In *Information and Software Technology* (Vol. 53, pp. 789–817). https://doi.org/10.1016/j.infsof.2011.02.007
Few, S. (2004). Data Presentation : TAPPING THE POWER OF VISUAL PERCEPTION. *Intelligent Enterprise*, *13*(7), 33.
Few, S. (2012). *Show me the numbers : designing tables and graphs to enlighten*. Analytics Press. Retrieved from http://library.wur.nl/WebQuery/clc/2175649
Freitas, C. M. D. S., Luzzardi, P. R. G., Cava, R. A., Winckler, M., Pimenta, M. S., & Nedel, L. P. (2002). On evaluating information visualization techniques. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 373–374.
Friendly, M. (2008). A Brief History of Data Visualization. *Handbook of Data*

Visualization, 15–56 (1–43). https://doi.org/10.1007/978-3-540-33037-0_2
Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., & Rao, R. P. N. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *Proceedings of the 26th SIGCHI Conference on Human Factors in Computing Systems*, 835. https://doi.org/10.1145/1357054.1357187
Guo, Q., & Agichtein, E. (2012). Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. *Www12*, 569–578. https://doi.org/10.1145/2187836.2187914
Guo, Q., & Agichtein, E. (2012). Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. *Www12*, 569–578. https://doi.org/10.1145/2187836.2187914
Gwizdka, J. (2009). Assessing Cognitive Load on Web Search Tasks. *The Ergonomics Open Journal*, *2*(4), 113–123. https://doi.org/10.2174/1875934300902020114
Gwizdka, J. (2010). Distribution of cognitive load in Web search. *Journal of the American Society for Information Science and Technology*, *61*(11), 2167–2187. https://doi.org/10.1002/asi.21385
Haga, S., Shinoda, H., & Kokubun, M. (2002). Effects of task difficulty and time-on-task on mental workload. *Japanese Psychological Research*, *44*(3), 134–143. https://doi.org/10.1111/1468-5884.00016
Harrison, L., Reinecke, K., & Chang, R. (2015). Infographic Aesthetics. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 1187–1190. https://doi.org/10.1145/2702123.2702545
Hart, Sandra, G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meting*, 904–908. https://doi.org/10.1037/e577632012-009
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9
Hartley, J. (1994). Designing instructional text for older readers: a literature review. *British Journal of Educational Technology*, *25*(3), 172–188.
Hartley, J. (2004). Designing instructional and informational text. *Handbook of Research in Educational Communications and Technology*, 917–947.
Heger, D., Putze, F., & Schultz, T. (2010). Online workload recognition from EEG data during cognitive tests and human-machine interaction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6359 LNAI*, 410–417. https://doi.org/10.1007/978-3-642-16111-7_47
Huey, B. M., & Wickens, C. D. (1993). *Workload transition: implication for individual and team performance*. Retrieved from https://www.google.com/books?hl=en&lr=&id=UXkRhCIM72AC&oi=fnd&pg=PT17&dq=workload+transition+implication+for+individual&ots=wu317UJI1G&sig=iMVfpC6aqIqLcT3TPplUGJ7c4PQ
Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, *33*(4), 470–516. https://doi.org/10.1145/503112.503114
Kamm, P. E. E., & Gevins, A. S. (1995). Neurocognitive adaptive computer interface method and system based on on-line measurement of the user's mental effort. *Accounting, Auditing & Accountability Journal*, *24*(6), 781–809. https://doi.org/10.1108/09513571111155546
Kendall, K. E., & Kendall, J. E. (2011). *Systems analysis and design*. Pearson Prentice Hall.

Klingner, J. (2010). *Measuring cognitive load during visual tasks by combining pupillometry and eye tracking*. *STANFORD UNIVERSITY*.
Kobsa, a. (2001). An empirical comparison of three commercial information visualization systems. *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.*, 123–130. https://doi.org/10.1109/INFVIS.2001.963289
Koua, E. L., Maceachren, a., & Kraak, M. - J. (2006). Evaluating the usability of visualization methods in an exploratory geovisualization environment. *International Journal of Geographical Information Science*, *20*(4), 425–448. https://doi.org/10.1080/13658810600607550
Longo, L. (2011). Human-Computer Interaction and Human Mental Workload : Assessing Cognitive Engagement in the World Wide Web. *Ifip International Federation For Information Processing*, (October 2015), 402–405. https://doi.org/10.1007/978-3-642-23768-3
Longo, L. (2012). Formalising human mental workload as non-monotonic concept for adaptive and personalised web-design. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7379 LNCS*(October), 369–373. https://doi.org/10.1007/978-3-642-31454-4_38
Longo, L. (2014). *Formalising Human Mental Workload as a Defeasible Computational Concept*. Trinity College Dublin.
Longo, L. (2015). A defeasible reasoning framework for human mental workload representation and assessment. *Behaviour & Information Technology*, *3001*(August), 1–29. https://doi.org/10.1080/0144929X.2015.1015166
Longo, L. (2016). Mental Workload in Medicine: Foundations, Applications, Open Problems, Challenges and Future Perspectives. *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, 106–111. https://doi.org/10.1109/CBMS.2016.36
Longo, L., & Barrett, S. (2010). A Computational Analysis of Cognitive Effort. In *Intelligent Information and Database Systems SE - 8* (Vol. 5991, pp. 65–74). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12101-2_8
Longo, L., & Dondio, P. (2015). On the Relationship between Perception of Usability and Subjective Mental Workload of Web Interfaces. *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 345–352. https://doi.org/10.1109/WI-IAT.2015.157
Longo, L., & Kane, B. (2011). A novel methodology for evaluating user interfaces in health care. *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*, (October 2015), 1–6. https://doi.org/10.1109/CBMS.2011.5999024
Longo, L., Rusconi, F., Noce, L., Barrett, S., Rusconi, F., Noce, L., & Barrett, S. (2012). The Importance of Human Mental Workload in Web Design. *WEBIST 2012 - Proceedings of the 8th International Conference on Web Information Systems and Technologies*, (October), 403–409. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84864868048&partnerID=40&md5=e556a67dae7ef680cc16d24381384c4f
Mandler, G. (1967). Organization and Memory. *Psychology of Learning and Motivation*, *1*, 327–372. https://doi.org/10.1016/S0079-7421(08)60516-2
Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, *49*(4), 41. https://doi.org/10.1145/1121949.1121979
Mayes, D. K., Sims, V. K., & Koonce, J. M. (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, *28*(6), 367–378. https://doi.org/10.1016/S0169-8141(01)00043-9

Mayhew, D. (1999). The usability engineering lifecycle. *CHI'99 Extended Abstracts on Human Factors in Computer Systems*, 147–148. https://doi.org/10.1145/632716.632805
Navarro-prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive strategies in web searching. *Proceedings of the 5th Conference on Human Factors & the Web*, (1999), 43–56.
Nielsen, J. (1993). *Usability engineering*. *IEEE Transactions on Visualization and Computer Graphics* (Vol. 14). AP Professional. https://doi.org/10.1109/TVCG.2008.24
Noyes, J. M., & Garland, K. J. (2003). VDT versus paper-based text: Reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics*, *31*(6), 411–423. https://doi.org/10.1016/S0169-8141(03)00027-1
Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411
Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1
Pantic, M., Nijholt, A., Pentland, A., & Huanag, T. S. (2008). Human-Centred Intelligent Human Computer Interaction (HCI^2): how far are we from attaining it? *International Journal of Autonomous and Adaptive Communications Systems*, *1*(2), 168. https://doi.org/10.1504/IJAACS.2008.019799
Patrick Rau, P.-L., & Hsu, J.-W. (2005). Interaction Devices and Web Design for Novice Older Users. *Educational Gerontology*, *31*(1), 19–40. https://doi.org/10.1080/03601270590522170
Pearson, J. M., Pearson, A., & Green, D. (2007). Determining the importance of key criteria in web usability. *Management Research News*, *30*(11), 816–828. https://doi.org/10.1108/01409170710832250
Pimenta, A., Carneiro, D., Neves, J., & Novais, P. (2016). A neural network to classify fatigue from human-computer interaction. *Neurocomputing*, *172*, 413–426. https://doi.org/10.1016/j.neucom.2015.03.105
Pimenta, A., Carneiro, D., Novais, P., & Neves, J. (2013). Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8073 LNAI, pp. 222–231). Springer. https://doi.org/10.1007/978-3-642-40846-5_23
Quesenbery, W. (2001). What Does Usability Mean : Looking Beyond " Ease of Use ." *Proceedings of the 48th Annual Conference, Society for Technical Communication*, 1–8.
Rizzo, L., Dondio, P., Delany, S. J., & Longo, L. (2016). Modeling Mental Workload Via Rule-Based Expert System: A Comparison with NASA-TLX and Workload Profile (pp. 215–229). https://doi.org/10.1007/978-3-319-44944-9_19
Rosson, M. B., & Carroll, J. M. (John M. (2002). *Usability engineering : scenario-based development of human-computer interaction*. *Interface*. Academic Press.
Ru, G., & Ming, Z. Y. (2014). Infographics applied in design education. In *2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)* (pp. 984–986). IEEE. https://doi.org/10.1109/WARTIA.2014.6976439
Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology*, *53*(1), 61–86. https://doi.org/10.1111/j.1464-0597.2004.00161.x

Saleh, B., Dontcheva, M., Hertzmann, A., & Liu, Z. (2015). Learning Style Similarity for Searching Infographics. *GI*, 3–5. Retrieved from http://arxiv.org/abs/1505.01214

Saraiya, P., North, C., & Duca, K. (2005). An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *11*(4), 443–456. https://doi.org/10.1109/TVCG.2005.53

Saraiya, P., North, C., Lam, V., & Duca, K. A. (2006). An Insight-Based Longitudinal Study of Visual Analytics. *IEEE Trans. Visualization and Computer Graphics*, *12*(6), 1511–1522.

Scholtz, J. (2006). Beyond usability: Evaluation aspects of visual analytic environments. *IEEE Symposium on Visual Analytics Science and Technology 2006, VAST 2006 - Proceedings*, 145–150. https://doi.org/10.1109/VAST.2006.261416

Silva, F. P. da. (2014). Mental Workload, Task Demand and Driving Performance: What Relation? *Procedia - Social and Behavioral Sciences*, *162*(Panam), 310–319. https://doi.org/10.1016/j.sbspro.2014.12.212

Siricharoen, W. V. (2013). Infogragraphics: The New communication Tools in Digital Age. *The International Conference on E-Technologies and Business on the Web*, (April), 169–174. https://doi.org/10.1177/1468794111399838

Tory, M., & Möller, T. (2005). Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, *25*(5), 8–11. https://doi.org/10.1109/MCG.2005.102

Tracy, J. P., & Albers, M. J. (2006). Measuring cognitive load to test the usability of web sites. *STC's 53rd Annual Conference Proceedings 2005*, (October), 256–260.

Tsang, P. S., & Vidulich, M. A. (2006). Mental Workload and Situation Awareness. In *Handbook of Human Factors and Ergonomics (Third Edition)* (pp. 243–268). Retrieved from http://dx.doi.org/10.1002/0470048204.ch9

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors*, *43*(1), 111–121. https://doi.org/10.1518/001872001775992570

van Steenbergen, H., & Bocanegra, B. R. (2015). Promises and pitfalls of Web-based experimentation in the advance of replicable psychological science: A reply to Plant (2015). *Behavior Research Methods*, (2015), 1–5. https://doi.org/10.3758/s13428-015-0677-x

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, *2*(3), 270–285. https://doi.org/10.1037/1076-898X.2.3.270

Ward, R. D., & Marsden, P. M. (2003). Physiological responses to different WEB page designs. *International Journal of Human Computer Studies*, *59*(1–2), 199–212. https://doi.org/10.1016/S1071-5819(03)00019-3

Wastell, D. G., & Newman, M. (1996). Stress, control and computer system design: A psychophysiological field study. *Behaviour & Information Technology*, *15*(3), 183–192. https://doi.org/10.1080/014492996120247

Wästlund, E., Norlander, T., & Archer, T. (2008). The effect of page layout on mental workload: A dual-task experiment. *Computers in Human Behavior*, *24*(3), 1229–1245. https://doi.org/10.1016/j.chb.2007.05.001

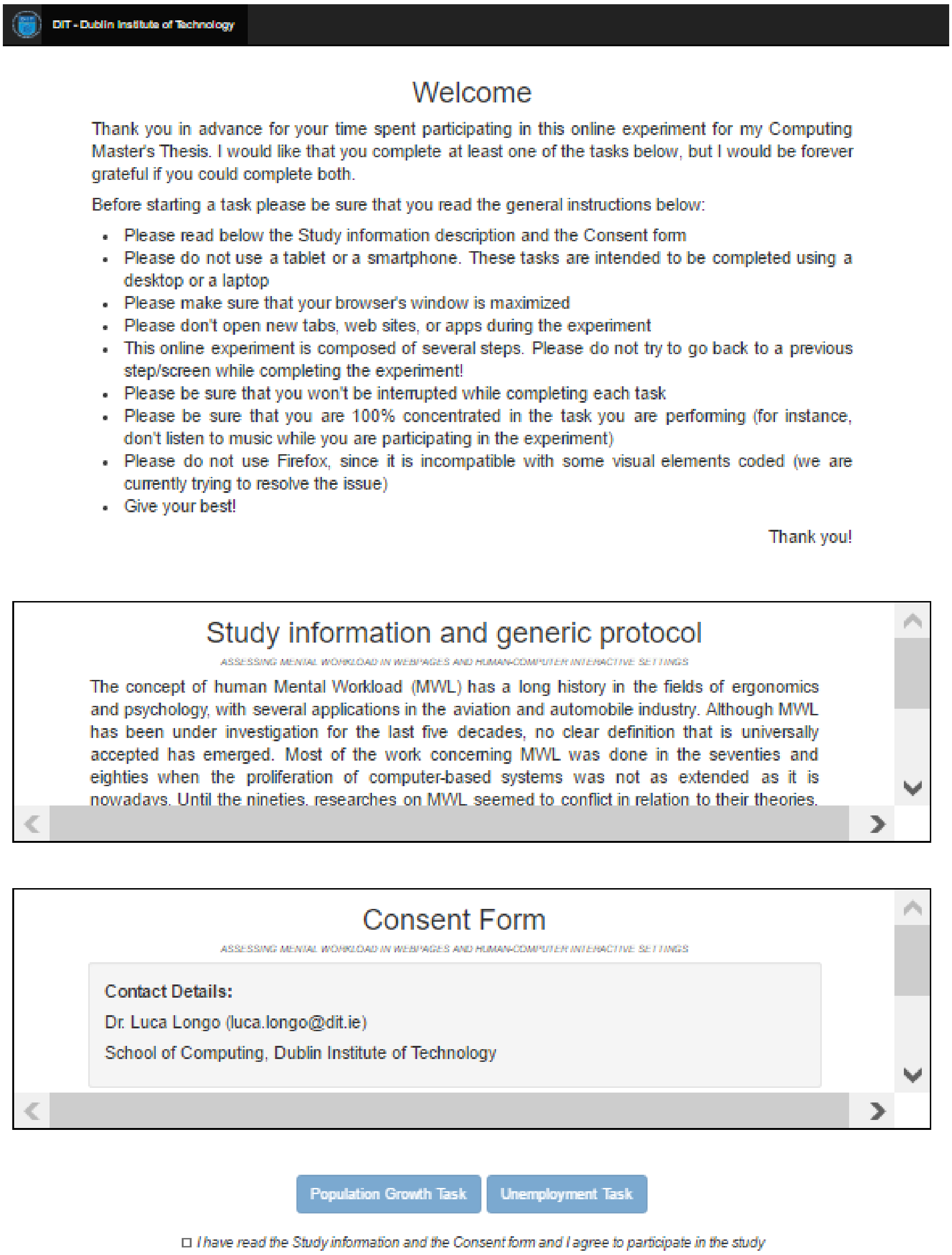
Wästlund, E., Reinikka, H., Norlander, T., & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior*, *21*(2), 377–394. https://doi.org/10.1016/j.chb.2004.02.007

Welie, M. Van, Veer, G. C. Van Der, & Eliëns, A. (1999). Breaking down usability. *Proceedings of INTERACT'99*, (SEPTEMBER 2001), 613–620. Retrieved from

http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.2666
Wickens, C. D. (1991). Processing Resources in Attention. *Multiple-Task Performance*, 61. https://doi.org/R. Parasuraman, R. Davies
Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806
Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. *Human Factors*, *35*(2), 263–281. https://doi.org/10.1177/001872089303500205
Womack, R. (2014). Data Visualization and Information Literacy. *Iassist*, 12–17. https://doi.org/10.7282/T3X92CZF
Woods, A. T., Velasco, C., Levitan, C. A., Wan, X. X., & Spence, C. (2015). Conducting perception research over the Internet: a tutorial review. *PeerJ PrePrints*, *3*, e1058. https://doi.org/10.7717/peerj.1058
Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work & Stress*, *14*(1), 74–99. https://doi.org/10.1080/026783700417249
Yamagishi, M., Kobayashi, T., Kobayashi, T., Nagami, M., Shimazu, A., & Kageyama, T. (2007). Effect of web-based assertion training for stress management of Japanese nurses. *Journal of Nursing Management*, *15*(6), 603–607. https://doi.org/10.1111/j.1365-2834.2007.00739.x
Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, *3*(2), 178–194. https://doi.org/10.1080/14639220210123789
Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring Cognitive Load Using Eye Tracking Technology in Visual Computing. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 78–85. https://doi.org/10.1145/2993901.2993908
Zhu, H., & Hou, M. (2009). Restrain mental workload with roles in HCI. *TIC-STH'09: 2009 IEEE Toronto International Conference - Science and Technology for Humanity*, 387–392. https://doi.org/10.1109/TIC-STH.2009.5444469
Zinovyev, A. (2011). Data visualization in political and social sciences. *International Encyclopedia of Political Science*. Retrieved from http://arxiv.org/abs/1008.1188

APPENDIX A ONLINE EXPERIMENT

A.1 Welcome Page


DIT - Dublin Institute of Technology

Welcome

Thank you in advance for your time spent participating in this online experiment for my Computing Master's Thesis. I would like that you complete at least one of the tasks below, but I would be forever grateful if you could complete both.

Before starting a task please be sure that you read the general instructions below:

- Please read below the Study information description and the Consent form
- Please do not use a tablet or a smartphone. These tasks are intended to be completed using a desktop or a laptop
- Please make sure that your browser's window is maximized
- Please don't open new tabs, web sites, or apps during the experiment
- This online experiment is composed of several steps. Please do not try to go back to a previous step/screen while completing the experiment!
- Please be sure that you won't be interrupted while completing each task
- Please be sure that you are 100% concentrated in the task you are performing (for instance, don't listen to music while you are participating in the experiment)
- Please do not use Firefox, since it is incompatible with some visual elements coded (we are currently trying to resolve the issue)
- Give your best!

Thank you!

Study information and generic protocol

ASSESSING MENTAL WORKLOAD IN WEBPAGES AND HUMAN-COMPUTER INTERACTIVE SETTINGS

The concept of human Mental Workload (MWL) has a long history in the fields of ergonomics and psychology, with several applications in the aviation and automobile industry. Although MWL has been under investigation for the last five decades, no clear definition that is universally accepted has emerged. Most of the work concerning MWL was done in the seventies and eighties when the proliferation of computer-based systems was not as extended as it is nowadays. Until the nineties, researches on MWL seemed to conflict in relation to their theories.

Consent Form

ASSESSING MENTAL WORKLOAD IN WEBPAGES AND HUMAN-COMPUTER INTERACTIVE SETTINGS

Contact Details:

Dr. Luca Longo (luca.longo@dit.ie)

School of Computing, Dublin Institute of Technology

Population Growth Task | Unemployment Task

☐ *I have read the Study information and the Consent form and I agree to participate in the study*

Fig. A.1 Welcome Page

A.2 Study Information and Consent Form

Study information and generic protocol

ASSESSING MENTAL WORKLOAD IN WEBPAGES AND HUMAN-COMPUTER INTERACTIVE SETTINGS

The concept of human Mental Workload (MWL) has a long history in the fields of ergonomics and psychology, with several applications in the aviation and automobile industry. Although MWL has been under investigation for the last five decades, no clear definition that is universally accepted has emerged. Most of the work concerning MWL was done in the seventies and eighties when the proliferation of computer-based systems was not as extended as it is nowadays. Until the nineties, researches on MWL seemed to conflict in relation to their theories, definitions, sources, measurement typologies as well as computational modelling techniques. Unfortunately, the situation nowadays is little different, and although several MWL-based applications have emerged in the first decade of the new millennium, these are still based on earlier theories and methodologies. This state-of-the-art research is justified by the fact that defining and modelling MWL is a non-trivial problem. This complexity was earlier acknowledged and researchers felt that no representative measure of mental workload was likely to have a general use. This complexity is also acknowledged by more researchers who, nowadays, still confirm that MWL is difficult to be uniquely defined, due to its multi-faceted and multi-dimensional nature. Despite these discouraging issues, it has been argued, in line with many researchers, that MWL remains an extremely important design concept that would benefit from a significant and challenging re-investigation. Since modern advances in technology have been driving human activity more cognitively oriented and less physical, this re-investigation should be mainly imprinted on the multidisciplinary domain of human-computer interaction. The focus will be on modelling mental workload within fast-growing areas such as the World Wide Web in contrast to traditional application areas such as in aviation, automobile and manufacturing/automation.

In the study under this generic protocol, your mental workload will be assessed while you perform one or two fact-finding tasks.

Mental workload will be assessed by gathering evidence employing two typologies of methods:

- Subjective measures: digital questionnaires using self-report scales and/or
- Primary Task Performance measure: A non-invasive piece of software for gathering human activity over a technological device (computer/mobile/tablet). This activity will include actions such as mouse clicking, scrolling, movement, as well as keyboard usage.

In either case, gathered evidence will be stored in a private database, password protected and will only be accessed by this study's researchers for research purposes. To guarantee the participant's privacy, sensitive personal data such as name, surname, and date of birthday will not be collected. However, some demographical info will be collected for analytical purposes and an email address will be requested to inform the participant of further studies or in case something needs to be communicated. A minimum of 20 responses will be collected for each planned task. The study aims to capture detailed pieces of information and/or detailed performed actions to automatically assess the imposed mental workload imposed on the participant by a given web task. All actions the participant performs over the experiment web site be recorded and saved.

The job of the participant is to perform a computer-based task interacting as natural as possible with the technology provided. The research that is performed under this protocol is conducted in accordance with the ethics guidelines set by Dublin Institute of Technology. The rights of a participant, including the right to withdraw at any point without penalty, are ensured. It is anticipated that the findings of the study conducted under this protocol will be written in the student's thesis report. All results will be anonymised and it will not be possible to identify individual participants' name or email.

For further information please feel free to contact the research supervisor: Dr. Luca Longo (luca.longo@dit.ie)

Frequently Asked Questions

1) <u>Is the study anonymous?</u>

Yes, the studies conducted under this protocol are totally anonymous, collected data will not be linked to the participant's identity.

2) <u>Will the participant's experience while executing a computer-based task be altered by any monitoring technology?</u>

No, the participant's experience while executing a computer-based task will not change. The monitoring technology, if applied, will be completely invisible and non-invasive.

3) <u>Will data the participant inputs online such as logins and passwords be captured and stored somewhere?</u>

No, data entered online such as form input, logins, passwords, or email addresses, will not be recorded.

4) <u>How is the participant's privacy guaranteed?</u>

The participant's personal data will not be stored. Logins, password, and any data entered into forms or over social networks, wikis, blogs, or other resources will never be recorded.

5) <u>Will recorded data be linked to the participant?</u>

No, recorded data will never be linked to the participant. Employed software will randomly generate a code to identify the participant's interactions with provided technologies or devices over time, but this code can never be associated with the participant's personal data, computer IP, or computer MAC address as such information is never stored.

6) Is the captured data stored in a public database?

No, the captured data will be stored in a private password-protected database.

7) <u>Who will have access to stored participants' data, and what about confidentiality?</u>

Only the researcher of the studies conducted under this protocol will have access to your interaction data, exclusively for research purposes. The researcher will never be able to associate any stored data with the identity of a specific participant, as this information is never stored. No one else will have the right to access any stored information.

8) <u>What does the monitoring software look like?</u>

The monitoring software employed in the studies conducted under this protocol is totally transparent. It comes as either a background application – such as a plug-in/add-on for the browsers (Firefox or Chrome) that is installed only once, or a non-invasive and hidden tracking software. Nothing further is required

Consent Form

ASSESSING MENTAL WORKLOAD IN WEBPAGES AND HUMAN-COMPUTER INTERACTIVE SETTINGS

Contact Details:
Dr. Luca Longo (luca.longo@dit.ie)
School of Computing, Dublin Institute of Technology

I consent to participate in this study. I have been informed that the confidentiality of the data I provide will be safeguarded. The electronic data that will be formed of the responses given will be permanently archived by the student/researcher and any sensitive data will be anonymised to prevent my identification.

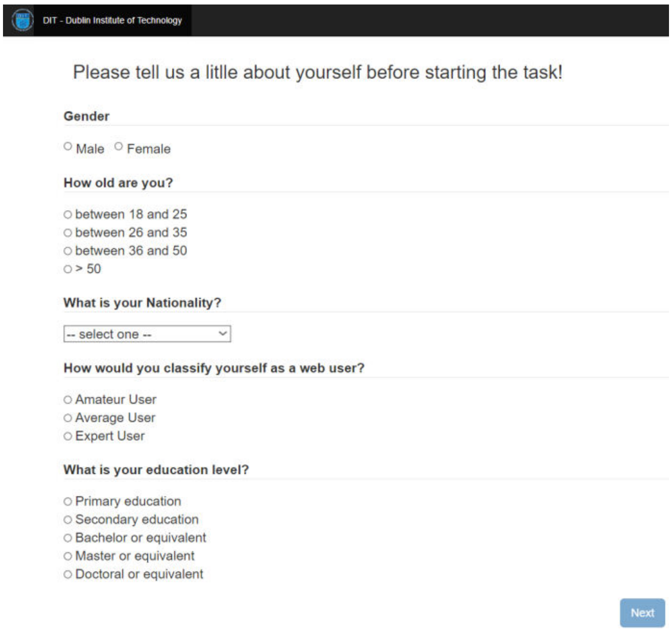
I understand that my behaviour will be gathered or monitored during the experiment and the generated data will be stored for statistical analysis in a password-protected database. This database is exclusively accessible by the student/researcher and it is placed within a password-protected server. Regarding the web task(s) I will be asked to perform ('web-based activity'), I understand that my behaviour will be monitored by means of a piece of software for experimental purposes. This piece of software is aimed at gathering my activity such as clicking, scrolling, mouse movements generated during my interaction with a web-page. The aim of the study is to assess my mental workload imposed by a web-based interactive task. I am free to ask any questions at any time before and after the study. I have not been coerced in any way to participate in this study and I understand that I may terminate my participation in the study at any point should I so wish. I am at least 18 years old.

Data Protection

- I understand that my participation is entirely voluntary, that I may refuse to answer any question and may withdraw at any time without prejudice.
- I agree to Master student Filipe Romero and Dublin Institute of Technology to the storing of any data resulting from this project. I agree to the processing of such data for purposes connected with this research as outlined to me.
- I understand that my participation is fully anonymous, no personal sensitive details will be recorded, no images or video will be stored and all information collected will remain confidential.
- I have been provided with a study information description that outlines the activities I will take part in, how data will be collected and stored and how I can contact the student/researcher.
- I agree that my data is used for scientific purposes and I have no objection that it is reported in a Master Thesis Report in a way that does not reveal my identity.
- I have read this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction.

I have understood the description of the research that is being provided to me.

A.3 Pre-questionnaire – Subject traits



DIT - Dublin Institute of Technology

Please tell us a litlle about yourself before starting the task!

Gender

○ Male ○ Female

How old are you?

○ between 18 and 25
○ between 26 and 35
○ between 36 and 50
○ > 50

What is your Nationality?

-- select one --

How would you classify yourself as a web user?

○ Amateur User
○ Average User
○ Expert User

What is your education level?

○ Primary education
○ Secondary education
○ Bachelor or equivalent
○ Master or equivalent
○ Doctoral or equivalent

Next

Fig. A.2 Pre-Questionnaire

A.4 Instructions

A.3.1 Task1

Instructions - Population Growth

You are going to have access to a dataset showing the population growth, for the past decades, of most territories in the euro zone (source: World DataBank).

At the top of the page you will be provided with a dropdown list (**Select Year:** 1967) which will allow you to choose the year you want to explore. After selecting a year a bar chart will be displayed, where each bar represents a single territory. The territories will be ordered from the territory with the largest Population Growth % Increase to the territory with the largest Population Growth % Decrease.

If you place your mouse pointer over a territory's bar you will be able to see more information regarding population statistics for the territory and for the previously selected year. The following image shows extra info for Ireland that can be visualized by "mousing hover" Ireland's bar:

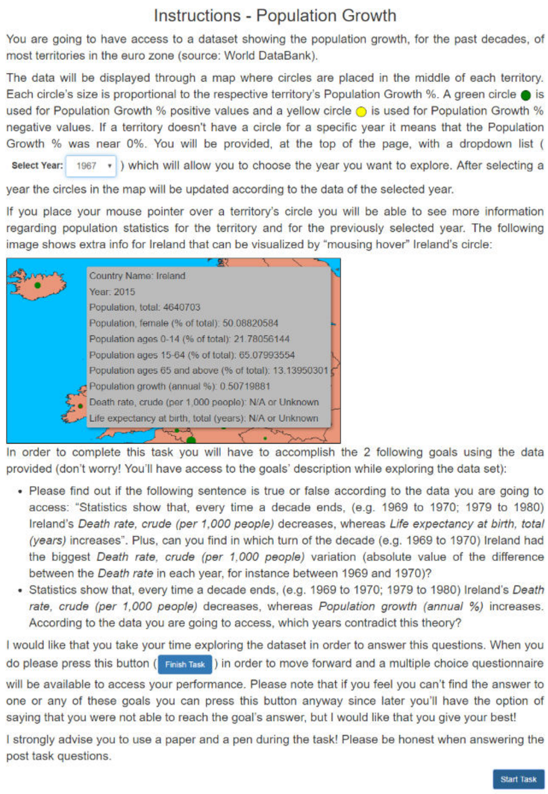
In order to complete this task you will have to accomplish the 2 following goals using the data provided (don't worry! You'll have access to the goals' description while exploring the data set):

- Please find out if the following sentence is true or false according to the data you are going to access: "Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's *Death rate, crude (per 1,000 people)* decreases, whereas *Life expectancy at birth, total (years)* increases". Plus, can you find in which turn of the decade (e.g. 1969 to 1970) Ireland had the biggest *Death rate, crude (per 1,000 people)* variation (absolute value of the difference between the *Death rate* in each year, for instance between 1969 and 1970)?
- Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's *Death rate, crude (per 1,000 people)* decreases, whereas *Population growth (annual %)* increases. According to the data you are going to access, which years contradict this theory?

I would like that you take your time exploring the dataset in order to answer this questions. When you do please press this button (Finish Task) in order to move forward and a multiple choice questionnaire will be available to access your performance. Please note that if you feel you can't find the answer to one or any of these goals you can press this button anyway since later you'll have the option of saying that you were not able to reach the goal's answer, but I would like that you give your best!

I strongly advise you to use a paper and a pen during the task! Please be honest when answering the post task questions.

Start Task



Instructions - Population Growth

You are going to have access to a dataset showing the population growth, for the past decades, of most territories in the euro zone (source: World DataBank).

The data will be displayed through a map where circles are placed in the middle of each territory. Each circle's size is proportional to the respective territory's Population Growth %. A green circle is used for Population Growth % positive values and a yellow circle is used for Population Growth % negative values. If a territory doesn't have a circle for a specific year it means that the Population Growth % was near 0%. You will be provided, at the top of the page, with a dropdown list (**Select Year:** 1967) which will allow you to choose the year you want to explore. After selecting a year the circles in the map will be updated according to the data of the selected year.

If you place your mouse pointer over a territory's circle you will be able to see more information regarding population statistics for the territory and for the previously selected year. The following image shows extra info for Ireland that can be visualized by "mousing hover" Ireland's circle:

In order to complete this task you will have to accomplish the 2 following goals using the data provided (don't worry! You'll have access to the goals' description while exploring the data set):

- Please find out if the following sentence is true or false according to the data you are going to access: "Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's *Death rate, crude (per 1,000 people)* decreases, whereas *Life expectancy at birth, total (years)* increases". Plus, can you find in which turn of the decade (e.g. 1969 to 1970) Ireland had the biggest *Death rate, crude (per 1,000 people)* variation (absolute value of the difference between the *Death rate* in each year, for instance between 1969 and 1970)?
- Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's *Death rate, crude (per 1,000 people)* decreases, whereas *Population growth (annual %)* increases. According to the data you are going to access, which years contradict this theory?

I would like that you take your time exploring the dataset in order to answer this questions. When you do please press this button (Finish Task) in order to move forward and a multiple choice questionnaire will be available to access your performance. Please note that if you feel you can't find the answer to one or any of these goals you can press this button anyway since later you'll have the option of saying that you were not able to reach the goal's answer, but I would like that you give your best!

I strongly advise you to use a paper and a pen during the task! Please be honest when answering the post task questions.

Start Task

Fig. A.3 Instructions - Task 1/Interface 1

Fig. A.4 Instructions - Task 1/Interface 2

A.3.2 Task2

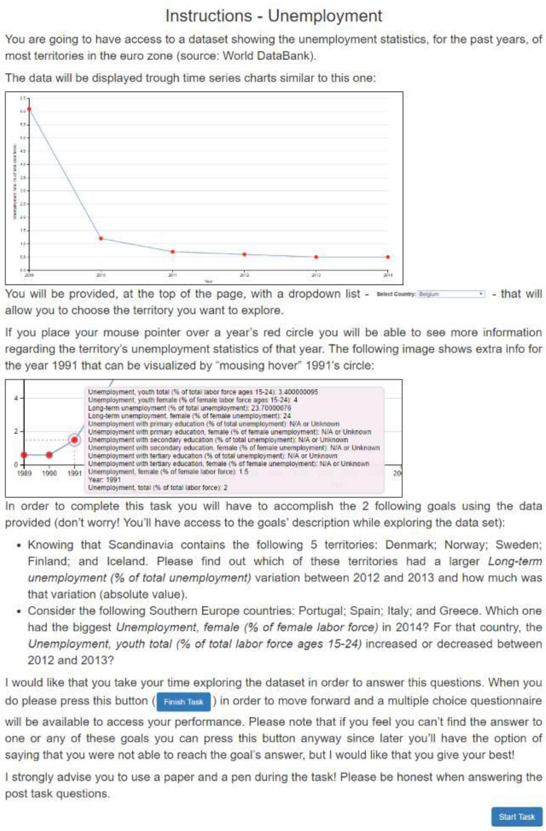


Fig. A.5 Instructions - Task 2/Interface 1

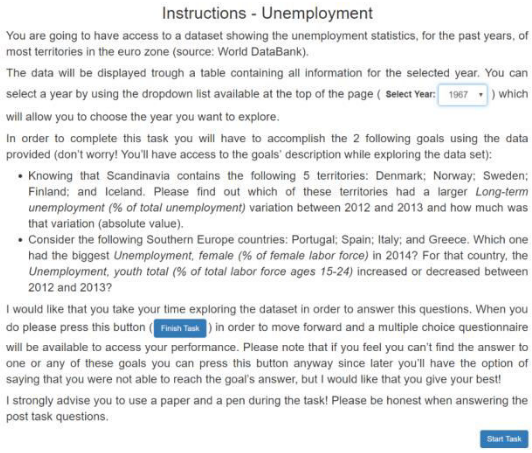


Fig. A.6 Instructions - Task 2/Interface 2

A.4 Tasks and Interfaces

A.4.1 Task1

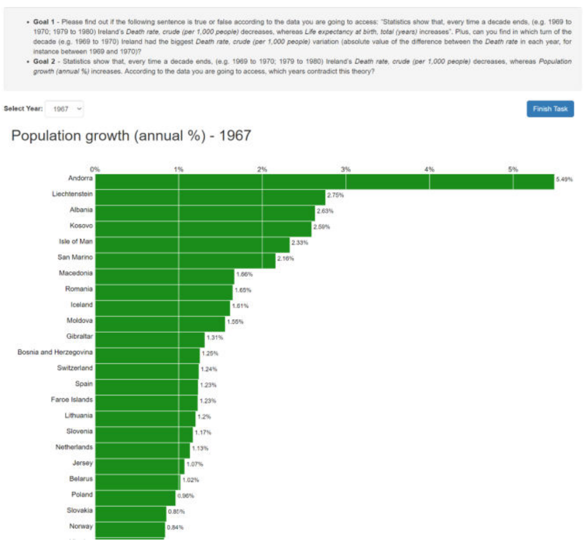


Fig. A.7 Task 1/Interface 1

Fig. A.8 Task 1/Interface 2

A.4.2 Task2

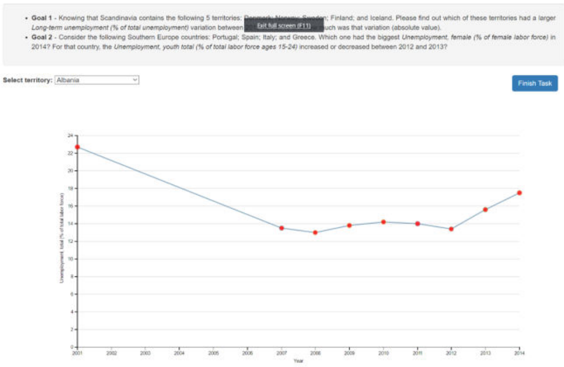


Fig. A.9 Task 2/Interface 1

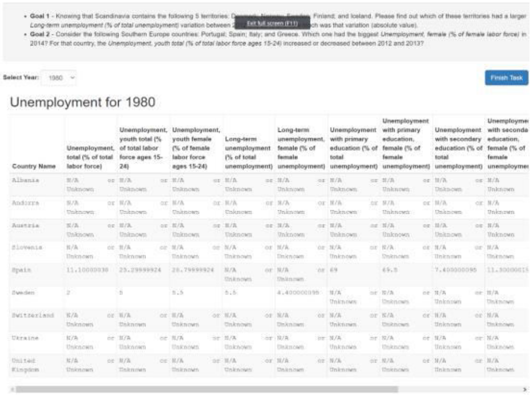
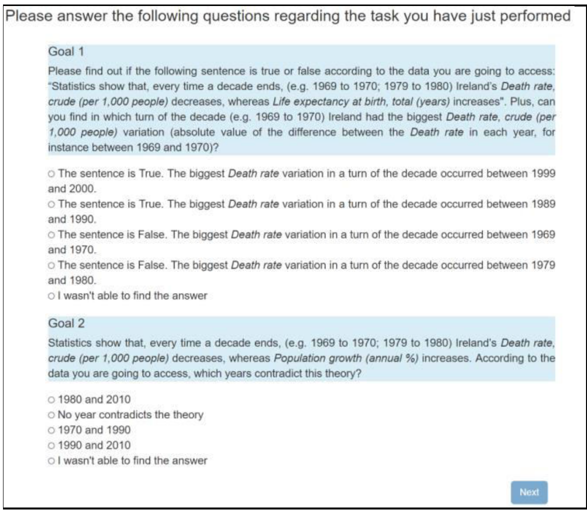


Fig. A.10 Task 2/Interface 2

A.5 Post-Questionnaire – Performance



Please answer the following questions regarding the task you have just performed

Goal 1

Please find out if the following sentence is true or false according to the data you are going to access: "Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's *Death rate, crude (per 1,000 people)* decreases, whereas *Life expectancy at birth, total (years)* increases". Plus, can you find in which turn of the decade (e.g. 1969 to 1970) Ireland had the biggest *Death rate, crude (per 1,000 people)* variation (absolute value of the difference between the *Death rate* in each year, for instance between 1969 and 1970)?

○ The sentence is True. The biggest *Death rate* variation in a turn of the decade occurred between 1999 and 2000.
○ The sentence is True. The biggest *Death rate* variation in a turn of the decade occurred between 1989 and 1990.
○ The sentence is False. The biggest *Death rate* variation in a turn of the decade occurred between 1969 and 1970.
○ The sentence is False. The biggest *Death rate* variation in a turn of the decade occurred between 1979 and 1980.
○ I wasn't able to find the answer

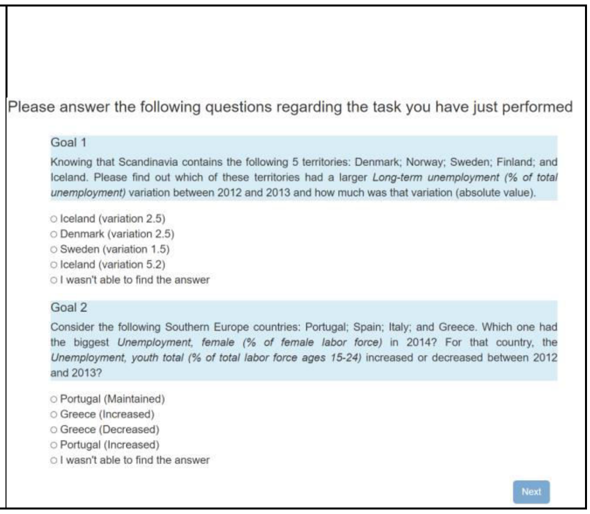
Goal 2

Statistics show that, every time a decade ends, (e.g. 1969 to 1970; 1979 to 1980) Ireland's *Death rate, crude (per 1,000 people)* decreases, whereas *Population growth (annual %)* increases. According to the data you are going to access, which years contradict this theory?

○ 1980 and 2010
○ No year contradicts the theory
○ 1970 and 1990
○ 1990 and 2010
○ I wasn't able to find the answer

Next

Fig. A.11 Task 1



Please answer the following questions regarding the task you have just performed

Goal 1

Knowing that Scandinavia contains the following 5 territories: Denmark; Norway; Sweden; Finland; and Iceland. Please find out which of these territories had a larger *Long-term unemployment (% of total unemployment)* variation between 2012 and 2013 and how much was that variation (absolute value).

○ Iceland (variation 2.5)
○ Denmark (variation 2.5)
○ Sweden (variation 1.5)
○ Iceland (variation 5.2)
○ I wasn't able to find the answer

Goal 2

Consider the following Southern Europe countries: Portugal; Spain; Italy; and Greece. Which one had the biggest *Unemployment, female (% of female labor force)* in 2014? For that country, the *Unemployment, youth total (% of total labor force ages 15-24)* increased or decreased between 2012 and 2013?

○ Portugal (Maintained)
○ Greece (Increased)
○ Greece (Decreased)
○ Portugal (Increased)
○ I wasn't able to find the answer

Next

Fig. A.12 Task 2

A.6 Post-Questionnaire – MWL subjective assessment

NASA Task Load Index

- **Mental Demand** - How much mental and perceptual activity was required by the task you have just executed? In other words, was the task easy or demanding, simple or complex?
- **Physical Demand** - How much physical activity was required by the task? In other words, was the task physically easy or demanding, slack or strenuous?
- **Temporal Demand** - How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow or rapid?
- **Performance** - How successful were you in the task? In other words, how satisfied were you with your level of performance?
- **Effort** - How hard did you have to work (mentally and physically) to accomplish your level of performance?
- **Frustration** - How irritated, stressed and annoyed versus content, relaxed, and complacent did you feel during the task?

Workload Profile

- **Solving and deciding** - How much attention was required by the task for activities like remembering, problem-solving, decision-making and perceiving (e.g. detecting, recognising and identifying objects)?
- **Task and space** - How much attention was required by the task for spatial processing (spatially pay attention around you)?
- **Verbal material** - How much attention was required by the task for verbal material (e.g. reading or processing linguistic material or listening to verbal conversations)?
- **Auditory attention** - How much attention was required for executing the task and executing its activities based on the information auditory received?
- **Speech response** - How much attention was required by the task for producing speech responses (e.g. engaging in conversation or talk or answering questions)?
- **Response selection** - How much attention was required by the task for selecting the proper response channel (manual or speech) and its execution?
- **Visual attention** - How much attention was required for executing the task based on the information visually received?
- **Manual activity** - How much attention was required for manually responding to the elements/activities of the task (writing, drawing, clicking, etc)?

A.7 Dataset attributes

The data collected related with **Population Growth** were the following:

- Population growth (annual %)
- Population, total
- Death rate, crude (per 1,000 people)
- Life expectancy at birth, total (years)
- Population ages 0-14 (% of total)
- Population ages 15-64 (% of total)
- Population ages 65 and above (% of total)
- Population, female (% of total)

The data collected related with **Unemployment** were the following:

- Unemployment, total (% of total labour force)
- Unemployment, female (% of female labour force)
- Unemployment, youth total (% of total labour force ages 15-24)
- Unemployment, youth female (% of female labour force ages 15-24)
- Long-term unemployment (% of total unemployment)
- Long-term unemployment, female (% of female unemployment)
- Unemployment with primary education (% of total unemployment)
- Unemployment with primary education, female (% of female unemployment)
- Unemployment with secondary education (% of total unemployment)
- Unemployment with secondary education, female (% of female unemployment)
- Unemployment with tertiary education (% of total unemployment)
- Unemployment with tertiary education, female (% of female unemployment)

A.8 Technology Used

The web site used to carry on the online experiment was developed from scratch by the student. A mix of technologies was used, as well as Web Frameworks:

- **Front-End** – Typical web programming technologies were used such as Hypertext Mark-up Language (HTML), Cascading Style Sheets (CSS), JavaScript and Scalable Vector Graphics (SVG). Moreover, Bootstrap[1], a front-web framework was used for the most part of the website's design (Welcome page; Instructions; Surveys; etc.).
- **Back End** – A PHP server was used in order to code and place server related processes. The experiment's data was stored by using a MySQL database.

Regarding each task interface, more specific technologies were used. For instance, task 1 interfaces (bar chart and two-dimensional map) were developed using the Data Driven Objects (D3) framework[2], whereas task 2's first interface was developed with the help of DimpleJS framework[3], which is basically a high level API for using D3. Furthermore, the data used to build the two-dimensional map of task 1's second interface was collected in Natural Earth's website[4]. However, the map downloaded in that web site was a worldwide map, and the experiment only needed the Euro Zone map. In order to filter the available data and remove polygons not belonging to the Euro Zone, Geospatial Data Abstraction Library (GDAL)[5] and MS4W (MapServer for

[1] http://getbootstrap.com
[2] https://d3js.org
[3] http://dimplejs.org
[4] http://www.naturalearthdata.com
[5] http://www.gdal.org

Windows) were used. GDAL is a translator library for raster and vector geospatial data formats. Using these tools it was possible to filter the data according with the experiment's needs, as well as converting the data provided by Natural Earth (Shapefile) to data that could be used directly with JavaScript (GeoJSON). Lastly, TopoJSON, an extension of GeoJSON, was used in order to eliminate the redundancy in the final map's data, which allowed a faster loading time when rendering the data in web browsers.

APPENDIX B RESULTS

B.1 Indicators of User Interaction

Table B-1 – Indicators of User Interaction

Type	Desc	All interfaces			Task 1 / Int 1			Task 1 / Int 2			Task 2 / Int 1			Task 2 / Int 2		
		min	avg	max	min	avg	max	min	avg	max	min	avg	max	min	avg	max
Clicks	Number of Clicks	4	35.1	164	7	40	97	6	46.96	164	9	24.32	44	4	27.5	52
	Number of Clicks per Minute	0.53	3.58	13.32	1.06	3	9.05	0.53	5.12	13.32	0.84	2.73	5.69	0.65	3.1	9.54
	Number of void clicks	0	9.83	134	0	10.33	52	0	11	134	0	5.39	28	0	15.79	39
	Number of void clicks per minut	0	0.97	7.69	0	0.93	7.59	0	0.99	7.69	0	0.57	3.25	0	1.8	5.14
	Average Time between Clicks (seconds	4	24.08	103	6	24.33	53	4	16.96	103	7	25.82	73	6	34	95
	Average Time between Void Clicks (seconds)	2	89.95	521	8	125	459	2	110.26	521	16	88.76	223	3	21.64	52
	First Click (seconds)	4	41.22	258	5	37.5	161	5	56.04	258	4	35.43	222	8	29	83
	First Void Click (seconds)	5	161.85	1321	5	159.62	569	17	166.29	525	8	193.67	1321	9	105.54	487
Keyboard	Number of Keys Pressed	0	9.66	343	0	5.4	45	0	23.83	343	0	1.53	27	0	4.56	43
	Number of Keys Pressed	0	1.24	51.71	0	0.45	3.69	0	3.41	51.71	0	0.18	2.37	0	0.3	1.87
	Average Time between Keyboard Used (seconds)	0	15.05	54	0	20	50	1	5.2	10	1	20	54	0	15	30
	First Time Keyboard Was Used (seconds)	1	155.4	940	1	155	543	25	188.67	940	8	135.25	421	17	149.33	361
Scrolling	Number of Scrolling Events	0	14.82	96	0	28	96	0	11.03	61	0	5.47	19	0	22.75	57
	Number of Scrolling Events Per Minute	0	1.49	8.75	0	1.99	6.18	0	1.37	8.75	0	0.75	3.86	0	2.47	4.95
	First Time Scrolling Was Used (seconds)	3	72.67	1469	3	28.6	116	4	36.91	195	3	162.18	1469	4	32.54	101
	Average Duration of Scrolling (seconds)	1	1.92	6	1	3.13	6	1	1.45	4	1	1.18	2	1	2.54	4
	Average Time between Scroll Events (seconds)	6	53.46	608	7	24.67	52	6	73.95	608	14	78.88	295	8	22.31	79
Cursor	Distance travelled with mouse (pixels)	9734	48041	181501	9734	71061	129693	13002	37513	181501	17731	35159	128555	27570	62500	123383
	Distance travelled with mouse per minute (pixels)	1030	4572.3	16877	2264	4983	8167	1136	4026.1	7342	1030	3713.6	6653	1980	6659	16877
	Average time in the same area (seconds)	3	8.42	28	4	9.25	28	3	6.86	18	4	8.5	23	3	10.06	22
	Number of fixations	13	59.16	176	13	71.05	141	15	59.34	143	26	53.5	176	29	54.56	89
	Number of fixations per minute	1.95	5.77	12.11	1.99	5.06	8.89	2.16	6.35	12.11	1.95	5.89	9.84	2.1	5.38	8.96
	Average Fixation Time (seconds)	2	8.6	28	3	9.8	25	2	7.62	23	3	8.57	28	2	8.94	25

B.2 Correlation between NASA-TX and indicators of User Interaction

Table B-2 Correlation between mouse click indicators and NASA-TLX (aggregated indicator and dimensions)

User Interaction	Task/Interface	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Number of clicks	All	-0.17	-0.12	-0.06	-0.15	0.01	-0.1
	1/1	-0.19	-0.41	-0.45	0.11	0.12	-0.13
	1/2	-0.31	-0.11	0.2	-0.3	-0.25	-0.26
	2/1	0.15	0.44	-0.16	-0.23	0.27	0.34
	2/2	-0.25	-0.12	-0.37	0.14	0.03	0.34
Number of clicks Per Minute	All	-0.21	-0.1	-0.22	-0.17	-0.03	-0.04
	1/1	-0.02	-0.22	**-0.5**	-0.03	0.18	-0.02
	1/2	-0.41	-0.16	-0.13	-0.32	-0.28	-0.15
	2/1	-0.22	0.25	-0.35	-0.4	0.04	0.18
	2/2	-0.33	-0.39	**-0.68**	**0.57**	-0.27	0.22
Number of Void clicks	All	0.11	0.1	-0.11	0.08	0.07	0
	1/1	0.28	-0.14	-0.36	0.36	0.27	0.09
	1/2	-0.09	0.07	-0.05	-0.04	-0.23	-0.37
	2/1	0.27	0.39	-0.11	-0.14	0.22	0.28
	2/2	0.01	-0.01	-0.03	0.27	-0.09	0.11
Number of Void clicks Per Minute	All	0.12	0.14	-0.17	0.06	0.07	0.02
	1/1	0.3	-0.18	-0.35	0.36	0.25	0.07
	1/2	0.01	0.22	-0.1	0.02	-0.11	-0.28
	2/1	0.17	0.42	-0.23	-0.26	0.16	0.29
	2/2	-0.29	-0.41	**-0.6**	**0.56**	-0.31	0.26
Average Time Between Clicks (seconds)	All	0.14	0.03	0.21	0.15	-0.02	0.05
	1/1	0.01	0.22	0.48	0.05	-0.13	0.02
	1/2	0.31	0	0.18	0.24	0.16	0.11
	2/1	0.19	-0.26	0.24	0.33	-0.05	-0.09
	2/2	0.32	0.34	**0.68**	**-0.54**	0.21	-0.24
Average Time Between Void Clicks (seconds)	All	-0.26	-0.19	-0.03	-0.16	-0.18	0.02
	1/1	**-0.68**	-0.24	0.07	-0.12	-0.36	-0.31
	1/2	-0.26	-0.39	-0.03	-0.32	-0.3	0.25
	2/1	0	-0.15	-0.15	0.22	-0.15	0.16
	2/2	**0.51**	**0.62**	**0.61**	-0.3	0.19	-0.06
First Click (seconds)	All	0.2	0.15	0.22	-0.12	0.11	-0.06
	1/1	0.27	0.31	0.32	-0.18	0.36	0.01
	1/2	0.07	0.11	0.01	-0.2	-0.11	-0.02
	2/1	0.29	-0.04	0.3	0.36	0.1	-0.06
	2/2	0.48	0.42	0.6	**-0.53**	0.43	0.11
First Void Click (seconds)	All	0.04	-0.05	0.25	-0.19	0.07	-0.07
	1/1	0.25	0.03	0.48	0.05	0.14	-0.04
	1/2	0.01	0.08	0.06	-0.34	0.17	0.25
	2/1	-0.17	-0.41	0.45	0.15	-0.12	-0.28
	2/2	-0.14	-0.05	0.03	-0.55	0.29	0.09

Table B-3 Correlation between keyboard indicators and NASA-TLX (aggregated indicator and dimensions)

User Interaction	Task/Interface	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Number of Keys Pressed	All	-0.15	-0.3	-0.09	0.04	-0.17	-0.06
	1/1	-0.34	**-0.55**	-0.26	0.47	-0.35	-0.13
	1/2	-0.15	-0.32	0.05	0.02	-0.18	-0.13
	2/1	-0.32	-0.2	-0.34	0.07	-0.15	-0.08
	2/2	0.3	0.18	0.22	-0.4	0.16	0.36
Number of Keys Pressed Per Minute	All	-0.16	-0.3	-0.09	0.03	-0.18	-0.06
	1/1	-0.35	**-0.55**	-0.26	0.49	-0.37	-0.13
	1/2	-0.15	-0.32	0.04	0.02	-0.19	-0.14
	2/1	-0.32	-0.2	-0.33	0.08	-0.13	-0.1
	2/2	0.29	0.17	0.22	-0.43	0.15	0.37
Average Time between Keyboard Use (seconds)	All	0.25	0.21	-0.06	0.35	0.13	0.11
	1/1	-0.4	**-0.89**	**-0.67**	0.34	**-0.67**	**0.62**
	1/2	**0.9**	0	0.22	-0.1	**0.5**	**0.71**
	2/1	0.1	-0.31	0.1	**0.87**	0.45	**-0.82**
	2/2	**1**	**0.89**	**0.77**	-0.21	**0.74**	**0.63**
First Time Keyboard Was Used (seconds)	All	0.13	-0.25	0.09	-0.46	0.03	0.21
	1/1	**0.6**	0.22	-0.11	-0.22	**0.87**	0.36
	1/2	0.12	-0.26	-0.31	-0.2	-0.41	0
	2/1	0.24	0.11	-0.1	-0.4	-0.12	**0.65**
	2/2	-0.23	**-0.52**	0.49	**-0.7**	**-0.72**	-0.03

Table B-4 Correlation between scrolling indicators and NASA-TLX (aggregated indicator and dimensions)

User Interaction	Task/Interface	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Number of Scrolling Events	All	0.1	-0.01	-0.01	-0.1	0.05	0.06
	1/1	-0.13	-0.01	-0.25	-0.14	0.23	0.13
	1/2	0.07	-0.03	-0.08	-0.06	-0.06	0.1
	2/1	0.28	0.15	0.08	-0.22	0.18	0.07
	2/2	0.31	-0.06	0.27	-0.05	-0.12	0.03
Number of Scrolling Events Per Minute	All	0.11	0	-0.02	-0.18	0.01	0.08
	1/1	0	0.14	-0.29	-0.25	-0.03	-0.03
	1/2	0.1	0.04	-0.02	-0.13	-0.02	0.27
	2/1	0.26	0.21	0.05	-0.3	0.12	0.06
	2/2	0.03	-0.43	0.12	-0.15	-0.23	0.14
First Time Scrolling Was Used (seconds)	All	-0.06	-0.12	0.16	-0.1	0.03	0
	1/1	0.09	-0.03	0.4	-0.07	-0.01	0.26
	1/2	-0.21	-0.13	0.24	-0.22	-0.08	-0.07
	2/1	-0.11	-0.29	0.21	0.09	0.05	-0.13
	2/2	0.08	-0.06	-0.1	-0.08	0.41	-0.1
Average Duration of Scrolling (seconds)	All	-0.12	-0.14	0.15	-0.02	-0.22	-0.15
	1/1	0.3	**0.73**	0.35	**-0.6**	0.04	0.17
	1/2	-0.2	-0.38	0.19	0.41	-0.25	-0.49
	2/1	-0.27	-0.07	-0.13	0.21	**-0.54**	0
	2/2	-0.06	-0.27	0.44	-0.36	-0.31	-0.37
Average Time between Scroll Events (seconds)	All	0.01	0.02	-0.05	0.37	-0.08	-0.06
	1/1	0	-0.25	0.01	0.42	-0.1	0.05
	1/2	0.03	-0.19	-0.05	0.33	-0.01	-0.27
	2/1	-0.33	-0.13	0.01	**0.69**	**-0.53**	-0.18
	2/2	0.26	**0.51**	0.04	0.32	0.13	0.08

Table B-5 Correlation between mouse cursor indicators and NASA-TLX (aggregated indicator and dimensions)

User Interaction	Task/Interface	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Distance travelled with mouse (pixels)	All	-0.08	-0.13	0.07	-0.08	0.05	0.03
	1/1	-0.38	-0.39	0.17	-0.04	0.12	0.06
	1/2	-0.23	-0.2	0.02	-0.29	-0.28	-0.25
	2/1	0.07	0.12	0.05	0.05	0.04	0.16
	2/2	0.18	0.07	-0.28	0.23	0.27	0.14
Distance travelled with mouse per minute (pixels)	All	-0.11	-0.07	-0.12	-0.36	-0.05	0.09
	1/1	-0.22	-0.16	0.03	-0.28	-0.09	0.13
	1/2	-0.2	-0.03	-0.12	-0.45	-0.22	-0.08
	2/1	0.02	**0.5**	-0.37	-0.48	0.13	0.25
	2/2	-0.15	-0.45	-0.3	0.04	-0.03	0.22
Average time in the same area (seconds)	All	0.07	0.04	0.23	0.24	0.01	-0.14
	1/1	0.23	0.21	0.11	0.22	0.1	-0.12
	1/2	0.17	0.09	0.21	0.36	0.14	0.09
	2/1	0.02	-0.43	0.38	0.4	-0.1	-0.23
	2/2	0.12	0.46	0.39	-0.13	0.11	-0.32
Number of fixations	All	-0.04	-0.12	0.09	-0.08	0.02	0.07
	1/1	-0.33	-0.39	0	0	0.07	0.2
	1/2	-0.2	-0.24	0.08	-0.24	-0.24	-0.11
	2/1	0.21	0.13	0.01	0.02	0.15	0.27
	2/2	**0.53**	0.37	0.22	-0.07	0.26	0.06
Number of fixations per minute	All	-0.08	-0.04	-0.22	-0.38	-0.09	0.19
	1/1	-0.17	-0.2	-0.26	-0.12	-0.28	0.12
	1/2	-0.19	-0.05	-0.08	-0.39	-0.18	0.05
	2/1	0.04	0.41	-0.4	**-0.56**	0.1	0.3
	2/2	-0.14	**-0.65**	-0.08	-0.33	-0.2	0.31
Average Fixation Time (seconds)	All	0.05	0.03	0.21	0.35	0.03	-0.18
	1/1	0.15	0.15	0.28	0.18	0.2	-0.07
	1/2	0.1	-0.08	0.1	0.4	0.07	-0.11
	2/1	-0.07	-0.47	0.38	**0.56**	-0.16	-0.31
	2/2	0.09	**0.57**	0.21	0.03	0.18	-0.29

Table B-6 Correlation between mouse click indicators and NASA-TLX (aggregated indicator and dimensions), for each profile

User Interaction	Profile	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Number of clicks	A	-0.19	-0.25	-0.28	-0.06	-0.08	-0.25
	B	-0.31	-0.39	-0.22	-0.06	0.19	-0.26
	C	**-0.61**	**-0.5**	**-0.61**	-0.28	0.01	-0.3
	D	-0.23	-0.21	-0.26	-0.15	-0.04	-0.24
	E	-0.31	-0.32	-0.27	-0.14	-0.01	-0.31
Number of clicks Per Minute	A	-0.35	-0.26	-0.42	-0.43	-0.1	0.12
	B	-0.35	-0.26	-0.25	-0.06	0.2	-0.29
	C	**-0.61**	-0.46	**-0.64**	-0.25	0.09	-0.24
	D	-0.49	-0.29	-0.45	-0.31	-0.16	-0.08
	E	-0.4	-0.29	-0.25	-0.24	0.02	-0.06
Number of Void clicks	A	0.07	-0.13	0.12	0.48	-0.2	-0.31
	B	0.17	0.04	-0.04	0.18	0.09	0.02
	C	0.25	-0.13	0.27	0.44	-0.09	0.1
	D	0.08	-0.03	0.11	0.3	-0.17	0.01
	E	0.04	-0.13	0.08	0.24	-0.29	-0.22
Number of Void clicks Per Minute	A	0.13	0.01	0.3	0.43	-0.18	-0.18
	B	0.18	0.11	-0.06	0.21	0.08	0.02
	C	0.18	-0.12	0.15	0.43	-0.05	0.14
	D	-0.04	-0.04	0.06	0.26	-0.27	0.02
	E	0.09	-0.02	0.22	0.21	-0.31	-0.1
Average Time Between Clicks (seconds)	A	0.18	0.01	0.34	0.38	-0.08	-0.09
	B	0.26	0.18	0.18	0.1	-0.27	0.3
	C	0.47	0.32	**0.66**	0.2	-0.16	0.22
	D	0.38	0.14	0.4	0.29	0.08	0.09
	E	0.22	0.09	0.17	0.25	-0.15	0.09
Average Time Between Void Clicks (seconds)	A	-0.13	-0.06	-0.13	-0.46	0.08	0.34
	B	-0.47	-0.41	-0.36	-0.22	-0.18	-0.15
	C	-0.4	-0.09	-0.36	**-0.5**	-0.04	-0.24
	D	-0.09	-0.04	-0.07	-0.41	0.12	0.07
	E	-0.06	0.02	-0.24	-0.32	0.3	0.25
First Click (seconds)	A	0.38	0.46	0	0.07	0.41	-0.02
	B	0.29	-0.02	-0.04	0.11	0.25	-0.3
	C	0.35	0.25	0.04	-0.48	0.47	0.18
	D	**0.5**	0.47	0.35	-0.27	**0.51**	-0.01
	E	0.24	0.19	-0.13	-0.22	0.35	-0.13
First Void Click (seconds)	A	-0.05	-0.08	-0.43	-0.34	0.2	-0.01
	B	-0.18	-0.37	-0.19	-0.48	-0.14	0.01
	C	0.07	-0.01	-0.23	**-0.5**	0.31	0.23
	D	0.34	0.16	0.07	-0.37	0.48	0.16
	E	-0.08	-0.18	-0.36	-0.37	0.21	-0.12

Table B-7 Correlation between keyboard indicators and NASA-TLX (aggregated indicator and dimensions), for each profile

User Interaction	Profile	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Number of Keys Pressed	A	-0.37	-0.41	-0.06	0.08	-0.28	-0.17
	B	0.02	-0.38	0.07	-0.33	0.04	0.22
	C	-0.12	-0.19	-0.11	0.16	-0.15	0.02
	D	-0.27	-0.29	-0.12	0.14	-0.24	-0.13
	E	-0.25	-0.27	-0.11	0.04	-0.09	-0.13
Number of Keys Pressed Per Minute	A	-0.4	-0.43	-0.08	0.09	-0.32	-0.18
	B	0.02	-0.38	0.08	-0.33	0.03	0.23
	C	-0.1	-0.17	-0.1	0.16	-0.14	0.04
	D	-0.28	-0.31	-0.13	0.15	-0.26	-0.14
	E	-0.27	-0.29	-0.12	0.05	-0.12	-0.13
Average Time between Keyboard Use (seconds)	A	0.2	**0.77**	0.13	**0.82**	-0.09	-0.28
	B	0.21	**0.73**	-0.3	**0.65**	**0.82**	-0.06
	C	-0.07	0.33	**-0.53**	**0.64**	0.22	-0.29
	D	0.16	0.12	-0.34	**0.86**	0.04	-0.25
	E	0.2	**0.84**	0.09	0.36	0.29	0
First Time Keyboard Was Used (seconds)	A	-0.18	**-0.53**	0.14	**-0.52**	0.3	0.23
	B	-0.23	-0.27	**-0.57**	-0.37	-0.33	-0.02
	C	**0.54**	-0.11	0.13	**-0.55**	**0.63**	**0.71**
	D	0.18	-0.28	0.16	-0.35	0.41	0.48
	E	0.15	-0.38	-0.14	**-0.54**	0.3	0.36

Table B-8 Correlation between scrolling indicators and NASA-TLX (aggregated indicator and dimensions), for each profile

User Interaction	Profile	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Number of Scrolling Events	A	-0.07	-0.18	-0.02	-0.12	-0.17	0
	B	-0.16	-0.41	0.11	-0.17	0.12	-0.22
	C	-0.01	-0.4	-0.09	-0.12	-0.18	0.3
	D	0.09	-0.14	0.03	-0.14	0	0.18
	E	0.08	-0.21	-0.01	-0.15	0.07	-0.07
Number of Scrolling Events Per Minute	A	-0.04	-0.11	0.04	-0.33	-0.23	0.09
	B	-0.37	**-0.52**	0.14	-0.38	-0.05	-0.23
	C	0.08	-0.28	-0.09	-0.26	-0.21	0.22
	D	0.1	-0.08	0	-0.28	-0.05	0.19
	E	0.18	-0.1	0.06	-0.29	-0.03	0
First Time Scrolling Was Used (seconds)	A	-0.11	-0.16	-0.14	-0.18	-0.03	0.23
	B	-0.2	-0.19	0.06	-0.12	0.02	-0.17
	C	-0.06	0	-0.14	0.2	0.27	-0.05
	D	0.05	-0.07	-0.08	-0.05	0.12	0.3
	E	0.06	-0.14	0.16	-0.26	0.13	0.23
Average Duration of Scrolling (seconds)	A	-0.17	-0.28	0.13	-0.18	-0.25	-0.27
	B	-0.29	-0.22	0.19	0.04	-0.3	-0.06
	C	0.33	0.09	0.39	-0.28	-0.29	0.08
	D	0.04	-0.22	0.2	-0.05	-0.15	-0.22
	E	0.06	-0.13	0.25	-0.03	-0.07	-0.26
Average Time between Scroll Events (seconds)	A	0.21	0.31	0.01	0.4	0.25	0.04
	B	0.39	0.47	-0.22	0.44	0.12	0.14
	C	0.04	0.15	-0.12	0.45	0.1	-0.11
	D	0.04	0.14	0.08	0.3	0.04	-0.03
	E	-0.1	0.18	-0.17	0.34	0.01	0.04

Table B-9 Correlation between mouse cursor indicators and NASA-TLX (aggregated indicator and dimensions), for each profile

User Interaction	Profile	TLX	Mental Demand	Temporal Demand	Performance	Effort	Frustration
Distance travelled with mouse (pixels)	A	-0.12	-0.2	-0.28	-0.15	0.14	0.12
	B	-0.2	-0.2	-0.21	-0.2	0.01	0.2
	C	-0.05	-0.24	-0.04	-0.03	0.09	0.33
	D	0.11	-0.05	-0.06	-0.05	0.22	0.12
	E	-0.01	-0.19	-0.12	-0.14	0.16	0.11
Distance travelled with mouse per minute (pixels)	A	-0.21	-0.14	-0.18	**-0.52**	-0.12	0.17
	B	-0.43	-0.35	-0.03	**-0.56**	-0.14	0.07
	C	-0.03	-0.09	-0.11	-0.2	0.08	0.14
	D	-0.09	-0.04	-0.24	-0.32	-0.05	0.04
	E	-0.05	-0.15	-0.08	-0.44	-0.1	0.17
Average time in the same area (seconds)	A	-0.11	-0.06	0.27	0.23	-0.19	-0.38
	B	0.32	0.25	0.39	0.44	0.02	-0.17
	C	0.46	0.34	**0.6**	0.08	-0.03	0.13
	D	0.2	0.09	0.39	0.28	-0.02	-0.07
	E	0.01	0.19	0.16	0.34	-0.06	-0.16
Number of fixations	A	-0.04	-0.2	-0.29	-0.01	0.15	0.2
	B	-0.12	-0.32	-0.19	-0.14	0.01	0.1
	C	-0.16	-0.26	-0.21	-0.15	0.02	0.23
	D	0.21	-0.03	0.04	-0.13	0.27	0.19
	E	-0.04	-0.24	-0.19	-0.17	0.21	0.15
Number of fixations per minute	A	-0.04	-0.1	-0.23	-0.42	-0.02	0.35
	B	-0.4	-0.46	-0.06	**-0.63**	-0.23	-0.02
	C	-0.32	-0.4	**-0.56**	-0.18	-0.07	0.05
	D	-0.12	-0.12	-0.33	-0.4	-0.07	0.23
	E	-0.03	-0.19	-0.15	-0.42	-0.02	0.25
Average Fixation Time (seconds)	A	-0.04	0.04	0.12	0.46	-0.06	-0.41
	B	0.43	0.48	0.09	**0.59**	0.17	0.08
	C	0.29	0.44	**0.58**	0.14	0.02	-0.07
	D	0.11	0.1	0.35	0.38	0.03	-0.23
	E	-0.06	0.12	0.08	0.44	-0.03	-0.3

B.3 Correlation between WP and indicators of User Interaction

Table B-10 Correlation between mouse click indicators and Workload Profile (aggregated indicator and dimensions)

User Interaction	Task Interface	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Number of clicks	All	-0.01	0.1	-0.06	0.05	-0.31	-0.14
	1/1	-0.14	-0.04	-0.44	0.21	-0.16	-0.35
	1/2	-0.06	-0.13	0.05	0.06	**-0.57**	-0.17
	2/1	0.2	0.29	-0.12	-0.15	-0.18	-0.1
	2/2	0.39	0.3	-0.25	0.35	-0.2	0.12
Number of clicks Per Minute	All	-0.01	0.16	-0.03	0.02	-0.22	-0.13
	1/1	-0.08	0.33	-0.22	0.31	-0.14	-0.37
	1/2	0.09	0.05	0.11	0.23	-0.33	0.01
	2/1	0.09	0.05	-0.18	-0.28	-0.05	-0.25
	2/2	-0.19	0.05	-0.46	-0.19	-0.38	-0.24
Number of Void clicks	All	0.35	0.13	-0.08	0.16	0	0.27
	1/1	0.18	0.08	-0.04	0.09	0.05	0.18
	1/2	0.28	-0.03	-0.22	0.06	-0.13	0.3
	2/1	0.42	0.24	0.01	0.05	0.01	0.27
	2/2	0.16	-0.09	-0.43	0.17	-0.1	0.44
Number of Void clicks Per Minute	All	0.37	0.2	-0.04	0.14	-0.01	0.27
	1/1	0.17	0.19	-0.04	0.13	0.04	0.16
	1/2	0.41	0.12	-0.06	0.14	-0.05	0.4
	2/1	0.41	0.16	-0.09	0.05	-0.02	0.18
	2/2	-0.17	0.04	-0.43	-0.18	-0.43	-0.16
Average Time Between Clicks (seconds)	All	-0.04	-0.18	-0.02	-0.06	0.2	0.1
	1/1	0.04	-0.27	0.19	-0.33	0.16	0.39
	1/2	-0.21	-0.14	-0.24	-0.33	0.33	-0.09
	2/1	-0.1	-0.04	0.11	0.21	-0.01	0.23
	2/2	0.15	-0.12	0.4	0.15	0.35	0.22
Average Time Between Void Clicks (seconds)	All	-0.46	-0.3	-0.14	-0.32	0.04	-0.27
	1/1	-0.48	**-0.56**	-0.27	-0.3	0.29	-0.29
	1/2	**-0.62**	-0.46	-0.07	-0.37	0.12	-0.46
	2/1	-0.31	0.07	-0.14	-0.01	-0.03	-0.05
	2/2	0.42	0.23	0.34	0.24	0.35	0.23
First Click (seconds)	All	0.05	-0.07	0.05	-0.04	0.15	0.13
	1/1	0.47	-0.05	0.1	0.28	0.19	0.24
	1/2	-0.15	-0.25	-0.18	-0.37	0.13	-0.01
	2/1	0.04	0.01	0.16	0.13	0.17	0.17
	2/2	0.47	0.37	**0.55**	**0.53**	0.19	0.35
First Void Click (seconds)	All	-0.16	-0.1	-0.06	-0.1	-0.14	-0.19
	1/1	0.11	0.29	-0.23	0.03	-0.41	-0.44
	1/2	-0.04	-0.17	0.06	-0.09	-0.27	-0.12
	2/1	-0.34	-0.36	-0.11	-0.05	0.03	-0.11
	2/2	0.26	0.09	0.41	0.34	0.42	0.2

Table B-11 Correlation between keyboard indicators and Workload Profile (aggregated indicator and dimensions)

User Interaction	Task Interface	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Number of Keys Pressed	All	-0.05	-0.04	0.01	-0.05	0.03	-0.05
	1/1	-0.46	-0.42	-0.45	-0.35	-0.21	-0.23
	1/2	0.27	0.15	0.02	0.25	0.17	0.17
	2/1	-0.23	0.02	-0.12	-0.21	-0.07	-0.12
	2/2	0.31	0.06	**0.76**	0.04	0.24	-0.18
Number of Keys Pressed Per Minute	All	-0.06	-0.05	0.01	-0.06	0.03	-0.06
	1/1	-0.47	-0.43	-0.45	-0.37	-0.21	-0.22
	1/2	0.27	0.16	0.01	0.25	0.18	0.18
	2/1	-0.23	0.02	-0.11	-0.22	-0.06	-0.12
	2/2	0.3	0.06	**0.74**	0.03	0.17	-0.22
Average Time between Keyboard Use (seconds)	All	0.08	0.22	0.04	0.08	-0.13	0.04
	1/1	-0.4	-0.2	0.36	**-0.97**	**0.62**	0.3
	1/2	**0.6**	-0.21	0.1	-0.36	0.22	0
	2/1	**0.7**	0.46	0.05	**0.87**	-0.1	**0.67**
	2/2	**1**	**0.6**	0.26	**0.8**	-0.32	-0.32
First Time Keyboard Was Used (seconds)	All	0.06	-0.21	-0.06	-0.23	0.16	-0.09
	1/1	**0.6**	**0.8**	0.1	**0.56**	0.36	-0.2
	1/2	0.17	-0.31	**-0.52**	**-0.63**	0.07	0.16
	2/1	-0.21	-0.23	0.04	-0.22	0.39	-0.12
	2/2	-0.26	**-0.88**	**-0.51**	-0.2	-0.27	-0.34

Table B-12 Correlation between scrolling indicators and Workload Profile (aggregated indicator and dimensions)

User Interaction	Task Interface	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Number of Scrolling Events	All	0.18	-0.12	-0.02	0.03	0.11	0.11
	1/1	0.07	-0.11	-0.29	0.18	-0.14	-0.11
	1/2	-0.04	**-0.5**	-0.15	-0.29	0.07	0.23
	2/1	0.37	-0.04	0.01	0.06	0.46	0.41
	2/2	0.25	-0.02	0	0.13	0.19	-0.07
Number of Scrolling Events Per Minute	All	0.17	-0.15	-0.04	-0.03	0.08	0.07
	1/1	0.25	-0.13	-0.11	0.31	-0.3	-0.3
	1/2	-0.11	**-0.51**	-0.15	-0.32	0.12	0.17
	2/1	0.35	0.01	-0.01	-0.01	0.47	0.41
	2/2	0.02	0.07	-0.29	-0.2	-0.14	-0.16
First Time Scrolling Was Used (seconds)	All	-0.14	-0.08	0.02	-0.09	-0.13	-0.08
	1/1	0.13	-0.14	0.1	-0.16	-0.03	0.35
	1/2	-0.27	-0.32	-0.36	-0.06	-0.04	-0.23
	2/1	-0.32	0.02	0.44	-0.18	-0.26	-0.19
	2/2	0.09	0.14	-0.15	0.36	-0.13	-0.16
Average Duration of Scrolling (seconds)	All	0.02	-0.01	0.05	0.21	-0.18	-0.2
	1/1	0.32	0.27	0.15	0.16	-0.08	-0.22
	1/2	-0.38	-0.08	0.02	-0.11	-0.41	-0.21
	2/1	0	-0.23	-0.44	0.25	0.03	0.12
	2/2	-0.2	-0.1	0.03	-0.24	-0.3	-0.42
Average Time between Scroll Events (seconds)	All	-0.14	0.2	0.12	0	0.06	0.06
	1/1	-0.2	-0.19	0.3	-0.43	0.35	0.44
	1/2	-0.13	**0.57**	0.43	0.23	-0.11	-0.31
	2/1	-0.36	0.25	-0.33	0.37	-0.34	-0.29
	2/2	0.2	0.14	**0.52**	0.22	0.22	0.32

Table B-13 Correlation between mouse cursor indicators and Workload Profile (aggregated indicator and dimensions)

User Interaction	Task Interface	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Distance travelled with mouse (pixels)	All	0.08	0.01	-0.01	0.14	-0.05	0
	1/1	-0.11	-0.24	-0.35	0.13	0.06	-0.01
	1/2	-0.11	-0.14	0	-0.12	-0.27	-0.17
	2/1	0	-0.08	-0.24	0.05	-0.25	-0.05
	2/2	0.34	0.15	-0.09	0.35	0.17	0.14
Distance travelled with mouse per minute (pixels)	All	0.11	0.03	-0.08	-0.01	0	-0.03
	1/1	0.19	-0.1	-0.02	0.37	0.03	-0.04
	1/2	0.05	-0.04	-0.19	-0.18	0.04	-0.06
	2/1	0.1	0.2	-0.13	-0.46	0.11	0.06
	2/2	-0.01	0.11	-0.32	0.04	-0.13	-0.05
Average time in the same area (seconds)	All	-0.02	0.02	0.21	0.15	-0.01	0.04
	1/1	-0.16	0.09	0.15	-0.34	0.03	0.1
	1/2	0.06	0.12	0.22	0.26	-0.09	0.09
	2/1	-0.07	-0.14	0.12	0.4	-0.08	-0.02
	2/2	-0.06	-0.07	0.29	0	0.16	0.07
Number of fixations	All	-0.06	-0.08	0.01	-0.02	-0.1	-0.02
	1/1	-0.11	-0.24	-0.38	-0.04	0.08	-0.01
	1/2	-0.16	-0.23	0.02	-0.19	-0.23	-0.04
	2/1	-0.05	0.03	0.02	-0.08	-0.34	-0.12
	2/2	0.46	0.01	0.28	0.25	0.42	0.22
Number of fixations per minute	All	-0.09	-0.08	-0.1	-0.31	-0.06	-0.07
	1/1	0.02	-0.17	-0.13	0.07	-0.12	-0.28
	1/2	-0.21	-0.2	-0.1	-0.31	0	-0.03
	2/1	-0.05	0.13	0.02	**-0.61**	-0.02	-0.05
	2/2	0.04	-0.02	-0.13	-0.12	-0.12	-0.16
Average Fixation Time (seconds)	All	-0.01	0.04	0.13	0.23	0.02	0.01
	1/1	-0.18	0.09	0.01	-0.2	0.07	0.16
	1/2	0.08	0.16	0.1	0.28	-0.05	-0.06
	2/1	-0.1	-0.15	0.04	**0.58**	-0.05	-0.07
	2/2	0.04	-0.01	0.34	0.12	0.16	0.1

Table B-14 Correlation between mouse clicks indicators and Workload Profile (aggregated indicator and dimensions), for each profile

User Interaction	Profile	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Number of clicks	A	**-0.54**	-0.11	**-0.56**	0.14	-0.46	**-0.56**
	B	-0.42	0.05	-0.49	-0.18	-0.31	**-0.52**
	C	-0.4	0.02	-0.35	-0.35	-0.2	**-0.52**
	D	-0.27	0.13	-0.32	0.07	-0.32	**-0.54**
	E	-0.35	-0.08	-0.41	-0.07	-0.13	-0.4
Number of clicks Per Minute	A	-0.47	-0.39	-0.47	-0.1	-0.2	-0.34
	B	**-0.67**	-0.22	**-0.51**	-0.2	-0.34	**-0.58**
	C	-0.09	0.31	-0.07	-0.17	-0.18	-0.34
	D	-0.09	0	-0.17	0.02	-0.15	-0.31
	E	**-0.52**	-0.34	-0.28	-0.08	-0.08	-0.33
Number of Void clicks	A	-0.33	0.02	-0.44	-0.04	-0.1	-0.16
	B	0.35	0.34	-0.35	-0.18	0.49	0.3
	C	0.1	0.08	0.12	0.07	0.27	0.27
	D	0.09	0.16	-0.06	0.21	0.07	0.01
	E	0	0.09	-0.37	-0.14	0.13	0.09
Number of Void clicks Per Minute	A	-0.28	-0.04	-0.39	-0.1	0.15	-0.09
	B	0.28	0.25	-0.4	-0.19	0.48	0.27
	C	0.27	0.29	0.25	0.13	0.12	0.28
	D	0.13	0.1	-0.07	0.22	0.06	0.02
	E	-0.02	0.02	-0.34	-0.07	0.21	0.11
Average Time Between Clicks (seconds)	A	0.2	0.36	0.28	-0.04	0.12	0.22
	B	**0.6**	0.25	**0.5**	0.25	0.32	0.56
	C	-0.05	-0.39	-0.06	0.08	0.19	0.23
	D	-0.04	0	0.05	-0.09	0.12	0.24
	E	0.31	0.33	0.08	-0.01	0.03	0.25
Average Time Between Void Clicks (seconds)	A	0.09	0.02	0.2	-0.13	-0.17	0.09
	B	**-0.54**	-0.24	-0.14	-0.03	**-0.57**	**-0.6**
	C	-0.45	-0.26	-0.37	**-0.52**	-0.09	-0.29
	D	-0.34	-0.17	-0.02	-0.41	-0.07	0
	E	0.01	0.12	0.21	-0.23	-0.03	0.05
First Click (seconds)	A	0.32	0.18	0.29	0.09	0.11	0.14
	B	-0.03	-0.11	-0.14	**-0.57**	0.26	-0.03
	C	0.19	0.08	0.17	0.1	**0.5**	0.15
	D	0.2	0.21	0.24	0.17	0.31	0.09
	E	0.45	0.28	0.28	-0.18	0.37	0.19
First Void Click (seconds)	A	0.04	-0.05	0.12	0.26	-0.06	-0.29
	B	-0.43	-0.47	-0.2	-0.42	-0.3	-0.45
	C	0.05	-0.06	0.03	0.03	0.3	-0.1
	D	0.06	0.05	-0.06	0.14	-0.02	-0.23
	E	0.15	-0.04	0.09	0.01	0.04	-0.14

Table B-15 Correlation between keyboard indicators and Workload Profile (aggregated indicator and dimensions), for each profile

User Interaction	Profile	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Number of Keys Pressed	A	-0.25	-0.25	-0.03	-0.02	-0.23	-0.41
	B	0.19	-0.07	0.23	-0.15	0.09	-0.02
	C	-0.25	-0.05	-0.13	-0.29	-0.27	-0.27
	D	-0.32	-0.25	-0.02	-0.17	-0.08	-0.18
	E	-0.02	-0.08	0.01	-0.03	-0.16	-0.3
Number of Keys Pressed Per Minute	A	-0.27	-0.28	-0.05	-0.05	-0.26	-0.43
	B	0.18	-0.07	0.25	-0.16	0.08	-0.04
	C	-0.23	-0.04	-0.11	-0.27	-0.28	-0.27
	D	-0.34	-0.27	-0.03	-0.19	-0.09	-0.19
	E	-0.04	-0.11	0	-0.05	-0.2	-0.33
Average Time between Keyboard Use (seconds)	A	-0.11	0.33	-0.06	-0.09	-0.28	-0.12
	B	0.36	**0.8**	0.05	**0.55**	0.29	0.27
	C	0.25	0.42	0.16	0.22	-0.31	0.46
	D	0.23	0.34	0.07	0.2	0.05	0.49
	E	0.23	**0.56**	0.16	0.42	-0.46	-0.32
First Time Keyboard Was Used (seconds)	A	0.19	-0.11	0.08	-0.04	0.3	-0.17
	B	-0.18	-0.42	**-0.78**	-0.36	-0.15	-0.2
	C	0.39	0.09	0.35	0.22	**0.52**	-0.32
	D	0.16	-0.04	0.21	-0.09	**0.55**	-0.17
	E	0.28	-0.05	-0.02	-0.14	0.41	-0.1

Table B-16 Correlation between scrolling indicators and Workload Profile (aggregated indicator and dimensions), for each profile

User Interaction	Profile	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Number of Scrolling Events	A	0.08	-0.27	-0.22	0.16	-0.17	-0.03
	B	0.3	0.02	0.11	0.01	0.14	0
	C	-0.24	-0.21	-0.04	**-0.5**	-0.02	-0.25
	D	-0.03	-0.19	-0.24	-0.04	-0.04	-0.11
	E	0.21	-0.19	-0.09	0.02	0.05	0.06
Number of Scrolling Events Per Minute	A	0.08	-0.34	-0.26	0.03	-0.29	-0.06
	B	0.12	-0.15	0	-0.18	-0.05	-0.11
	C	-0.08	-0.23	0.06	-0.42	-0.12	-0.34
	D	-0.01	-0.3	-0.31	-0.06	-0.11	-0.13
	E	0.19	-0.34	-0.1	0.01	-0.09	-0.01
First Time Scrolling Was Used (seconds)	A	-0.17	0.08	-0.27	-0.31	-0.04	0.29
	B	-0.18	-0.46	-0.03	-0.07	-0.19	-0.02
	C	0.26	0.49	0.23	0.24	0.21	0.14
	D	-0.03	0.2	0.02	-0.11	-0.06	0.07
	E	-0.16	0.15	0.02	-0.22	-0.23	-0.01
Average Duration of Scrolling (seconds)	A	0.21	-0.14	0.06	0.21	-0.25	-0.2
	B	-0.07	0.07	0.27	0.03	-0.29	-0.11
	C	0.2	-0.34	0.2	0.23	-0.28	-0.45
	D	0.19	0	-0.02	0.28	-0.23	-0.32
	E	0.26	0.04	0.22	**0.5**	-0.28	-0.31
Average Time between Scroll Events (seconds)	A	-0.12	0.46	0.19	-0.02	0.19	-0.04
	B	0	0.33	-0.04	0.01	0.28	0.16
	C	-0.05	0.26	-0.01	-0.22	0.18	**0.51**
	D	-0.09	0.35	0.31	-0.05	0.19	0.19
	E	-0.03	0.46	0.03	0	0.22	0.18

Table B-17 Correlation between mouse cursor indicators and Workload Profile (aggregated indicator and dimensions), for each profile

User Interaction	Profile	WP	Solving Deciding	Verbal Material	Response selection	Visual attention	Manual activity
Distance travelled with mouse (pixels)	A	-0.01	0.04	-0.16	0.14	-0.09	0.05
	B	0.19	-0.02	-0.08	0.31	0.05	-0.04
	C	-0.23	0.03	0.06	-0.22	-0.01	-0.09
	D	0	0.24	-0.08	0	-0.04	-0.01
	E	0.09	0.03	-0.14	0.14	-0.06	0.14
Distance travelled with mouse per minute (pixels)	A	-0.15	-0.2	-0.32	-0.13	-0.28	-0.16
	B	0.05	-0.33	-0.25	0.14	-0.29	-0.13
	C	0.05	0.13	0.25	-0.03	-0.27	-0.21
	D	-0.04	-0.05	-0.29	-0.04	-0.24	-0.15
	E	-0.09	-0.32	-0.35	-0.03	-0.27	-0.08
Average time in the same area (seconds)	A	0.23	0.12	0.42	0.22	0.19	0.03
	B	0.09	0.41	**0.52**	0.05	0.17	0.2
	C	0.16	-0.24	-0.05	0.26	0.21	0.21
	D	0.13	0.07	0.37	0.14	0.24	0.09
	E	0.13	0.31	0.37	0.24	0.07	-0.09
Number of fixations	A	-0.11	0	-0.17	0.08	-0.08	0.04
	B	0.04	-0.1	0	0.15	0.1	-0.19
	C	-0.46	-0.06	0	**-0.62**	0.19	-0.07
	D	-0.15	0.12	0.01	-0.18	-0.01	0.11
	E	0.06	0.02	-0.08	-0.02	0.04	0.15
Number of fixations per minute	A	-0.28	-0.15	-0.4	-0.27	-0.2	-0.07
	B	-0.26	**-0.51**	-0.17	-0.13	-0.11	-0.32
	C	-0.27	0.01	0.17	**-0.59**	0.02	-0.17
	D	-0.29	-0.19	-0.32	-0.38	-0.19	0
	E	-0.22	-0.32	-0.34	-0.25	-0.15	-0.04
Average Fixation Time (seconds)	A	0.16	0.07	0.35	0.22	0.22	0.03
	B	0.09	0.42	0.23	0.02	0.11	0.21
	C	0.14	-0.1	-0.17	0.44	0.04	0.12
	D	0.15	0.1	0.33	0.29	0.2	0.01
	E	0.11	0.26	0.34	0.19	0.1	-0.07

B.4 Scatter Plots – MWL vs. Performance

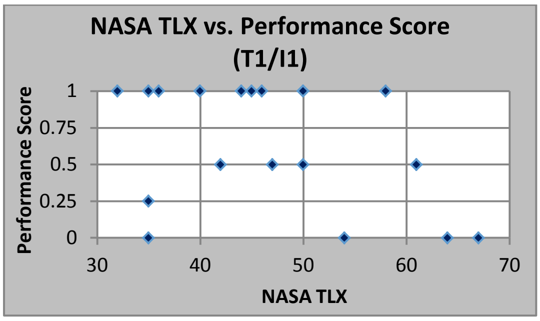


Fig. B.1 NASA-TLX vs. Performance (T1/I1)

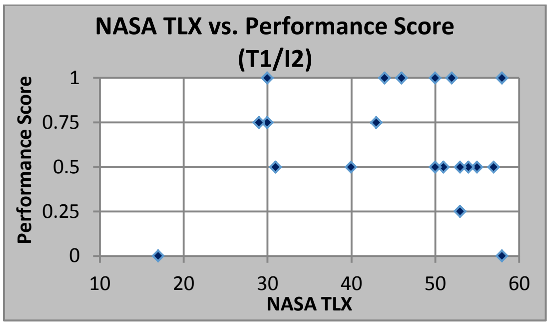


Fig. B.2 NASA-TLX vs. Performance (T1/I2)

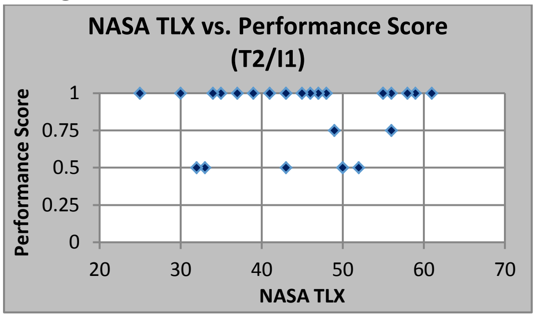


Fig. B.3 NASA-TLX vs. Performance (T2/I1)

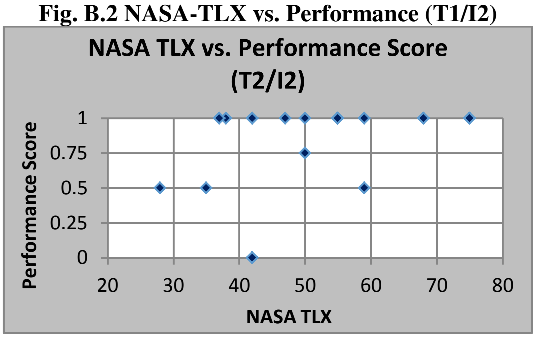


Fig. B.4 NASA-TLX vs. Performance (T2/I2)

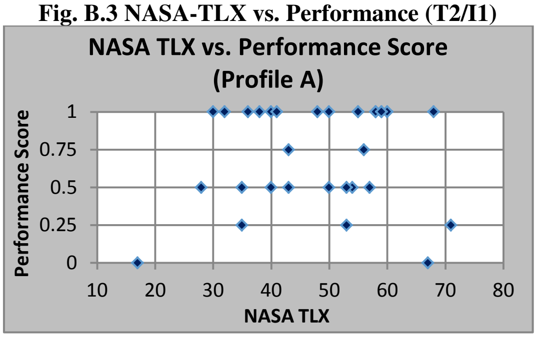


Fig. B.5 NASA-TLX vs. Performance (Profile A)

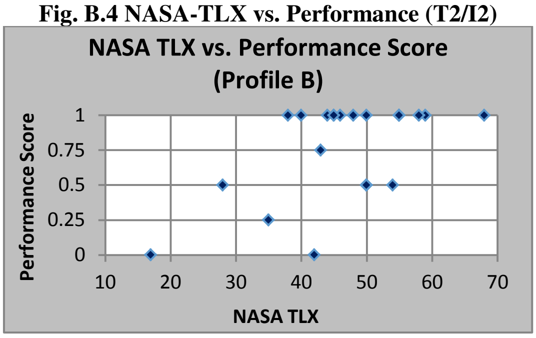


Fig. B.6 NASA-TLX vs. Performance (Profile B)

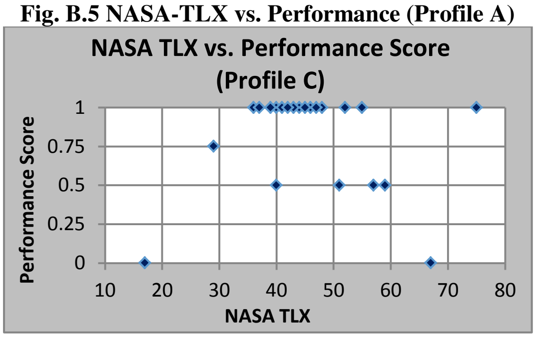


Fig. B.7 NASA-TLX vs. Performance (Profile C)

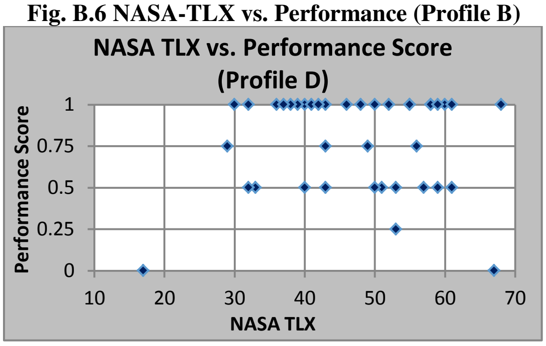


Fig. B.8 NASA-TLX vs. Performance (Profile D)

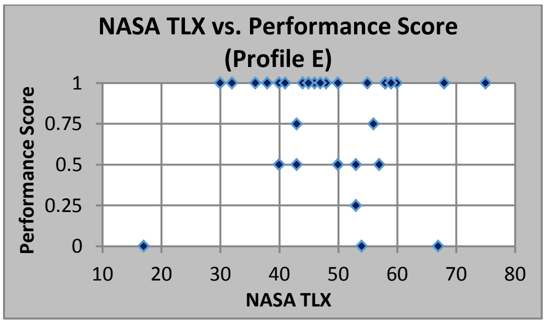


Fig. B.9 NASA-TLX vs. Performance (Profile E)

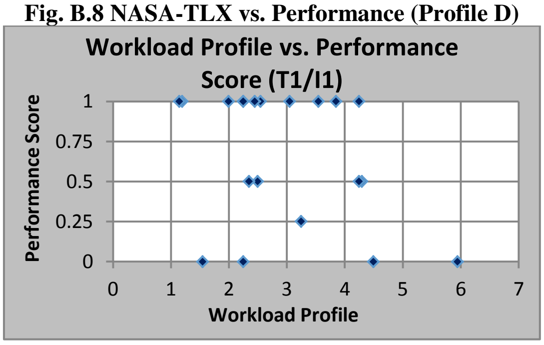


Fig. B.10 WP vs. Performance (T1/I1)

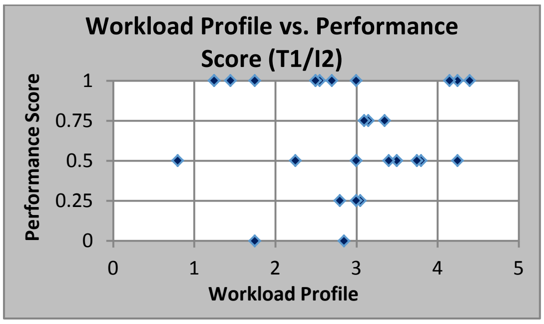


Fig. B.11 WP vs. Performance (T1/I2)

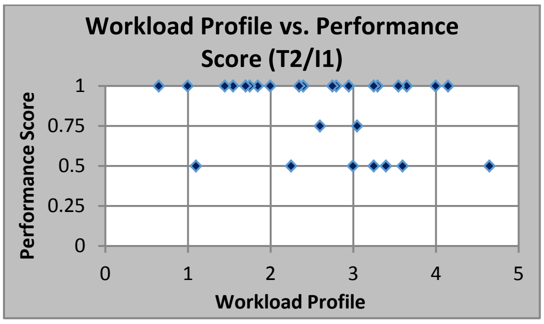


Fig. B.12 WP vs. Performance (T2/I1)

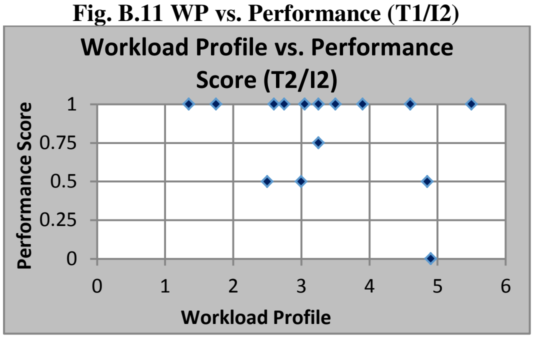


Fig. B.13 WP vs. Performance (T2/I2)

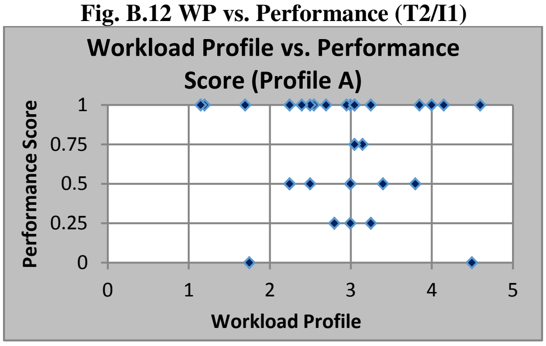


Fig. B.14 WP vs. Performance (Profile A)

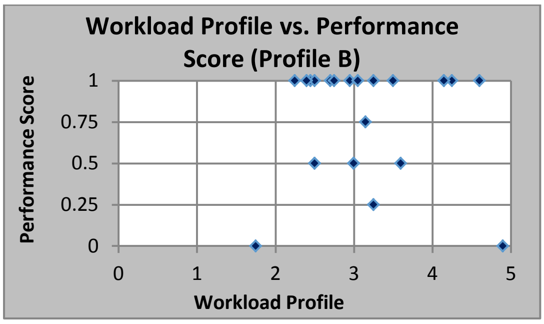


Fig. B.15 WP vs. Performance (Profile B)

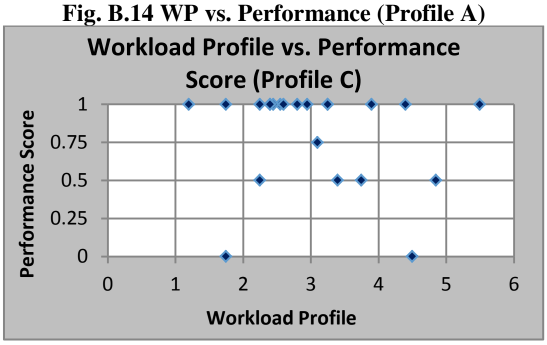


Fig. B.16 WP vs. Performance (Profile C)

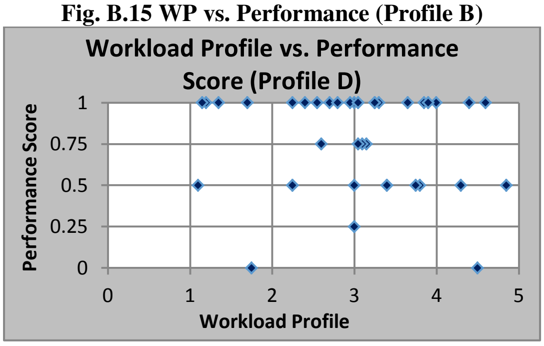


Fig. B.17 WP vs. Performance (Profile D)

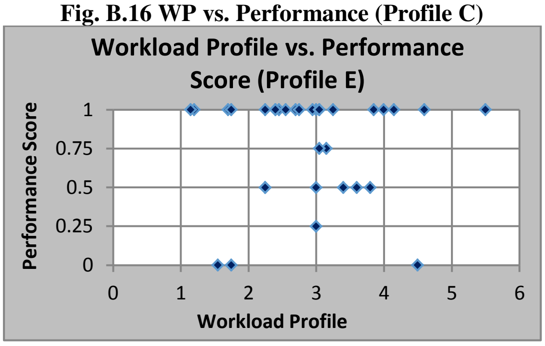


Fig. B.18 WP vs. Performance (Profile E)