USER EXPERIENCE ON THE WEBSITE OF SKAT

Author: Anne Skov Bærentsen

31. MAJ 2016

SUPERVISOR: TANJA SVARRE JONASEN
Information Architecture, Aalborg University

Table of Contents

1 Introduction

In 2013 the second wave of the mandatory self-service strategy was introduced (The Danish Agency for Digitisation, 2014). As a part of this strategy e-government infrastructure must allow citizens to report on advance tax assessment and tax returns via IT-solutions. In 2014 digital communications with the public sector became mandatory in Denmark. The goal set by government was that 80% of the population must be able to receive digital mail from the public sector (Lauterbach, 2015b, p. 6).

By 2015 96% of the Danish population that were of working age had NemID (Lauterbach, 2015a) and were thus ready to take part in the e-government strategy. During 2015 40% of 16-89 year olds contacted public sector by digital self-service or via public organisation homepages (Lauterbach, 2015a; Lauterbach, 2015b, p. 9). The people within the age range of the "working age" constitute the largest percentage of the people that have used public organisation homepages or digital self-services, whereas the percentages decrease among people of the pensioner age ranges.

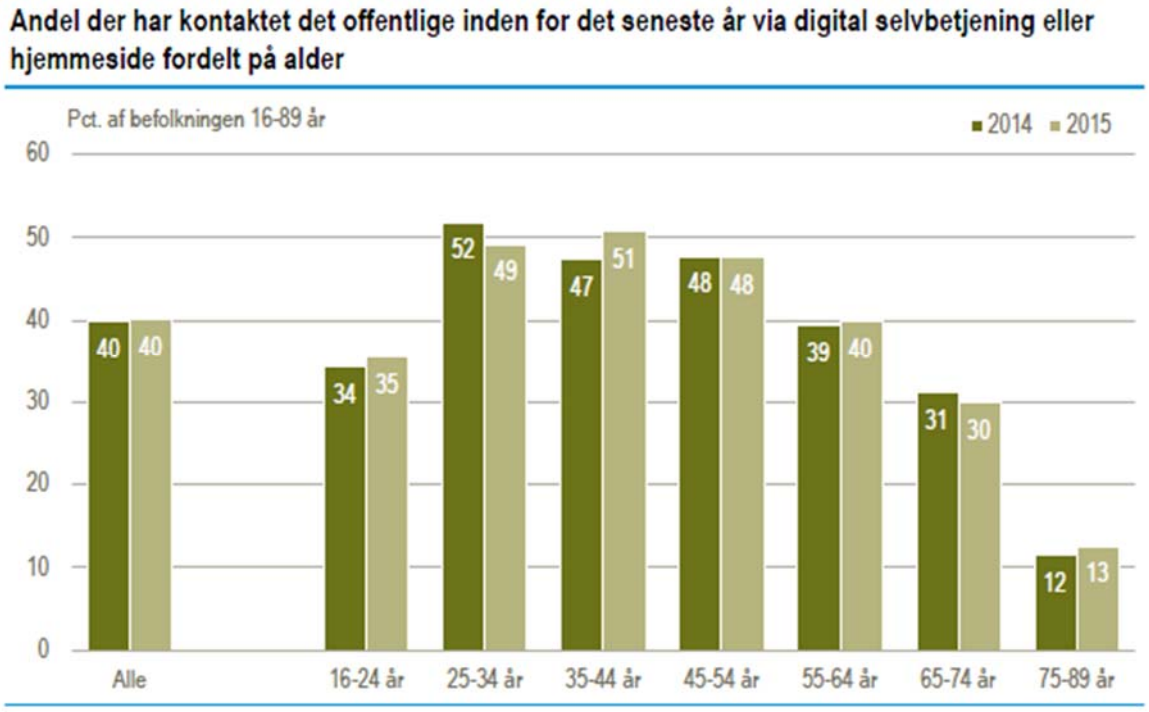


Figure 1: Percentage of Population That Contacted the Public Sector in 2014-2015 via Digital Self-Service or Homepage Distributed by Age (Lauterbach, 2015b, Figure8, p. 9)

An increasing number of people search for information on homepages of public organisations (Lauterbach, 2015b, p. 7). The number of people who find it difficult to use the homepages provided by the public sector is decreasing (ibid.). There was an increase in the number of people that have communicated digitally with the public sector and who find it easy to find information, from 79% in 2014 to 85% in 2015 (ibid., p. 8). There is a similar increase in the number of people that find it easy to use the homepages provided by the authorities, from 70% that found it easy to obtain information through the internet based services of public authorities in 2014 to 75% in 2015 (ibid.). There was a slight decrease in

the number of people who contacted public sector in person, from 22% in 2014 to 20% in 2015.

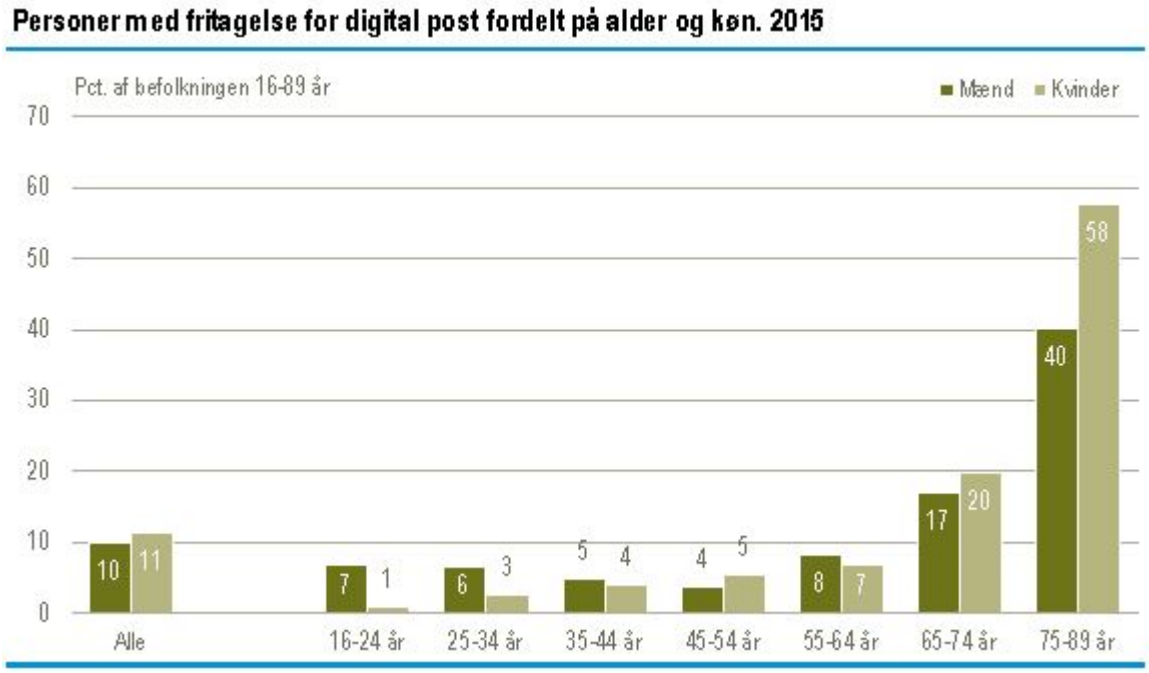


Figure 2: Persons Exempt from Digital Mail Distributed by Age and Gender (Lauterbach, 2015b, p. 10, Figure 10)

Although it is now mandatory to receive mail digitally from the public sector it is possible for citizens to be exempt if they are, for some reason, unable to receive digital mail. 11% of 16-89 year olds have been exempted from receiving digital mail from the public sector (Lauterbach, 2015b, p. 10). The most common reasons for exemption are lack of access to a computer in the citizen's own home, or “other reason”. However, as seen in Figure 2: Persons Exempt from Digital Mail Distributed by Age and Gender (Lauterbach, 2015b, p. 10, Figure 10), 7% of men between the ages of 16-24 have been exempt from using the digital mail service. In contrast only 1% of women in that group have been exempt. Among the 25-34 year olds 6% of men have been exempt, while 3% of women in the same age range have been exempt. 19% of men in the age range of 16-24 have been exempt because they do not own a computer, while 9% of men between 25-34 have been exempt for the same reason (ibid., p. 11-12). None of the women in these groups state the same reason for their exemption (ibid., p. 12). Other reason for exemption reported include “other” (24% among men, 22% among women), “disabilities” (13% among men, 21% of women), “do not know” (13% men, 10% women), “difficulties concerning NemID” (2% men, 4% women), “do not wish to respond” (6% men, 2% women), and “language barrier” (2% men, 1% women).

In the report on IT use among the population of Denmark (Lauterbach, 2015b) a chapter is dedicated to IT use among the elder citizens (65-89 year olds). However, data on the younger population is sparse. By the nature of things these next generations will likely have to use the IT-systems provided by the public sector for many more years than the older generation. Perhaps it is now relevant to take a greater interest in young peoples’ experiences with e-government services.

1.1 Motivation

Getting citizens to accept and use e-government services may require additional effort than the efforts that have already been made (Hung et al., 2006; Baumgarten & Chui, 2009; Madsen & Kræmmergaard, 2015). Nonetheless, with the introduction of e-government strategies, it is gradually becoming mandatory to use the e-government services provided. For this reason, it is becoming increasingly relevant to discuss the experience in regards to using the e-government services.

It is perhaps surprising that 7% of young men (16-24 year olds) have been exempted from receiving digital mail from the public sector. As the data from Statistics Denmark (Lauterbach, 2015b) focus on the general population, and on the elderly specifically, the question of why these young men have been exempt is left unanswered. With the emergence of terms like "digital native" (which will be explained further in chapter 2, section 2.2.2), and focus on the struggles elderly users may face in regard to IT, it seems that younger generations are expected to be able to use IT-services intuitively (Prensky, 2001), nevertheless that may not be possible for all of them (Helsper & Eynon, 2010).

However, as the number of users that report that they find it easy to obtain information from the websites of public authorities seems to be increasing perhaps it is now time to move to the next level, studying and improving the experience that the users get when interacting with public websites.

The issues surrounding the prospects of digitalisation of the Danish Public administration make it natural to investigate how the web-based services are received by the users. One public service the most Danes will need to deal with is SKAT, which has therefore been chosen as the object for investigation in the present thesis.

1.2 Problem Statement

Which kind of User Experience (UX) is offered at SKAT's homepage? Where can the UX on the existing page be improved? Is there reason to study the UX of SKAT's homepage further?

1.3 Delimitation

The problem statement will be answered through developing and executing a User Experience Test of SKAT's homepage. The test consists of a set of four tasks, that participants must solve, as well as a set of questions for each task, and a comprehensive set of questions regarding the User Experience of using SKAT's website. Questionnaire data was collected, and video and audio was recorded during the test in addition. The test was performed with participants within the age range of 20-26 years, who are fluent in Danish. The sample is not representative of the general population, and only limited information about the test participants was collected. The test focuses only on public pages of SKAT.dk, so users were not required to use their NemID to log in to any pages in order to solve tasks for the test.

2 Theory

The transition towards a digitalisation of the public sector belongs within the field of e-government. Therefore, a definition of the term e-government will be made in this chapter, after which an overview of the current tendencies within literature of the g-government field will be made as well.

The users of a system possess valuable insight regarding where the system could be improved. However, different types of users have different perspectives on the system. Therefore, it is important to consider which types of users and user behaviour to study. In this chapter the significance of user expertise in relation to information seeking behaviour will be outlined. After which, the concepts of Digital Natives and Digital Immigrants will be discussed.

Lastly, definitions of Usability and User Experience will be outlined, and the difference between User Experience and Usability will be touched upon.

2.1 The Current State of E-Government Research

In this section a definition of the word e-government will be made. It will be substantiated that e-government is a growing field, and that some users may have difficulty accepting the new e-government services. Some factors critical for user acceptance of e-government services will be identified.

The term e-government covers the concept of providing information and communication technologies in order to improve the interactions between citizens and the public-sector (Baumgarten & Chui, 2009, p. 26). Baumgarten and Chui (ibid.) report that by 2009 many countries had achieved that more than 70% of taxpayers filed their taxes online. However, many of the e-government projects have failed to generate as much interest among users as anticipated (ibid.). In addition, they have not resulted in clear improvements of efficiency.

According to Madsen and Kræmmergaard (2015, p. 3) a gap exist between "the availability and uptake of online public services". Even though the public sector has been moving towards digitalisation for decades, citizens continue to use traditional channels in their interaction with the State, be it as primary channels or as a supplement to online channels (ibid.). Hung, Chang & Yu (2006, p. 98) support this claim, stating that: "Since the late 1990s, numerous governments have made huge investments in electronic government (e-Government) services […] low levels of user acceptance of these services are recognized as an endemic problem for government policy makers, government agencies, and e-Government services providers.".

In their study of the determinants for acceptance of an online tax filing and payment system in Taiwan, Hung et. al. (ibid., p. 111) identify nine critical factors for user acceptance for e-government services: "perceived usefulness, ease of use, perceived risk, trust, compatibility, external influence, interpersonal influence, self-efficacy, and facilitating condition".

Baumgarten and Chui (2009, p. 26) claim that there is a hesitance to allow user participation during the creation of e-government services, which clashes with users' growing

expectations for internet based services. By not adopting the concept of Web 2.0 (which places an emphasis on usability, two-way communication, and interoperability) government websites run the risk of reinforcing a perception among citizens, that these sites "offer a vastly diminished experience" (ibid., p. 27). Baumgarten and Chui (ibid.) suggest that competences in areas such as customer insights, usability (among others) should be improved.

In his review, of the literature referenced in the Electronic Government Reference Library, Scholl (2014, p. 3) found that the study domain within the field of e-government has thrived since its beginnings. By analysing the frequency of use of certain keyword clusters an overview of "topical directions and scholarly interests in the study domain" (ibid., p. 11) was created. The overview can be seen in Table 3.

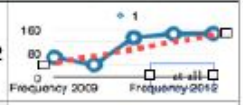
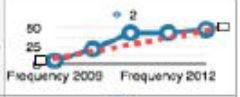
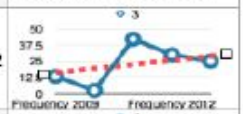
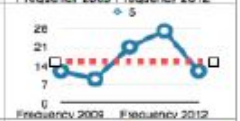
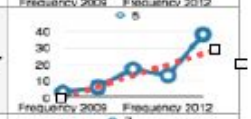
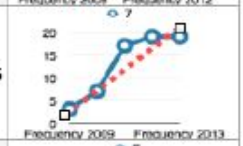
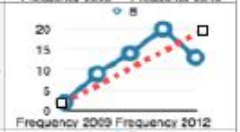
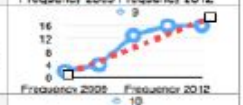
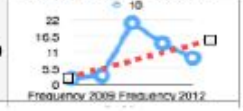
	Keyword/Cluster	Frequency 2009	Frequency 2010	Frequency 2011	Frequency 2012	Frequency 2013	Frequency 2009 to 2013	Frequency Charts 2009 to 2013
1	e-government, electronic government, internet in public administration, transformational government	67	42	132	145	146	532	
2	information & communication technologies	5	21	43	43	48	160	
3	democracy, digital democracy, e-democracy, e-voting, internet voting	13	2	42	30	25	112	
4	participation, e-participation, political participation, civic engagement, citizen engagement	5	3	34	28	20	90	
5	information, government information, electronic-government information, public-sector information	12	9	21	27	12	81	
6	government, public administration	3	6	17	13	38	77	
7	access to information, accessibility, digital divide	3	7	17	19	19	65	
8	services, information services, web services, public services, service delivery	2	9	14	20	13	58	
9	adoption, technology adoption, diffusion	2	4	13	16	16	51	
10	acceptance, user acceptance, technology acceptance model	2	3	21	14	9	49	

Figure 3: Most Frequent Topical Clusters Based on Keywords, 2009 to 2014 (Scholl, 2014, table 7, p. 12)

As shown in Table 3 "*electronic and transformational government*, *information and communication technologies* […] *participation, services,* and the *digital divide*" are the main focus of the literature.

Interestingly, the frequency of keywords such as "access to information" and "accessibility" seem to be stagnating. Likewise, keywords such as "user acceptance", as opposed to "user experience" or "user satisfaction", are present. This may indicate a focus on merely getting citizens to conform to using the e-government resources they are provided with, as opposed to making the experience enjoyable for users.

2.2 Users

The users are the people who use the IT-systems, or services, that designers make. According to Morville and Rosenfeld (2007, p. 246) users are the "ultimate designers of the Web". If a website or service confuses the users, they will stop using it, and find alternatives to help achieve their goals. This could lead systems to fail, and result in expensive redesigns. Therefore, Morville and Rosenfeld (ibid.) emphasise the importance of user-sensitive design.

However, users are unpredictable, which means that lessons learned in one system do not necessarily fit another system (ibid.). Therefore, the nature of the particular system and its users should be considered (ibid.). There are a number of different approaches to study the users, such as interviews, or observational studies (ibid., p. 247). However, as there is no single right approach to discovering information about who the users are, what their goals are, and how they search for information, many different perspectives may be needed in order to collect adequate information (ibid.).

2.2.1 Information Seeking Behaviour

Russell-Rose and Tate (2013, p. 4) state that expertise plays an important role in how users behave when seeking information. Rubin and Chisnell (2008, p. 119) stress the importance of defining what is meant by terms like "expert" and "novice". Two types of expertise are defined by Russell-Rose and Tate (2013, p.4): domain expertise, and technical expertise. Domain expertise concerns how familiar the user is with the subject, while technical expertise is about the user's proficiency when it comes to the internet, search engines, and computers in general. If both types of expertise are present, the user is more likely to succeed. It is further stated that domain novices are more likely to experience difficulty in discerning if a source is relevant or reliable, unlike domain experts (ibid.). According to Russell-Rose and Tate (ibid.) technical novices have a tendency to use a "breadth-first strategy" and "orienteering", whereas technical experts have a tendency to use a "depth-first approach" and "teleporting".

These two types of users' information seeking behaviour combine into four categories of users: Double experts, Domain expert/Technical novices, Domain novice/Technical experts, and Double novices (Figure 4).

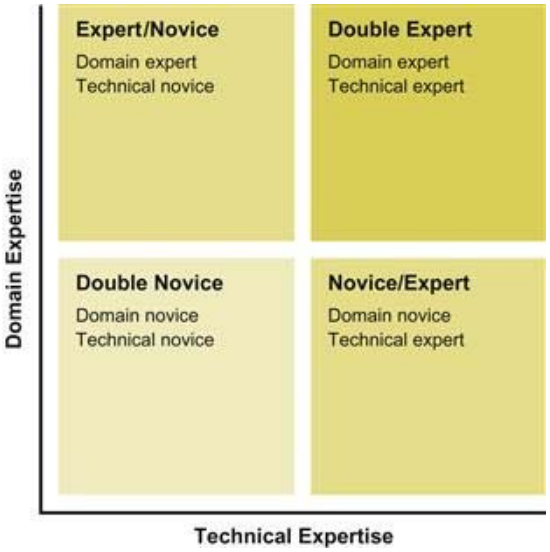


Figure 4: Two Dimensions of Expertise: Domain and Technical (Russell-Rose and Tate, 2013, figure 1.1, p. 5)

Russell-Rose and Tate (ibid., pp. 4-5) describe the typical behaviour of the different types of users thusly:

2.2.1.1 Double novices

Double novices tend to "orienteer", as the road to fulfil their information goals is most treacherous to double novices as opposed to other user types.

1. **Frequently reformulate queries:** Although novices tend to visit fewer pages than experts, they often perform more queries. They frequently rephrase their queries, but only slightly, and the changes they make to their search terms are often inconsequential.
2. **Going back:** Novices tend to return to the search pages, even after they click a search result. This is ascribed to a fear of venturing too far from the starting point.
3. **Spending more time:** Novices are likely to spend more time than experts due to making more queries and returning more frequently to previous pages.

Russell-Rose and Tate (ibid., p. 5) suggest helping double novices by supporting them in reformulating their search terms (e.g. by showing related searches), as well as letting them return to previous pages more easily (e.g. By adding breadcrumbs to the website).

2.2.1.2 Double experts

Unlike the tentative double novices, double experts tend take a "depth-first approach", attempting to "teleport" directly to their desired destination (ibid.).

The assumption is that most users will be domain novices, but will fall into either the category of technical novice or technical expert.

1. **Examine more pages:** Double experts tend click more search results than novices.
2. **Going deeper:** While double novices often return to previous pages, double experts are more likely to move further into the information space, by following links from page to page (ibid., p.6)

3. **Spending less time:** Double experts tend to spend less time looking for relevant information, as they do not rephrase their search queries as often, and are able to determine which search results are most relevant more rapidly than novices (ibid.).

Russell-Rose and Tate (ibid.) state that experts construct queries, that enables them to "teleport" directly to the relevant information, unlike the novices that have to reformulate their queries often. Thus, Russell-Rose and Tate (ibid., p. 7) suggest that expert-friendly systems should allow experts to rapidly narrow their search, by supporting advanced syntax.

2.2.1.3 Domain Expert/Technical Novices

Domain experts/technical novices are known to use their in-depth knowledge to formulate effective queries, and are quick to evaluate the relevance of search results. However, they are more likely to use a breadth-first strategy (as opposed by a depth-first approach), due to their lack of technical knowledge (ibid., p. 8).

1. **Advanced terminology:** Domain experts/technical novices possess an extensive vocabulary which enables them to construct more successful queries than domain novices.
2. **Effective evaluation:** Their domain knowledge makes them able to evaluate pages faster.
3. **Going back:** However, because of their lack of technical knowledge domain experts/technical novices tend to return to the search pages, rather than exploring the system in depth.

2.2.1.4 Domain novice/technical experts

Users that fit the domain novice/technical experts category tend to have difficulty discerning the relevant content, but are confident in the technical aspects (ibid.).

1. **Advanced formatting:** Unlike technical novices, technical experts are more likely to use query formatting (e.g. double quotes, Boolean operators).
2. **Confident exploration:** Because of their technical expertise they tend not to worry about becoming lost in navigating the system, which leads them to explore the system more deeply (ibid., p. 9).
3. **Difficulty with evaluations:** Since these users have a lack of domain knowledge they tend to have difficulty assessing the relevance of a page (ibid.).

Russell-Rose and Tate (ibid.) emphasise the importance of not prioritising one category of users at the expense of another, even though these different types exhibit unique approaches, unless it is necessary. Russell-Rose and Tate (ibid.) stress the value of understanding the level of expertise of the intended users, but conclude that it is often necessary to design for both novices and experts.

2.2.2 Digital Natives and Digital Immigrants

As the users tested in the user experience test were selected based on their age it is useful to explain the distinction between Digital Natives and Digital Immigrants. Digital Natives belong to the generations that grew up to become ""native speakers" of the digital language

of computers, video games and the Internet." (Prensky, 2001, p. 1). In contrast, Digital Immigrants have learned the use of computers when they were older, and thus retain an "accent" (ibid., p. 3). These two groups have different needs and expectations when it comes to using IT systems. Digital Immigrants may be slower in turning to the internet for information, and more likely to read a manual for a program, whereas Digital Natives are used to getting information much faster (ibid.). Prensky (ibid., pp. 3-4) lists some characteristics of Digital Natives. They:

- "like to parallel process and multi-task."
- "prefer their graphics before their text".
- "prefer random access (like hypertext)."
- "thrive on instant gratification and frequent rewards."
- "prefer games to "serious" work."

The difference between digital natives and non-digital natives is often seen as a question of year of birth (Helsper & Eynon, 2010, p. 503). However, Helsper and Eynon (ibid.) separate "doing" ("digital activities […] that indicate digital nativeness") and "being" ("types of people […] most likely to demonstrate these characteristics"). The validity of viewing the concept of "digital natives" as a product of being from a certain generation is increasingly questioned in academic research (ibid., p. 505). Subscribers to the concept of digital natives/digital immigrants tend to "assign broad characteristics […] to an entire generation […] and suggest all young people are expert with technology." (ibid.). Although young people are more likely to use Internet and new technologies, Helsper and Eynon (ibid.) state that the reason why, and the mode in which and how effectively they use these technologies, differ significantly. Several writers have emphasised the complexity and diversity of young people's use of technologies (ibid.). According to Helsper and Eynon (ibid., p. 504) "breadth of use, experience, self-efficacy and education" are more important than age when it comes to becoming a digital native.

2.2.3 User Group

Kuniavsky (2003) and Rubin and Chisnell (2008) suggest that user group should belong to a specific, relatively narrow demographic. However, in the case of SKAT the potential user group is a very large percentage of the population. The users of SKAT are not as such customers, as they have little to no influence over the provider of the service. As described in chapter 1 96% of Danes of working age have NemID, and thus they are in principle able to log in to SKAT's homepage. This means that more or less anyone that has dealings with the Danish Tax and Customs Administration are potential users.

Discussion of the usefulness (or lack there off) of the shift from physical interaction to internet mediated interaction, between citizens and state, often focus on problems the elderly may encounter. In the report regarding IT use among the Danish population (Lauterbach, 2015b) data on young citizens' use of IT was limited. However, a chapter was dedicated solely to presenting data on IT use among the elderly. This may be an indication that young people are assumed to be Digital Natives, unlike people above the age of 64, who may be assumed to need help when it comes to IT use. However, the younger

generations will especially, and increasingly, be expected to use technology for their interactions with the state. If the expectations of IT users are indeed growing, as suggested by Baumgarten and Chui (2009), the next generations may hold interesting insights regarding the User Experience on websites provided by the public sector.

2.3 Usability

Kuniavsky (2010, p. 14) defines Usability as "the practice of making things easy to use". The terms of Usability and User Experience are often confused as two separate terms for the same concept (ibid.). However, they are not the same. Where poor Usability may "break" a product, good Usability will not in itself provide a good user experience (ibid.).

Rubin and Chisnell (2008, p. 3-4) state that usability is only an issue when it is lacking, and absence of frustration is what makes a product usable. Rubin and Chisnell (2008, p. 4) further state that "when a product or service is truly usable, the user can do what he or she wants to do the way he or she expects to be able to do it, without hindrance, hesitation, or questions."

Rubin and Chisnell (ibid.) lists six important aspects of usability:

- **Usefulness:** Concerns how useful the product is to the user. If the product is not useful for achieving the relevant goals, there is no point for the user to engage with the product. Rubin and Chisnell (ibid.) claim that this aspect is most often overlooked when studying usability.
- **Efficiency:** Whether or not the user will be able to achieve their goal accurately and completely, generally within a short amount of time, depends on the efficiency of the system.
- **Effectiveness:** Concerns the degree to which the product behaves as expected, and how easily the user is able to use the product to reach their goals.
- **Satisfaction:** This aspect concerns the user's perceptions, feelings, and opinions of the product. Rubin and Chisnell (ibid., pp. 4-5) state that this aspect is generally measured through both oral and written questioning, and typically through asking users to rate and rank the product.
- **Learnability:** This regards the user's ability to use the system "to some defined level of competence after some predetermined amount and period of training" (ibid.).
- **Accessibility:** How usable the product is to people with disabilities.

The discipline of making things usable is part of the field of user-centered design, which is part of the concept of experience design (ibid., p. 5). When no troubles are encountered and tasks can be successfully solved the user may not even notice that a system has good usability (ibid., p. 6). As Rubin and Chisnell (ibid.) state: "True usability is invisible". Usability is measured through the number of problems users face when interacting with the system (ibid.).

2.4 User Experience

According to Hassenzahl and Tractinsky (2006, p. 91) the concept of User Experience is a counter reaction to the Usability paradigm. Hassenzahl and Tractinsky (ibid.) state that the

ideas surrounding the concept of User Experience are not original, as they are already to some extent represented in areas such as Usability. It is further stated that there is a lack of empirical research on the subject of User Experience, which slows down the understanding and development of User Experience as a concept (ibid., pp. 91-92).

Kuniavsky (2010, p. 14) states that there is no universal definition of User Experience. According to Kuniavsky's (ibid.) any definition of User Experience must take into consideration all aspects of what users of the system regard as part of their experience. Kuniavsky (2010, p. 14) describes User Experience as:

> *"The user experience is the totality of end users' perceptions as they interact with a product or service. These perceptions include effectiveness (how good is the result?), efficiency (how fast or cheap is it?), emotional satisfaction (how good does it feel?), and the quality of the relationship with the entity that created the product or service (what expectations does it create for subsequent interactions?)."*

Similarly, Hassenzahl and Tractinsky (2006, p. 95) define User Experience as "a consequence of the user's internal state […], the characteristics of the designed system […] and the context (or the environment) within which the interaction occurs". They further state that User Experience is about "designing for pleasure, rather than for absence of pain" (ibid.).

Buxton (2007, p. 127) regards Experience Design as the most human centric of the terms in the field of human computer interaction. As he (ibid.) states: "It is ultimately experiences we are designing, not things.". Although physical objects are often the most tangible outcome of the design process, their primary function is to engage in an experience. Experience is shaped by affordances (Gibson, in Buxton, 2007, p. 127) and character embedded into the product itself (Norman; Gaver, in Buxton, ibid.). Buxton (ibid.) emphasises that aesthetics and functionality play an important role as well.

Buxton (ibid., p. 129-131) exemplify how the use of two seemingly similar products may result in very different experiences through his experiences with different juicers. Although the usability of the different juicers was similar, the experience of using one of the two products was significantly better, than with the other (ibid., p. 129).

Hassenzahl and Tractinsky (ibid., p. 92) identify three major perspectives in User Experience literature (as shown in Figure 5Figure 5: Facets of User Experience (Hassenzahl & Tractinsky, 2006, p. 95)): "addressing human needs beyond the instrumental", "affective and emotional aspects of the interaction", and "the nature of experience".

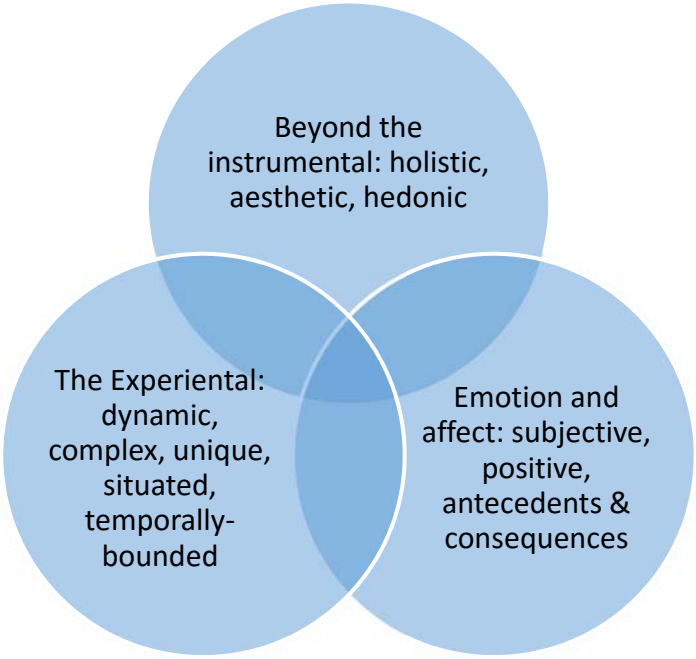


Figure 5: Facets of User Experience (Hassenzahl & Tractinsky, 2006, p. 95)

Within the field of Human Centred Informatics ensuring the instrumental value of a product has been the major focus (ibid., p. 92). However, this focus has been repeatedly challenged. Hassenzahl and Tractinsky (ibid.) exemplify this with the notion of beauty as an important quality aspect, which clearly goes beyond the instrumental value of a system; “Beauty is an end rather than a means.”.

It is argued that Human Centered Interaction must deal with both how systems may fit behavioural goals (Hassenzahl and Tractinsky refer to this as “pragmatic aspects”), as well as hedonic aspects for instance; stimulation, identification, and evocation (ibid.).

Hassenzahl and Tractinsky (ibid., p. 93) refer to The “Affective Computing” project, which was a pioneering attempt to create a discourse on affect in relation to Human Centered Informatics. However, it is noted that the perspective of affective computing is computer centered rather than user centered, i.e. affective computing concerns “how computers can sense user affect, adapt to it, or even express its own affective response” (ibid.). Furthermore, affective computing mainly depicts negative emotions in the interaction between humans and technology (ibid.).

In User Experience, Hassenzahl and Tractinsky (ibid.) argue, the perspective should be human centered. In addition, they (ibid.) state that User Experience is primarily focused on positive emotions. User Experience concerns the understanding of “the role of affect as an antecedent, a consequence and a mediator of technology use” (ibid.). Consequently, User Experience focuses on creating “*positive* emotional outcomes”, whereas Human Centered Informatics simply focuses on preventing negative emotions.

Hassenzahl (2002, 1 Introduction to User Experience and Experience Design) is not keen on the focus on usage and the user in User Experience, as this focus is rooted only in action. He

regards experience as a broader field. Hassenzahl (ibid.) believes that User Experience should be thought of in terms of experiences and experiencing, as a way of creating experiences, and as Experience Design. He regards User Experience as a field of study, focused on studying experiences and the user's needs. The term User Experience originated in the usability community, where the focus was on usage. However, "experience" is more focused on the user, their psychology, and emotions (ibid.). Over time people started to take the term "experience" more seriously, not just as a term for denoting closeness to users (ibid.).

According to Hassenzahl (ibid.) there are two ways of looking at experience:

1. Moment-by-moment experience: What is felt, done, and thought at a particular moment combines to the experience you have of the moment. Focus is on the "how", both in regards to tactility, aesthetics, interactivity (e.g. How the product feels, etc.). Moment-by-moment experience occurs over a short time.
2. Memories-of-experience: What is remembered. Constructing story of usage, what is remembered, and communicated to others. Episodic memories.

Through these two perspectives on experience Hassenzahl (ibid,) distinguish between experiencing and experience (as a narrative). In Hassenzahl's definition the objective of Experience Design is understanding what is important in the stories created through the product within the user. The design task thus becomes how to "re-tell" those stories through the product.

2.4.1 Needs of the User

In addition, Hassenzahl (2002, 2 Advice on Designing with Experience in Mind) emphasises the importance of considering at the actual needs of the user. He (ibid.) calls this the "Why of product use". If designers were to design with the actual goals in mind they would have to ask "why", i.e. what actual goal the user wants to fulfil.

Examples of actual needs:

- Feeling relatedness
- Feeling competent
- Feeling popular
- Helping others
- Competing

Hassenzahl (ibid.) claims that asking people how they feel about products is futile. However, they love to talk about themselves. For this reason, he suggests that designers ask questions like "Why do you love the iPhone?". He claims that enjoyment is essential, and that positive and worthwhile experiences are key-words, when discussing User Experience design.

Experience Design should be a dialogue between the designer and user Hassenzahl (2002, 3 Main Guidelines and Ethical Considerations). Where Usability focuses on objective measurements, User Experience is more subjective (ibid.). One way of guiding an experience

design process is the Three Level Hierarchy of Goals model (Hassenzahl, 2010, p. 12; Figure 6).

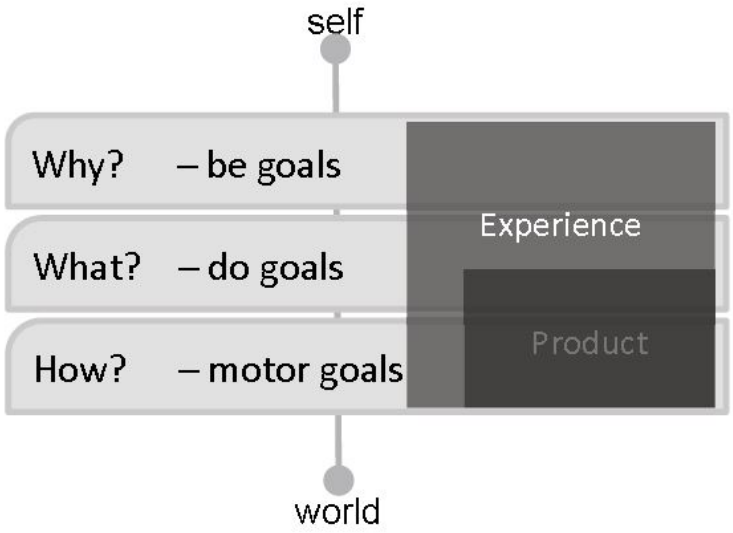


Figure 6: A Three Level Hierarchy of Goals (Hassenzahl, 2010, figure 2.1, p. 12)

- **Why:** Needs of people, emotions (be goals). Be goals are what "motivate the action and provide it with meaning.". Hassenzahl (2010, p. 13) claims that the action of "Making a telephone call" is not a meaningful action in itself. However, if a person is feeling lonely they may call someone in order to obtain a feeling of relatedness.
- **What:** The core functionality (do goals). The "concrete outcome, an actor wants to attain" (e.g. "making a telephone call"). Do goals are not necessarily tied to one particular technology (e.g. "making a phone call" can be achieved through a number of different types of phones, or through different types of computer applications). However, the goals are dependent on some form of technology. Hassenzahl (ibid., p. 12) claims that do goals are "*born* out of a general technology". Without the technology the goal would simply not exist.
- **How:** The interaction (motor goals). The process needed to achieve the do goal can be decomposed into individual steps, e.g. "pressing buttons" (ibid.).

2.4.2 Experience

Hassenzahl (2010, p. 2) states that the underlying principles of User Experience are not fundamentally different from experience in general. However, User Experience focuses on experience that is mediated by "interactive products" (ibid.). This distinction between general experience and User Experience may not be significant from the user's perspective, since the experience is an experience regardless of the medium (ibid.). The distinction is important to User Experience designers nonetheless, as the designer must strive to design for certain experiences, and thus have to put experience before product (ibid., pp. 2-3).

According to Hassenzahl (Various authors in Hassenzahl, 2010, p. 3) "emotion, cognition, motivation, and action" are intricately intertwined, and furthermore that emotion is linked to action and motivation. Based on this he concludes that emotion is crucial for experience.

Hassenzahl (Various authors in Hassenzahl, ibid., p. 3-4) states that it is important to note that “emotions are best viewed from the perspective of our goals”, which implies “motivated action”. He concludes that experience is a result of a combination of “perception, action, motivation, emotion, and cognition in dialogue with the world”. Experience is thus created through a combination of many elements, but as stated by Hassenzahl (ibid.) it is centered around emotion. Experience is a result of a combination of sensory input, emotion, motives, actions (ibid., p. 8). The greatest difference between experience and User Experience is that in User Experience the focus is on interactive products as “creators, facilitators, and mediators of experience” (ibid.).

Traditionally Human Centered Informatics has based recommendations on the “objective”, i.e. what could be measured through observations and User Testing, as opposed to the opinions of the users (ibid., p. 9). However, Hassenzahl (ibid.) states that if the positive qualities of a product are not experienced, then the product may subjectively be considered inadequate, even if a product is objectively good (for instance based on measurements of task solving times). Hassenzahl (Tversky & Kahneman in Hassenzahl, 2010, p. 10) further notes that initial values may affect people’s judgement. That is, people may adjust their final judgment on previous judgments, known as “anchoring and adjustment” (ibid.).

2.5 Summary

The strategy of providing e-government services is a growing field, and has been the focus of study for some time. However, getting users to accept the services provided may require more, than simply making them available. Nevertheless, it has been claimed that there is a hesitance to allow users to participate in creation of these services. In a literature review it was found, that keywords such as “user experience” and “user satisfaction” were absent from the list of most frequently used keywords.

Selecting the right group of users to test is important, as the test results will be dependent on which users are tested. Different types of users may display different information seeking behaviours, and users belonging to different demographics may behave differently as well. Four types of information seeking behaviour were outlined. In chapter 4, section 4.4, the users' behaviour will be considered, in order to consider which groups the tested users are likely to belong to. Users may fit the description of either the term “Digital Natives” or “Digital Immigrant”. However, it has been argued, that age alone is unlikely to guarantee “expert level” users. This will be further touched upon in the discussion of the findings in chapter 5.

Usability is a part of User Experience, since the Usability of a product influences how the overall product is experienced by the user. Where the main focus of Usability is on minimising frustrations and obstacles for the user, User Experience Design goes beyond the interactions that can be measured objectively, by focusing on creating a positive experience for the user.

3 Method

In this chapter the methods for data collection and data analysis will be described. First a general outline will be made of the approaches taken in the thesis. After this there will be a brief presentation of how the study deals with the overlaps and differences between User Experience and Usability. In order to investigate the User Experience of SKAT.dk a User Experience Test was designed. The present chapter will to some extent describe the considerations and process of the different aspects of the test design and final execution of the test.

In the next part of the present chapter considerations regarding the users will be outlined, such as reflections on which test participants to recruit for the User Experience Test, as well as an overview of the recruitment process. Hereafter the setting in which the tests took place is described.

Reflections on the task design are then presented, after which an explanation is made of how the individual tasks can be solved. The examples of task solutions serve to exemplify the differences and similarities of the tasks, as well as to enable a basic understanding of the website for anyone unfamiliar with the site. After this a description is made of the software used for the test and its limitations. The User Experience Questionnaire that was implemented into the test design will also be described.

Prior to conducting the User Experience Test pilot tests were made, after which the test design was adjusted. This process will be outlined before making an account of the procedure of conducting the final test. Following the account of the procedure a brief statement is made regarding the use of the recordings from the tests, as well as how the collected data will be analysed. Lastly, some of the ethical considerations of the test design and process are presented.

3.1 Data-collection

The present investigation is concerned with an attempt at understanding the experience of users of SKAT's homepage. Understanding the experience of people is a core issue for hermeneutics, and the present investigation may thus be regarded as belonging to the hermeneutical tradition, although its ultimate aim is rather pragmatic and related to the use of SKAT's homepage.

"Hermeneutics is the study of the interpretation of texts" (Kvale, 1996, p. 46). Additionally, obtaining valid and common understanding of a text is the goal of hermeneutical interpretation. Since the origin of the concept of hermeneutics there has been a broadening of the term "text" within the context of hermeneutics. In the context of this thesis the "text" to be interpreted is the data collected through a User Experience Test. Like Kvale's (ibid., p. 50) research interviews the data that is the object of interpretation in this study, is not finished texts, which is what hermeneutics traditionally has been concerned with interpreting. The subject of interpretation (i.e. the data) is created by the researcher among others, who is to interpret it. Thus, it is "not a pre-given literary text, but emerges in the same process as its interpretation" (ibid.).

Kvale (ibid., p. 47) states that the "interpretation of meaning is characterized by a *hermeneutical circle.*". "The understanding of a text takes place through a process in which the global meaning of the separate parts is determined by the global meaning of the text, as it is anticipated." (ibid.). Eventually a closer determination of meaning of separate parts may change the "originally anticipated meaning of the totality, which again influences the meaning of the separate parts, and so on" (ibid.).

3.1.1 Approach

The approach taken in the present thesis combines aspects of exploratory, descriptive and explanatory studies. According to Kelly (2009, p. 25) **exploratory studies** typically have broad, open-ended research questions. They "often employ a variety of research methods". Instead of making predictions the goal is to learn more about a phenomenon. The methods used in exploratory studies are often less structured than in "descriptive or explanatory studies". Findings from exploratory studies can sometimes lead to other types of studies (ibid.).

On the other hand, **descriptive studies** document and describe particular phenomenon (Kelly, 2009, p. 26). The purpose "is to provide benchmark descriptions and classifications.", and results can "inform other studies.". Sometimes descriptive studies lead to "a weaker form of prediction via correlation analysis", although it will not "explain why a relationship exists between two variables." (ibid.).

Kelly (2009, p. 26) further states that **explanatory studies** aim to "examine the relationship between two or more variables" in order to become able to predict or explain the subject that is being studied. Often explanatory studies are concerned with demonstrating causality between different variables, which requires systematic study and isolation of the variables of interest. This type of study is often conducted in laboratory setting, in order to allow the researcher to have the most control over the situation. Therefore, explanatory research generally uses more structured methods than exploratory and descriptive studies. Explanatory studies often include hypothesis testing. They do not necessarily offer explanations for the observations and statistics they provide (ibid).

The problem statement for the present thesis is relatively broad, as the objective is to examine the general experience for users of a particular website. As there was no hypothesis to test initially, an explorative approach was sought at the beginning of the process, that allowed collection of a variety of data. The research design created has similarities with Usability tests, and aims to explore a variety of different aspects regarding the User Experience of SKAT.dk, as well as examine the connection between some of these variables. Among other things, the test makes use of asking of verbal questions, asking the test subjects to think-aloud while solving specific tasks, measuring the time participants need in order to complete tasks, and a questionnaire about the general user experience for data collection. Through the process of making the research design, a notion emerged, that the experience of the users interacting with SKAT.dk may be tied with aspects like difficulty with using the site. This notion led to a decision for the research design: Data for different variables, such as time needed to complete tasks, level of frustration, perceived task

difficulty, user experience aspects, was noted separately. This was done in order to be able to examine the relationship between different variables, in order to be able to give possible explanations. The analysis of the collected data is descriptive. No predictions will be made based on the test findings.

A semi-naturalistic research design was sought, where tests took place in an environment familiar to test participants. However, the test design is relatively structured as tasks and questions have been prepared, so that specific variables can be examined to see if there is any connection that may explain the patterns of participants' ability to solve tasks on SKAT.dk and their experience with using the website.

3.1.1.1 The Evolving Map of Design Practice and Design Research

Sanders' (2008) Evolving Map of Design Practice and Design Research (as shown in Figure 7) offers another way of illustrating the approach taken in the User Experience Test. The map has two dimensions, the X-axis describes the mind-set of the design research, while the Y-axis illustrates the approach (ibid., p. 13).

The approaches axis has research-led perspective at the bottom and design-led perspective at the top. While the research-led perspective is the oldest of the two, the design-led perspective has recently become more visible (ibid.)

On the mind-set axis an expert mind-set is placed at the left, and a participatory mind-set on the right. In the expert mind-set researchers take on the role as experts, referring to the people that use the system as "subjects", "users", or other similar terms (ibid.). As opposed to this perspective, the participatory mind-set regards the people as the true experts, and people are considered as co-creators in the process (ibid.).

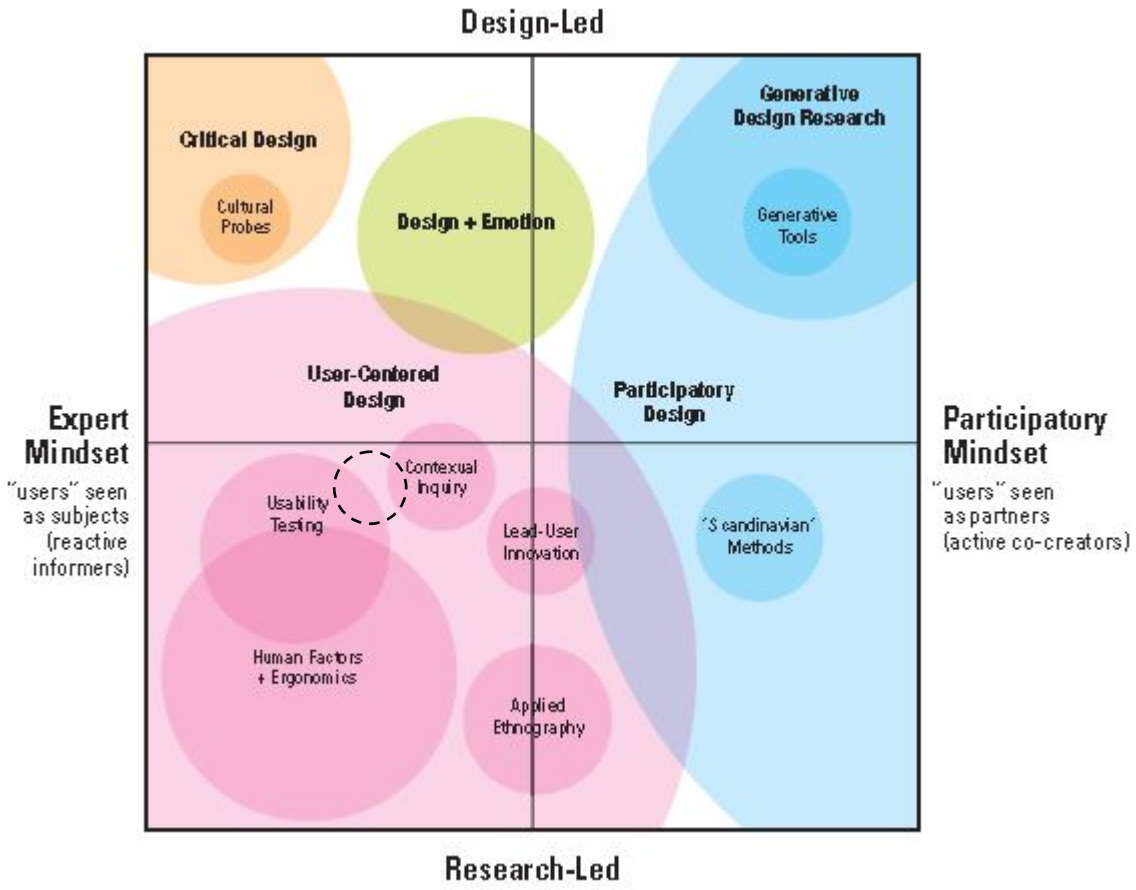


Figure 7: Evolving Map of Design Practice and Design Research (Sanders, 2008, figure 2, p. 14)

The present study is placed between the Usability Testing and Contextual Inquiry on the map (Figure 7), as marked by the black circle. In order to answer the problem statement an expert mind-set/research-led approach was called for. The problem statement concerns conduction of research on the current state of User Experience on the website of SKAT, as opposed to designing a new website. The development and execution of the User Experience test will be led by a researcher, who is familiar with the system to be tested, and who thus takes on the position of an expert. The users will be given specific tasks to solve, enabling collection of data on troubles they may encounter when trying to solve the tasks, similar to a usability test. In addition to answering questions on the overall User Experience of using the system, they will also be asked to interact with the system, while voicing their reflections on their interactions.

3.1.2 User Experience vs. Usability

Since there is a certain overlap between the area of Usability, and the area of User Experience, it is appropriate to substantiate that the present research is not simply a Usability test. Usability chiefly concerns which obstacles the user faces in achieving goals through the system, quantitative data such as time taken to complete tasks, number of mouse-clicks needed and similar information.

User Experience research offers a broader perspective by aiming to uncover the experience and impressions that the user is left with after they interact with the system.

While the present test contains data that makes it possible to measure some aspects of the level of Usability on SKAT's website, questions regarding the experience of the participants enter regarding the interaction with the system were asked so as to allow study of the experience itself. During the test participants are asked to rate, and if possible to verbalise, their level of frustration (or lack thereof) in conjunction with whether or not they found solving the specific task easy or not. The combination of data gathered should make it possible to substantiate which Usability issues lead to bad User Experience for some users, as well as some perspective on the User Experience on a more general level. At the end of the interaction with the system participants were asked to rate the system on several scales reflecting the overall experience. The questions will be listed in section 3.1.7 of this chapter.

3.1.3 Users

In order to conduct meaningful tests, it is necessary to figure out which demographic the users belong to, as well as what their physical and employment characteristics are, what their web use, and -experience level is, along with their technological experience level, and experience with computers in general (Kuniavsky, 2003, p. 85: Rubin & Chisnell, 2008, p. 116). Kuniavsky (2003, pp. 86-87) further suggests taking into consideration which segments would be most affected by the research, and whether to split segments into multiple groups, since the users of a product can in some cases have contradicting characteristics. Rubin and Chisnell, (2008, p. 115) also stress the importance of finding the right users to test. Otherwise any results of the test may be misleading or in the worst case they may even be incorrect. Therefore, effort should be put into identifying and describing the behaviour, skills, and knowledge of the users, in order to create a user profile of the target audience.

For the purpose of this thesis, the user group selected was Danish people between the ages of 18-26. However, in reality only users between the ages of 20-26 were tested. This group of users was chosen because they are likely to use e-government services to a wider extent than the older generations, as mentioned in chapter 1. No occupational demographics were chosen, as people from all educational and occupational backgrounds are expected to use SKAT's homepage. However, it is expected that test participants are students for the most part, due to the age range and the locations used for participant recruitment. The geographical location of test participants is the central part of Jutland for practical reasons, although they may come from various parts of Denmark.

3.1.3.1 Recruitment

Regarding recruitment, Kuniavsky (2003, pp. 88-90) suggests sending out an email questionnaire to find people that fit the target group and ask for contact information.

Recruitment was carried out in three waves. However, only the last wave was successful.

In the first wave a number of high schools in the middle part of Jutland were approached, either by sending emails to the administration, or by contacting friends and acquaintances through Facebook, that are employed as teachers in different high schools. One high school teacher agreed to send out a preliminary survey to his class, which 22 of the students filled out.

In the second wave, the administration at Business Academy Aarhus (Erhvervsakademi Aarhus) was contacted in order to get permission to set up test equipment and test students. Visits to the Business Academy Aarhus' facility, as well as the University campus at Aalborg University, were made. A stand was set up by the canteen, with a representative present to answer questions and carry out tests at the spot.

Lastly in the third wave, a recruiter at Skjoldhøj Kollegiet was employed. The recruiter posted a link to a preliminary survey on the Facebook group of Skjoldhøj Kollegiet, which two respondents filled out. This approach proved the most effective. The recruiter continuously found participants among the residents and guests at Skjoldhøj Kollegiet, securing a continuous succession of tests being conducted. It should be noted that the recruiter was also a participant. The test where the recruiter was participant took place at a different time and place.

During the first and third wave, a preliminary survey was distributed to a high school class and shared at Skjoldhøj Kollegiet's Facebook group, in which respondents were asked whether they were interested in participating in a User Experience test. They were also asked about their age, as well as a few questions regarding how much prior experience they have with SKAT's homepage. Furthermore, respondents were asked to submit their email-address. However, although all respondents had initially expressed an interest in participating, none of them actually participated in the end.

In addition, two participants were recruited through Facebook. These test participants are personal acquaintances of the researcher. These particular participants where tested at the home of a mutual friend, and the home of the researcher respectively. In the end a total of 15 people were successfully recruited and tested.

3.1.4 Setting

The setting in which the tests were to take place was carefully considered. Kelly (2009) lists laboratory studies and naturalistic studies as possible types of settings. While laboratory studies allow researchers to be more in control of the test situation, they are often criticised for being artificial, and not representing real life (ibid., p. 28). Naturalistic studies allow researchers to observe interactions as they occur in their natural setting (ibid.). However, in naturalistic studies the researcher has limited control over the setting, thus making crossover comparison more difficult (ibid.). In addition, Kelly (ibid.) notes that naturalistic studies are more intrusive than laboratory studies, and that participants have to be willing to give up some privacy in such studies.

A semi-naturalistic test design was used, that sought to mimic the natural setting in which interaction with the website might occur in real life. However, the study was designed to monitor the participants' interaction with the system, and present them with tasks in a way similar to how such a test might be structured in a laboratory setting.

The test equipment used was a laptop computer with a corded mouse, and an external audio-recorder. The software used to record individual tests was Morae Recorder, which will be further described in section 3.1.6 of this chapter. The participants solved tasks using

the Microsoft Edge 10 browser app in Windows 10, since this is generally the default browser on new Windows systems, unless the user specifically chooses another browser. It should be noted that browser choice may affect how a website performs. For instance, different browsers may display content differently. Therefore, all tests were conducted using the same browser.

Test equipment was set up in places where the participants were already present, so as to make participation as easy for participants as possible. The test was designed so that it could take place in settings where the participants presumably felt somewhat at home (i.e. at the participants' place of education, or close to their home). However, in order to protect the privacy of the participants, the tests were not to be carried out in their own homes. Since some participants were initially expected to be high school students it was especially important that the design was as respectful of participants' privacy as possible. Although participants were fully aware of the researcher's involvement in the test, and had been briefed on the purpose of the test, it was deemed important that the test cause as little annoyance and disruption to the participants' daily routines as possible. The researcher would first let participants try to solve the tasks on their own, and only interfere when participants expressed doubt regarding whether they had solved the task or not, or if they had very clearly misunderstood the tasks objective.

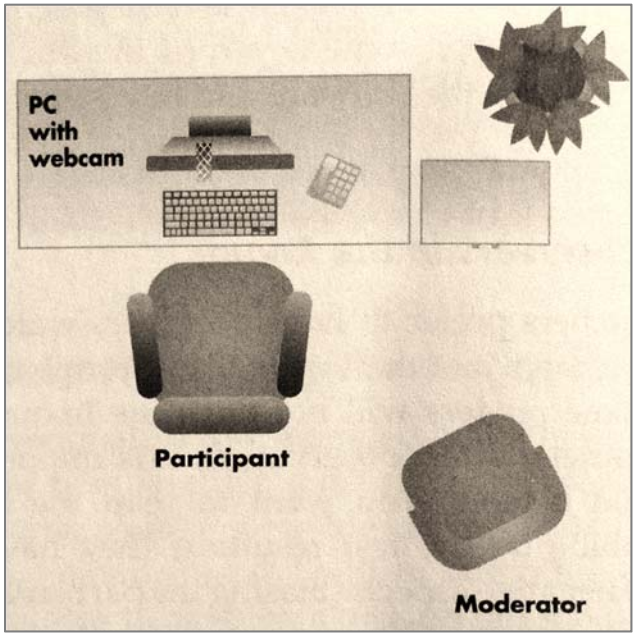


Figure 8: Monitoring Configuration with In-Room Observers (Section) (Rubin & Chisnell, 2008, figure9-4, p. 222)

The participant sat in front of the computer, with the researcher placed on their right side, in a similar setup as illustrated in Figure 8. The audio-recorder was placed roughly where the telephone is placed in the figure. Most participants were tested at the same apartment at Skjoldhøj Kollegiet. However, one participant was tested at her boyfriends' apartment, one was tested at the apartment of the researcher, and three were tested in the common room of the apartment building at Skjoldhøj Kollegiet.

3.1.5 Task Design

During the test participants were asked to solve four tasks using SKAT's homepage. The task descriptions are in Danish, as the tests were conducted in Danish. Initially a total of five tasks was created. However, four tasks were chosen for the final User Experience Test:

1. In the first task participants were asked to imagine they were about to order a t-shirt on a website. The product is shipped from outside the EU, and thus special taxation rules may apply. Participants were presented with the price of the product, as well as shipping costs, and asked to find the rules for taxation on the website, as well as calculate the total amount of customs duties they may apply.
2. The second task required participants to find a definition of the term "grænsegænger" (Transfrontier worker).
3. In the third task participants were asked to calculate their SU (State Education Grant) after taxes had been paid. The SU rate, income tax rate, and tax deduction rate, were stated in the task description.
4. In the last task participants were asked to find the next seminar regarding tax and VAT scheduled to take place either in Aarhus or online.

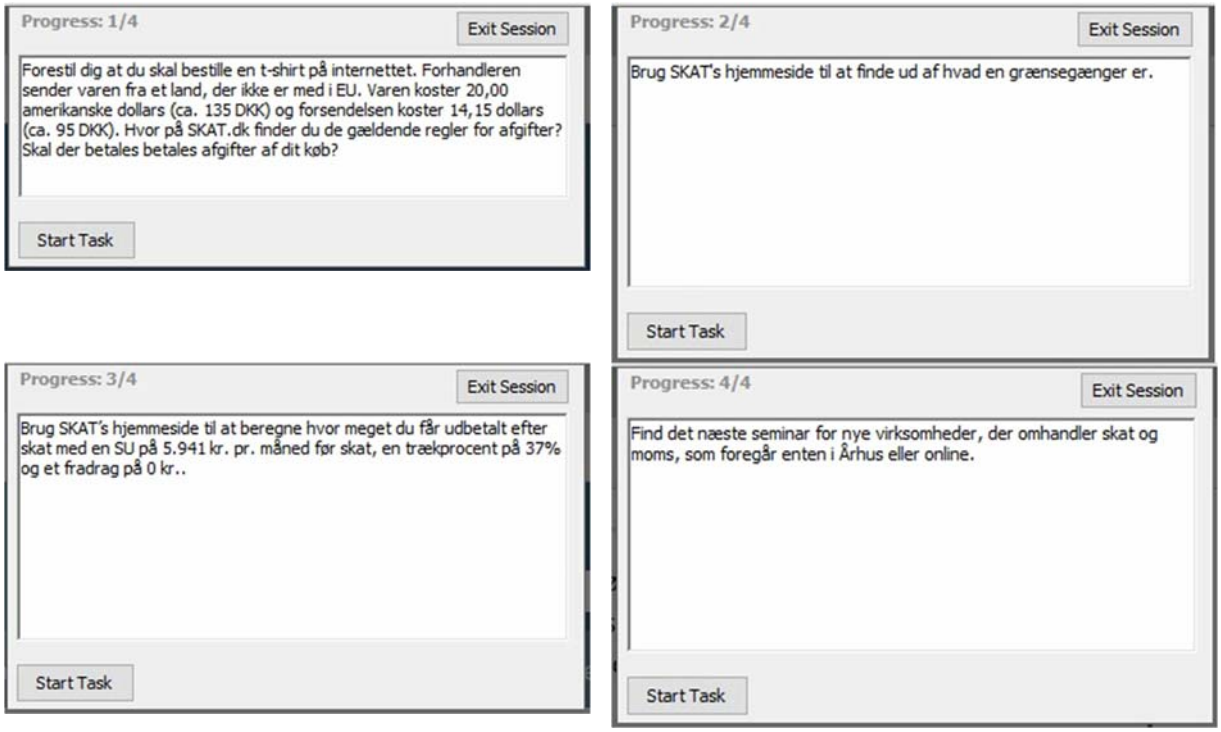


Figure 9Figure 9: Task Descriptions as Presented to the Participants.

These tasks were designed to reflect different information needs, and encourage different information-seeking behaviours. They also required participants to visit different pages of the website. Some tasks were designed so that participants were required to visit pages where the styling is notably different from that of the front-page of the website. This way participants were likely to experience a wider selection of types of subpages of the website.

Different goals were taken into consideration, as suggested in Hassenzahl's Three Level Hierarchy of Goals (Figure 6) as outlined in chapter 2, section 2.4.1. In the context of this

thesis the order in which the goals are considered is changed, so that the “What” and “How” take precedence over the “Why”, due to the nature of SKAT’s existing website. The do goal (“What”) in the tasks is to find specific information through the website.

Different motor goals (“How”) were considered. In the process of designing the tasks steps needed in order to find different information on the website were investigated. Some tasks were selected because the steps required to solve them were somewhat similar, while other tasks were selected, because the steps were significantly different from the other tasks.

Be goals (“Why”) were considered secondary to do goal and motor goal in the context of creating tasks for the User Experience Test, since the User Experience Test is not part of a design process. Furthermore, in the context of this thesis it is assumed to be unlikely that the motivating factor for using the e-government services is tied to emotion (e.g. to feel relatedness or to feel popular). However, when completing tasks on a website the user may experience some form of emotion. For instance, the user may feel competent, if the task is easily solved, or frustration, if a task is difficult to solve. Therefore, questions regarding these factors were incorporated in the test design. The list of questions can be found in TestQuestionsSkema.pdf and TestQuestionsResults.pdf.

Tasks 1 and 3 each have a specific end page and answer, which can be found by browsing through the menus of the website. Both tasks are difficult to complete through the website’s search function. The subjects of the information the users are required to find through these tasks were chosen because they were considered likely to be relevant to the selected user group.

In task 2 there is no clearly defined right answer. The task is completed, when the participant is able to make a description of the term that is deemed adequate by the researcher. It is difficult, if not impossible, to find definitions of the term without using the search function. The task design should thus allow collection of data regarding information seeking behaviour (as outlined in chapter 2, section 2.2.1).

The last task has a specific end-page, a calendar detailing where information meetings will take place. The answer can be found either through browsing or searching. However, using the term (“Seminar”), that participants are presented with in the task description, in a search string does not bring up the desired page in the search-results. The task was designed so that participants must either consider using synonyms for “seminar”, or find the answer through browsing the website. This was an intentional design decision, to see how users would act, when the system does not immediately provide usable results to a search.

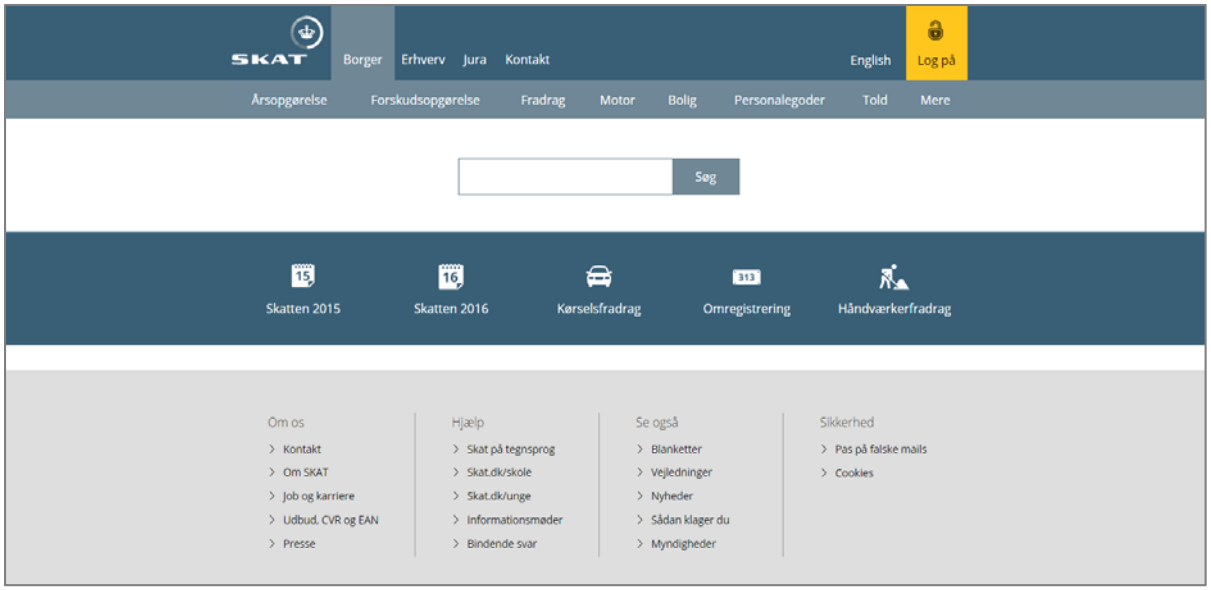


Figure 10: Frontpage of SKAT's Website

3.1.5.1 Path for Solving Task 1

In the first task participants are asked to calculate the amount of customs duty to be paid for items bought through a web-shop. It is possible to solve the task with six clicks, starting on the front-page (Figure 10). First the user must open the dropdown for the menu-point "Told" (Figure 11).

Figure 11: "Told" Dropdown Menu

In the dropdown the user must click the "Internethandel" menu item.

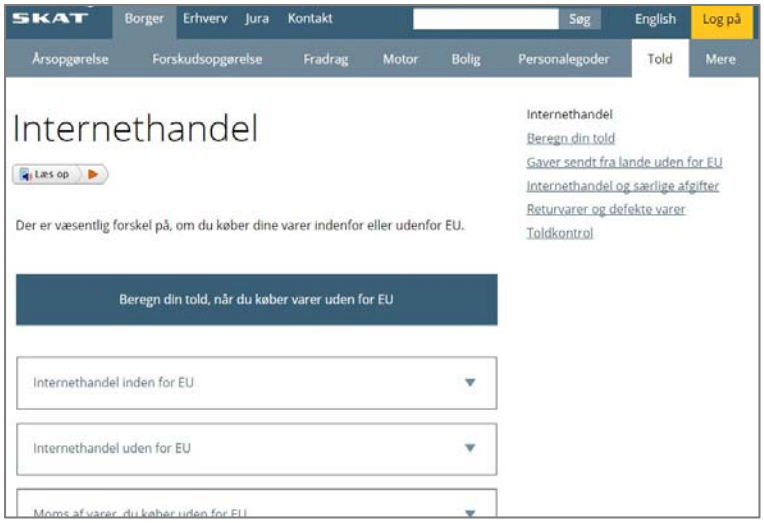


Figure 12: "Internethandel" Subpage

From the “Internethandel” subpage there are two ways to access the calculation form. Users can either click the “Beregn din told, når du køber varer uden for EU” box, or they can choose “Beregn din told” in the right hand list.

Figure 13: Step 1 of Choosing Product Type

In the “Beregn din told” sub-page the user must choose which type of product they intend to buy (Figure 13). Selecting “Tøj og sko” opens up the next step in the navigation. Here selecting “Tøj” (Figure 14) will open up another step, requiring the user to specify the type of fabric of the clothes (Figure 15).

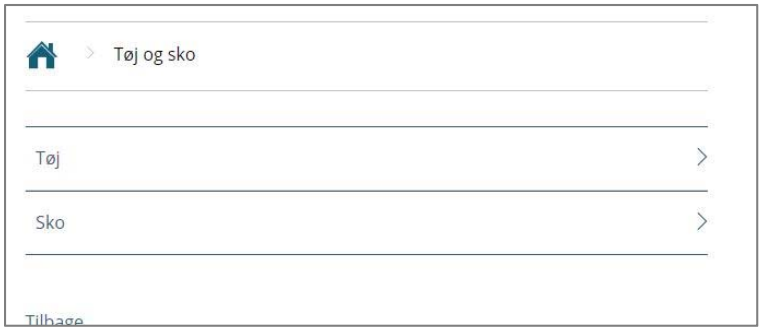


Figure 14: Step 2 of Choosing Product Type

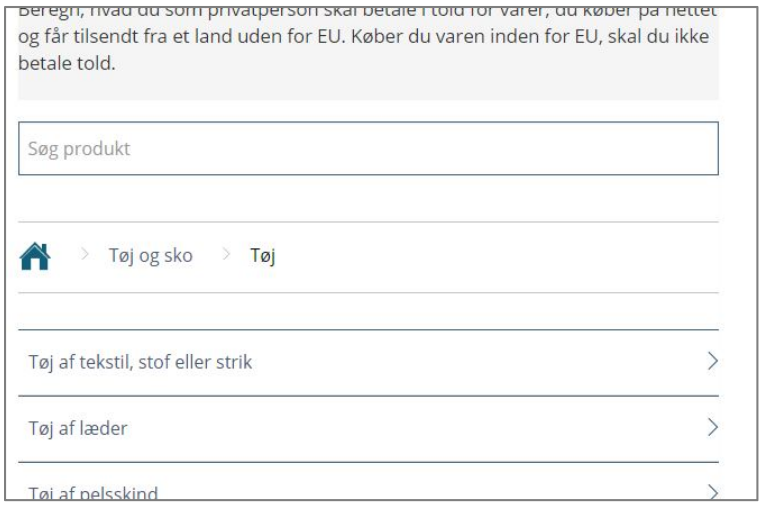


Figure 15: Step 3 of Choosing Material of Clothing

After the type of fabric has been specified, the user is presented with a form, in which the information specified in the task description must be entered (Figure 16).

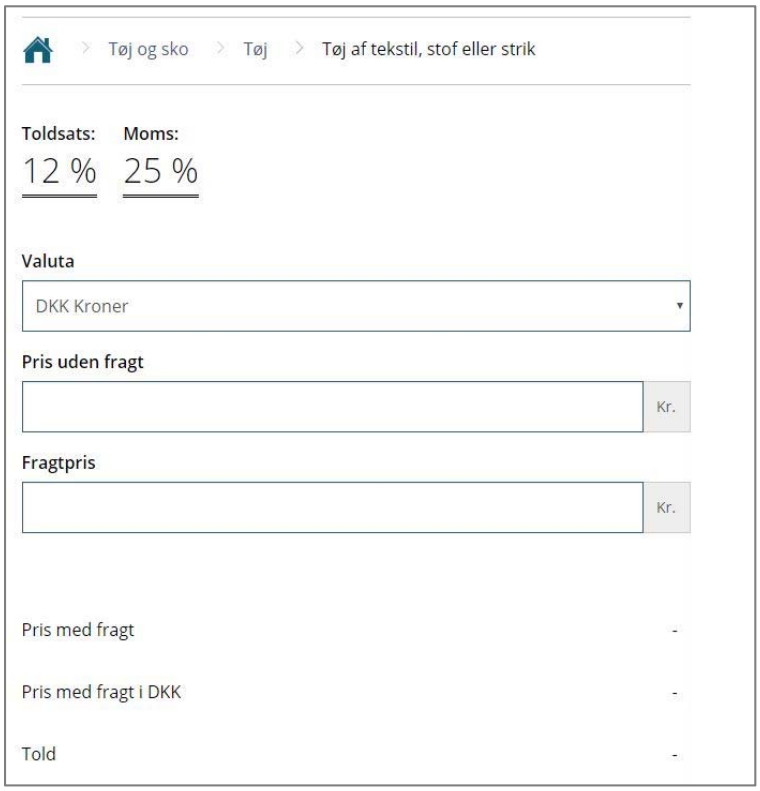


Figure 16: Calculate Amount of Customs Duty

3.1.5.2 Path for Solving Task 2

In task 2 participants must find information on what a transfrontier worker is.

This task can be solved with two clicks, as well as using the search bar. First the search bar is clicked, and then “Grænsegænger” is typed in the search field (Figure 17).

Figure 17: Searching for "Grænsegænger".

Choosing the first search result ("SKAT: Grænsegængerreglen") will lead the user to a subpage with information on the concept of transfrontier workers (Figure 19).

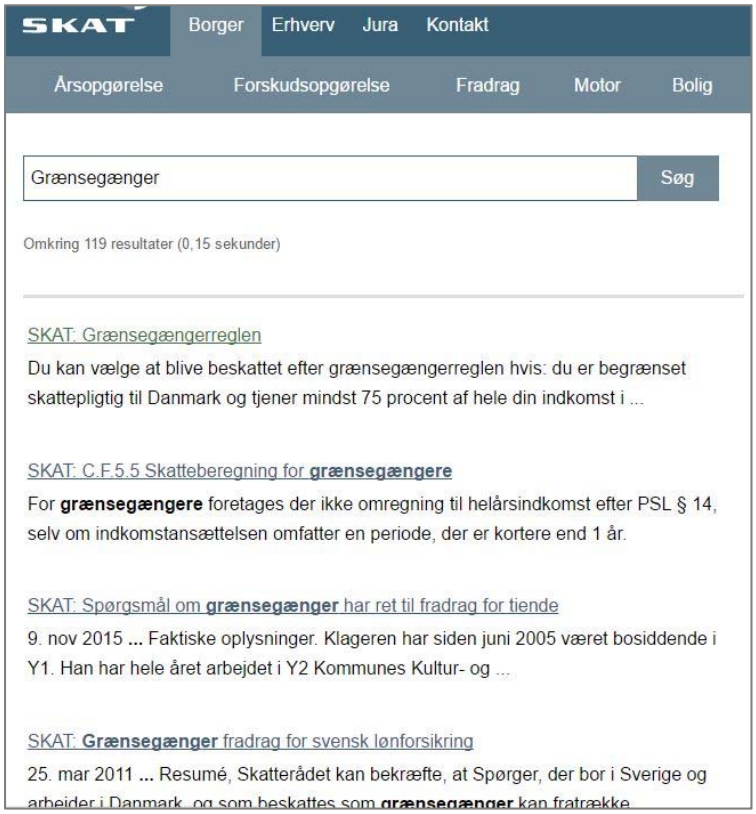


Figure 18: Search Results for "Grænsegænger".

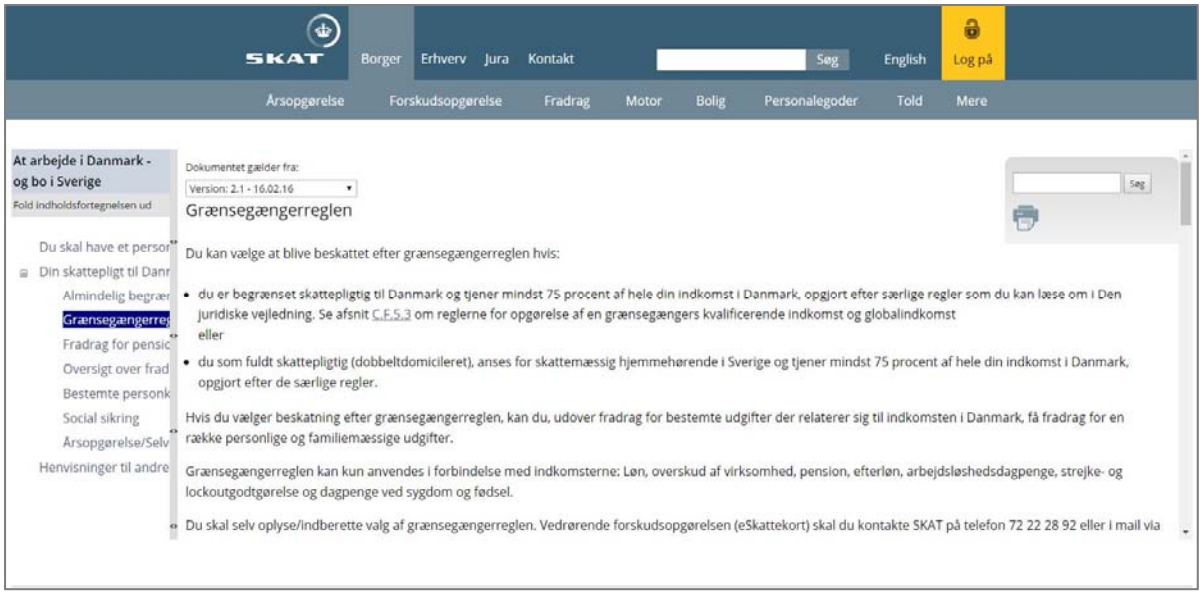


Figure 19: One of the Possible Ways to Find the Answer for Task 2.

3.1.5.3 Path for Solving Task 3

In task 3 the participant is asked to calculate the expected pay, after taxes, for a person receiving SU.

Task 3 can be solved in three steps from the frontpage. First the user must select the menu point “Forskudsopgørelse” to open the dropdown menu (Figure 20). Here “Regn ud, hvad du får udbetalt” is selected to open a sub-page, where the income information provided in the task description can be entered into the form (Figure 21).

Figure 20: Dropdown Menu for "Forskudsopgørelse"

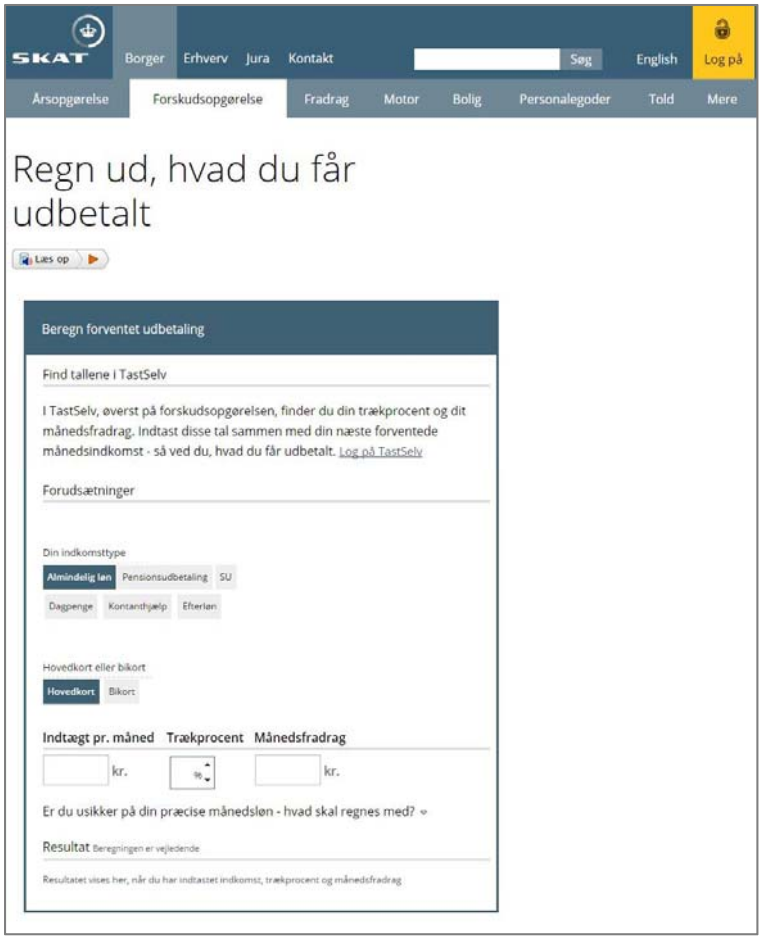


Figure 21: Form for Calculating Payment

3.1.5.4 Path for solving Task 4

In task 4 the participant is asked to find the date for the next seminar taking place in either Århus or on the web.

The task can be solved relatively easily by finding the link “Informationsmøder” in the page footer (Figure 22), and then clicking the “Aktuelle informationsmøder” link in the “Informationsmøder og webinarer” page (Figure 23), which links to the “Aktuelle informationsmøder og webinarer” subpage (Figure 24).

Figure 22: Footer Menu

Figure 23: "Informationsmøder" Subpage

Figure 24: Calendar Subpage

However, if the user does not think to look for the link in the page footer, there is another way to solve the task. Clicking the “Erhverv” menu point in the main menu opens up a submenu on the “Erhverv” subpage (Figure 25).

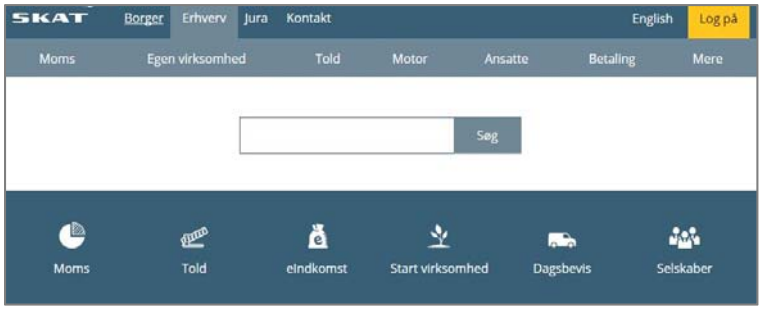


Figure 25: "Erhverv" Page

On the “Erhverv” page the “Egen virksomhed” menu point will open up a dropdown, from where the “Start virksomhed” menupoint leads to the “Start virksomhed” subpage (Figure 26). In the “Start virksomhed” page an accordion menu lists different subjects.

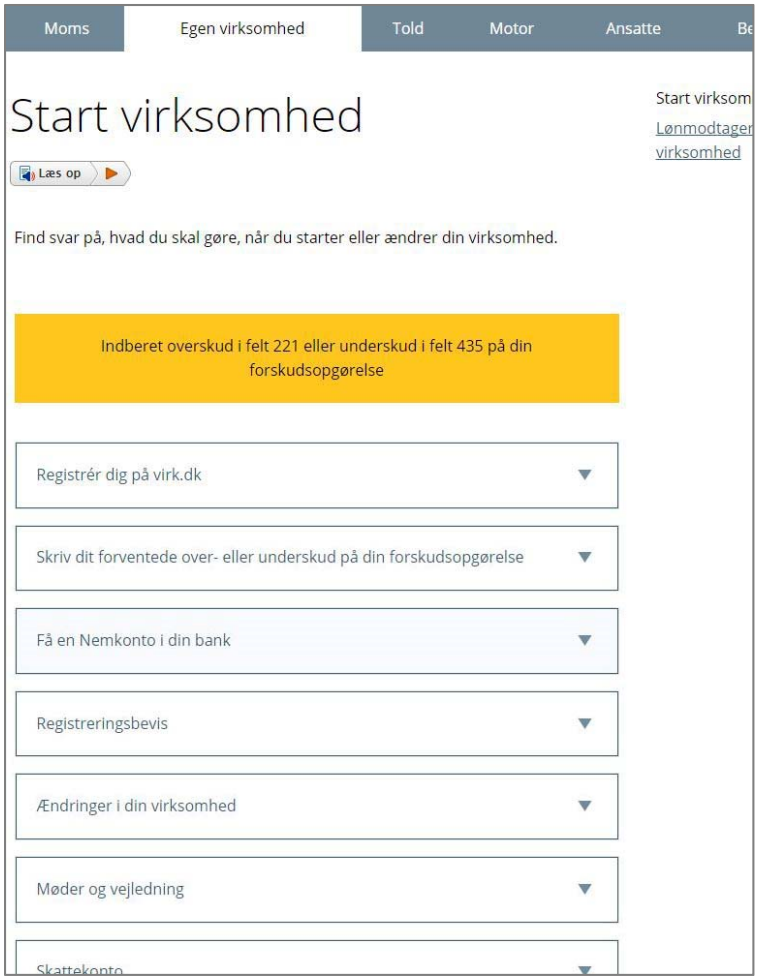


Figure 26: "Start virksomhed" Subpage

When “Møder og vejledninger” is clicked a list will become visible (Figure 27). The link in the first item of the list (“informationsmøder for nystartede virksomheder”) will open up the “Informationsmøder og webinarer” subpage (Figure 23) from where the calendar can be accessed.

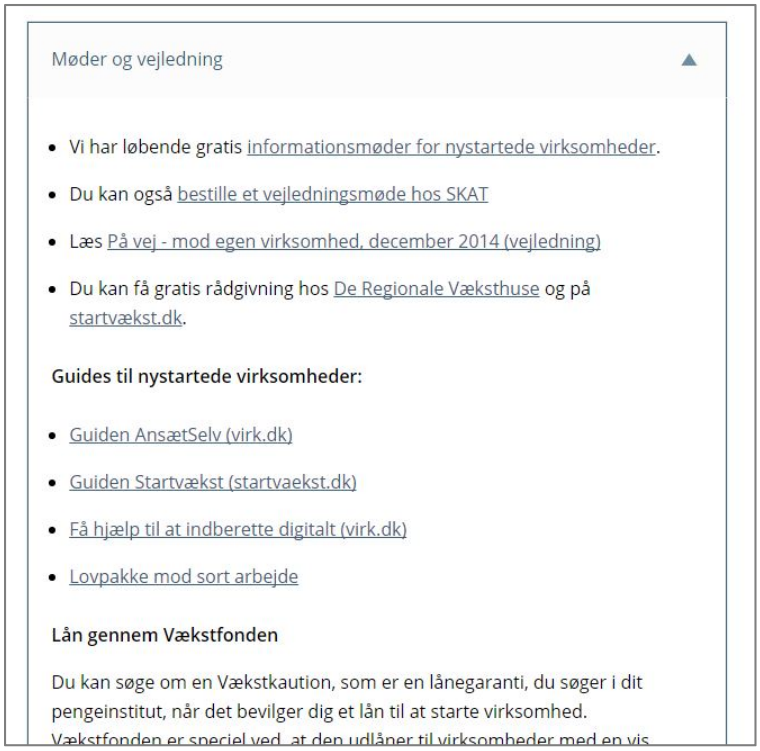


Figure 27: "Møder og vejledning" Accordion Menu

3.1.5.5 Questions for tasks

After each task participants were asked to verbally rate the experience of the task. Their answers were noted by the researcher, along with whether or not they solved the individual tasks. In some cases, short comments were added as well, for instance one participant stated at the start of the test that she was dyslexic, and another participant was confused by Morae's user interface and ended a task without reading the task description.

3.1.6 Morae

The software used for the User Experience Test was Morae Recorder. Morae can be used for various types of testing like User Experience testing. The program can record input from the computer's web camera, keyboard, mouse, and microphone. A test in Morae can be set to automatically present tasks, and surveys, to the test participant. In principle this allows the participant to complete tasks without much interference from the person conducting the test. Morae Recorder can be linked to Morae Observer, allowing the facilitator to control the presentation of tasks and surveys to the test person, without the facilitator being present in the same room as the test person. However, only Morae Recorder was used in this test. Having the researcher present during the test enabled the researcher to ask for verbal information on participants' thoughts and experiences.

During the test data on participants' age, gender, and when they last visited the homepage of SKAT was collected, as well as footage from the computer's webcam, footage captured from the screen of the computer, mouse-click data, audio recordings of the tests, and time taken to complete each task. After each task was completed participants were asked to indicate whether or not they thought the specific task was easy or difficult and simple or frustrating, on a scale from 1-5, and whether or not they were familiar with the subject of the question before starting the specific task. After completing all tasks, the participants were asked to fill out a User Experience Questionnaire.

3.1.6.1 Limitations of Morae Recorder

Randomising tasks was considered, in order to test if task sequence would influence participants' ability to solve individual tasks. However, as randomising tasks is not possible in Morae a fixed sequence was decided upon, based on pilot tests conducted prior to conducting the final User Experience Test.

The User Interface of Morae presented some challenges. For instance, the font used in the interface is small, both when Morae presents the user with task descriptions, and on the buttons used for controlling the progress in tasks: "Start Task", "Show Instructions/Hide Instructions", "End Task" (Figure 28-Figure 30). There was no option to mark a task as unsolved, possibly because Morae Observer was not used to start/end tasks remotely. Once a task has been ended it is not possible to return to that task. This means that if a task is marked as ended by mistake, there are two options: Either go through the whole test again, or proceed to the next step of the task without data on the task that was incorrectly ended.

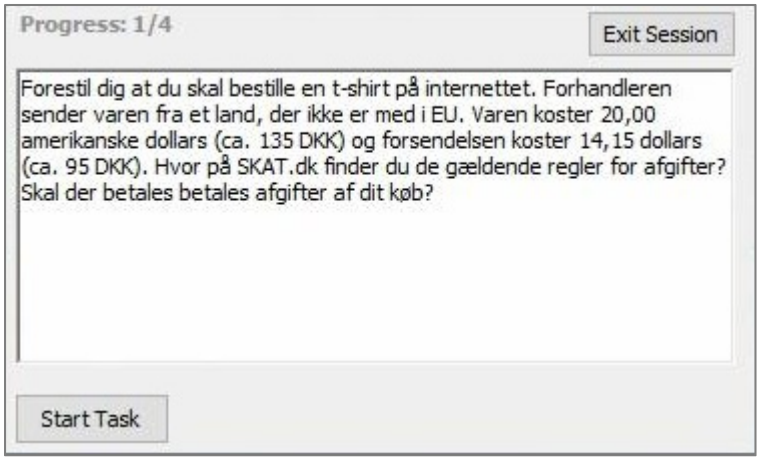


Figure 28: Morae User Interface - Start Task

Figure 29: Morae User Interface - Instructions Hidden

Figure 30: Morae User Interface - Instructions Visible

The styling of the surveys created in Morae is not centered, resulting misalignment of scale items on the 7-point scale used in the User Experience Questionnaire (Figure 31). Morae automatically numbers questions. When creating the items in Morae it was not possible to leave the "Question" field empty. However, the nature of the User Experience Questionnaire is such, that no question is asked for the individual items, the user simply chooses a number on a 7-point scale. As a result, each item was numbered twice, automatically by Morae, and manually during the test-design (Figure 31). In addition, the alignment of the different items on the 7-point scales varied due to the styling used by Morae.

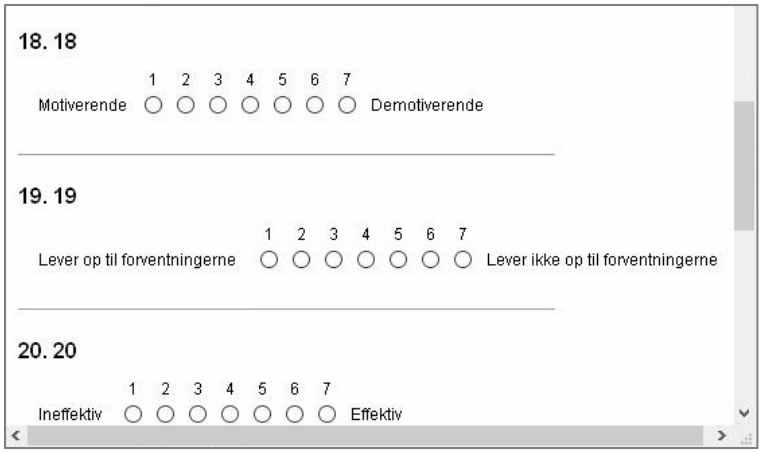


Figure 31: Morae User Interface - User Experience Questionnaire

Furthermore, Morae could not get access to the microphone input. This necessitated the use of an external audio recorder.

3.1.7 User Experience Questionnaire

The User Experience Questionnaire was used as a part of the User Experience Test. According to Laugwitz, Held and Schrepp (2008, p. 63) questionnaires can be an effective way to collect quantitative data to measure product features. However, it is also stated (ibid.) that user questionnaires should in general be combined with other assessment methods, in order to interpret the results. Thus, the questionnaire was incorporated into the User Experience Test, so that the test combined different data collection methods.

The questionnaire was created in 2005 (Schrepp, 2015, p. 1). It was created by having Usability experts collect as many terms and statements as possible on User Experience and Usability (Laugwitz et al., 2008, p. 63). Through a selection process a set of 26 questions, also referred to as “items”, were chosen for the final questionnaire (Laugwitz et al., 2008; Schrepp, 2015). The questions cover different aspects of User Experience and Usability; Attractiveness, Perspicuity, Efficiency, Dependability (ergonomic quality aspects), Stimulation, and Novelty (hedonic quality aspects). The aspects are also referred to as “scales”.

Schrepp (2015, p. 2) explains the different aspects:

- **Attractiveness** covers the overall impression of the product.
- **Perspicuity** covers users’ ability to learn to use the product.
- **Efficiency** regards whether or not users are able to solve their tasks easily.
- **Dependability** measures users’ perception of whether they are in control of the interaction.
- **Stimulation** concerns whether the product is exciting and motivating to use.
- **Novelty** regards whether or not the product is interesting to the user, and whether it is innovative and creative.

The questionnaire was created from a framework, where perceived ergonomic quality, perceived hedonic quality, and perceived attractiveness are distinguished between

(Laugwitz et al., 2008, p. 65). Ergonomic quality focuses on “the goal oriented or task oriented aspects of product design”, which allow users to efficiently achieve their goals, while hedonic quality concerns the “non-task oriented quality aspects of a software product”, meaning the originality and attractiveness of the user interface and design of the product (ibid.). Laugwitz et al. (ibid., p. 66) further state, that the perceived attractiveness results from “an averaging process from the perceived quality of the software concerning the relevant aspects in a given usage scenario”. On this basis the User Experience Questionnaire was constructed to measure the perceived attractiveness of the product, as well as the quality of relevant aspects of the product (ibid.).

Because there is a tendency that people choose neutral options when presented with a scale the creators used a seven-point scale for the User Experience Questionnaire (ibid.). Laugwitz et al. (2008, p. 65) state the importance of allowing the user to rate the product “as immediately and spontaneously as possible”, while an “explicit evaluation demanded by the user retrospectively” may not be reliable. Therefore, participants were asked to fill in the questionnaire as soon as they had answered the questions about the last task.

Scale	English translation	
Attractiveness	annoying	enjoyable
Perspicuity	not under-standable	understandable
Novelty	creative	dull
Perspicuity	easy to learn	difficult to learn
Stimulation	valuable	inferior
Stimulation	boring	exiting
Stimulation	not interesting	interesting
Dependability	unpredictable	predictable
Efficiency	fast	slow
Novelty	inventive	conventional
Dependability	obstructive	supportive
Attractiveness	good	bad
Perspicuity	complicated	easy
Attractiveness	unlikable	pleasing
Novelty	usual	leading edge
Attractiveness	unpleasant	pleasant
Dependability	secure	not secure
Stimulation	motivating	demotivating
Dependability	meets expecta-tions	does not meet expectations
Efficiency	inefficient	efficient
Perspicuity	clear	confusing
Efficiency	impractical	practical
Efficiency	organized	cluttered
Attractiveness	attractive	unattractive
Attractiveness	friendly	unfriendly
Novelty	conservative	innovative

Figure 32: English User Experience Questionnaire Items

As the User Experience Questionnaire was originally created in German and translated into English, and the participants in the present User Experience Test were Danish, a Danish translation of the questionnaire was made for this test. The translation was based on both the original German version and the English translation, in an effort to best convey the intended meaning of each item. The English translation of all 26 items can be seen in Figure 32, and the Danish translation can be seen in Table 1.

Table 1: Danish translation of User Experience Questionnaire

Item	Danish translation	
1	Irriterende	Behagelig
2	Uforståelig	Let forståelig
3	Kreativ	Fantasiløs
4	Let at lære	Svær at lære
5	Værdifuld	Inferiør
6	Kedelig	Spændende
7	Uinteressant	Interessant
8	Uforudsigelig	Forudsigelig
9	Hurtig	Langsom
10	Opfindsom	Konventionel
11	Hindrende	Støttende
12	God	Dårlig
13	Kompliceret	Enkel
14	Utilfredsstillende	Tilfredsstillende
15	Almindelig	Nyskabende
16	Ubehagelig	Behagelig
17	Sikker	Usikker
18	Motiverende	Demotiverende
19	Lever op til forventningerne	Lever ikke op til forventningerne
20	Ineffektiv	Effektiv
21	Let forståelig	Forvirrende
22	Upraktisk	Praktisk
23	Organiseret	Overfyldt
24	Attraktiv	Uattraktiv
25	Venlig	Uvenlig
26	Konservativ	Innovativ

While the questionnaire was originally designed for collecting quantitative data, it was also used as a way to get users to talk about their experience of the website. Participants were encouraged put into words how they had experienced the website, while filling out the questionnaire. The participants were presented with the questionnaire immediately after finishing the interactions with the website.

3.1.8 Pilot Testing

Prior to conducting the final tests, a series of pilot tests were conducted, as suggested by Kelly (2009, p. 60) and Rubin and Chisnell (2008, p. 215). In addition, the test design and tasks were the subject of ongoing discussion and feedback, with the supervisor for the project.

Rubin and Chisnell (2008, p. 215) suggest that “real” participants (i.e. the actual users of the system) be used for the pilot tests. However, employees may be recruited as participants as well, if absolutely necessary. They further state that the entire test should be conducted during the pilot test, just as it will be in the final tests. In accordance any questionnaires included in the test design should be filled out by participants. This way all relevant data collection methods can be practiced. The importance of conducting at the very least one

pilot test is strongly emphasised by Rubin and Chisnell (ibid.), lest the first couple of actual tests become the place where flaws in the test design becomes apparent.

The pilot tests were independent of each other, and were conducted with three separate participants, at three separate times. This was done in order to get an estimate on how much time the participants were likely to spend completing the test. Furthermore, the pilot tests served to uncover problems with the task descriptions, and the last pilot test was used for testing, and getting familiar with, the equipment used for the final tests, as it is suggested by Rubin and Chisnell (ibid.).

In the first two pilot tests tasks were presented to participants on paper, while time was recorded manually with a digital watch. In all pilot tests the sequence of the tasks was randomised. This was done to see whether or not the sequence in which the tasks were presented was likely to affect participants' ability to solve tasks. After each session the tasks were discussed with the participants to uncover any difficulties in interpreting the goal of the task. Small adjustments were made to the final design, due to the findings of the pilot tests. Changes included removing one task from the test design, partly due to time issues, and partly because it did not seem to contribute significantly to the findings. A printed template for the questions about the individual tasks was made, to ensure consistency in the written notes, and to make sure questions were asked in the same order.

Among other things it was found that randomising task sequence was not possible in the program chosen for conducting the tests. Therefore, a fixed sequence was decided upon. Task 1 and 3 (in the final test design described in section 3.1.5) have similar solutions (i.e. finding calculators in different parts of the "Borger" subpage). For that reason, it was decided that these two tasks should be separated by one task (Task 2, finding a definition for "Transfrontier worker"). From the pilot tests it was assumed that participants would likely use the search function of the website to solve Task 2.

3.1.9 Procedure

The tests were designed to be carried out individually, one participant at the time. Before the start of each test there was a short oral introduction to the test design, where the participants were asked whether they would consent to participating. Participants were then asked to sit in front of the computer to begin the test, and audio-, video-, and screen capture recordings were initiated.

The participants were first presented with short a questionnaire (Figure 33). Here they were asked about age, gender, and how recently they have visited SKAT.dk.

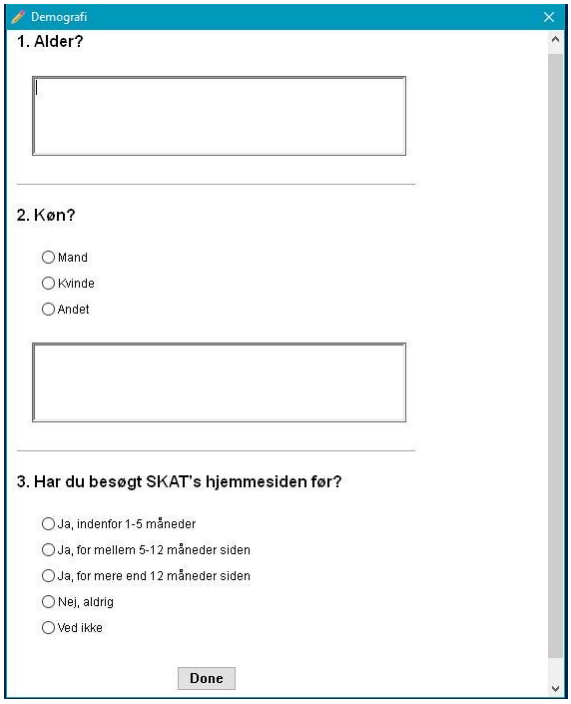


Figure 33: Demography

After each task participants were asked to verbally rate the experience on two scales, ranging from 1-5, one scale using the terms “Easy” to “Difficult”, and the other “Simple” to “Frustrating”. The answers were written down by the researcher, along with a few notes on interesting observations or insights that had not been taken into account during the design of the test (Figure 34). In addition, they were asked whether they knew the problem they were asked to solve beforehand (e.g. if they had ever had to calculate customs duties on goods bought outside of the European Union).

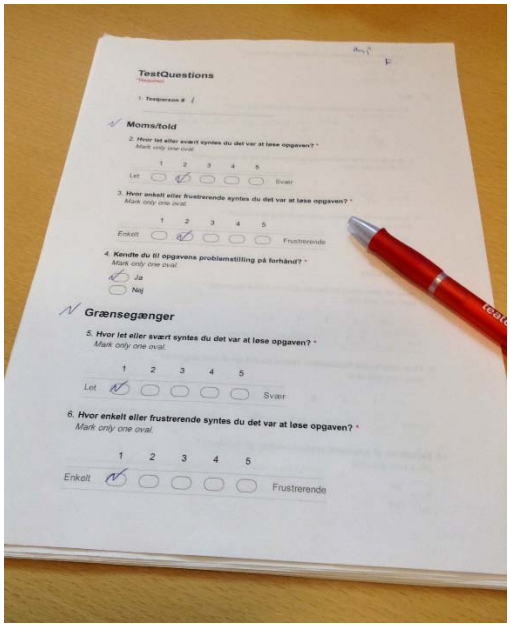


Figure 34: Questions for Individual Tasks

When all the tasks had been solved, participants were presented with the User Experience Questionnaire (Figure 35). The tests were recorded using Morae Recorder, as well as a voice recorder.

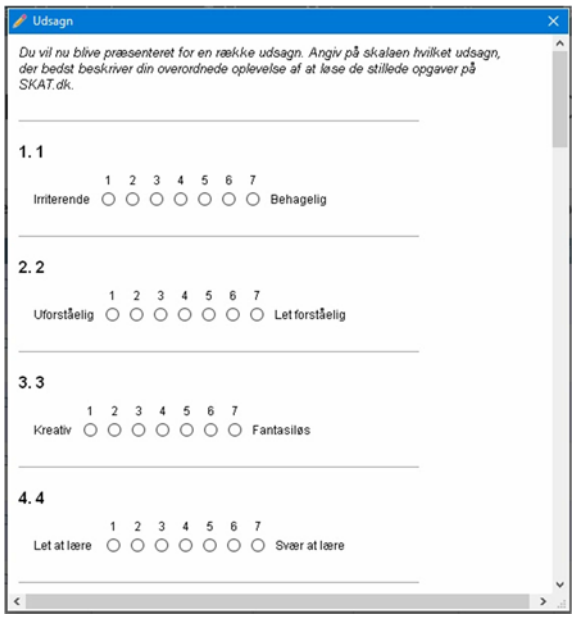


Figure 35: User Experience Questionnaire in Morae Recorder

Participants were asked to solve tasks in order to make sure they have a recent first-hand experience of using the website. As described in section 3.1.5 tasks were designed so that they reflected different aspects of using the website. For instance, some tasks could be solved by using the navigational elements of the website, while another task could only be solved through the website's search function. Phrasing of task presentations was considered as well, thus some task descriptions contained keywords that were known to provide good search results, whereas other task descriptions did not contain keywords recognised by the page's search function. A set of different tasks were created and tested. From these tasks four tasks were selected for the final User Experience Test, that would lead participants to explore different sub-pages of SKAT.dk, and use different means of navigation in order to solve the tasks.

In order to take into consideration whether there was any relation between possible frustration or difficulties with solving tasks, and bad overall experience with the webpage participants were asked to rate the experience of each task individually, in addition to rating the overall experience of using the website in the User Experience Questionnaire.

3.1.10 Use of Recordings

Audio-, video-, and screen capture-recordings were used to supplement the data collected through the questionnaires and notes taken during the test. After a preliminary data-analysis was made, the recordings were played back. Statements, made by the participants, that could shed light on aspects relevant to answering the problem statement, where written down.

3.1.10.1 Transcription

A few of the audio-recordings from the user experience tests were transcribed (Appendix 1). However, only sequences that were deemed relevant for answering the problem statement were transcribed. The audio-files have been included in the folders called AudioParticipant_1-5, AudioParticipant_6-10, and AudioParticipant_11-15.

Halcomb & Davidson (2006, p. 38) state that verbatim transcription is "word-for-word reproduction of verbal data, where the written word is an exact replication of the audiorecorded words". According to Poland, 1995, p. 290 it seems to be widely acknowledge that transcripts must be verbatim, however, Poland claims that this is often not an empirically verified claim. Poland (ibid., p. 291) further states, that much of the nonverbal communication during an interview session are not captured by audio recorders. As Kvale (2007, p. 93) states, these nonverbal cues are not accessible to the reader of the transcript. Furthermore, the verbal cues to emotional state that may be recorded, are difficult to transcribe (Poland, 1995, p. 292). Kvale (2007, p. 92) calls transcription "an interpretive process, where the differences between oral speech and written texts give rise to a series of practical and principal issues". If verbatim transcription is sought, the transformation of interview data from oral to written form, may render the transcription inadequate for both oral and written text (Kvale, 2007, p. 92).

Extensive note taking may lighten the task of transcription. However, Kvale (2007, p. 94) states that: "Taking extensive notes during an interview may, however, be distracting, interrupting the free flow of conversation", whereas notes written after the interview may fail to record important details (Britten, 1995). Britten further estimates that transcription of an hour's interview may take six-seven hours to transcribe. Transcription is thus extremely time consuming, as is also noted by Kvale (2007, p. 95), and Fasick (1977-1978, p. 549). Wellard and McKenna (2001, p. 182) claim that transcription "can be either beneficial or problematic for the researcher, depending on the situation". On one hand it may give the researcher in depth knowledge of what was said in the interview and how (Ibid., p. 182-183), although, this insight may prove difficult to reflect in written form. On the other hand, it may be costly (ibid., p. 182).

Due to the estimated time needed to do verbatim transcription of the audio-recordings, as well as an expected limited value of extensive transcription of whole sessions, only small parts of the audio-files, deemed to add to the understanding of individual participants' experience of the User Experience, were transcribed.

A very simple transcription key was created to transcribe by. The transcriptions (Appendix 1) are written as comprehensible text, and are not phonetically accurate representations of the spoken content of the recordings. However, care has been taken to make them as accurate as possible with regards to what the participants are saying.

Table 2: Transcription key used for transcriptions

Transcription key	
[…]	Content left out
…	Pause

3.2 Data Analysis

The analysis presented in this thesis is mainly descriptive. An overview of the demographic data collected (the participants gender, and age distribution) will be made, as well as data on when participants last visited SKAT.dk. In addition, data on how much time participants spent solving each task and the average time needed pr. task, as well as which tasks the individual participants solved (or left unsolved) will be presented. Connections between these different aspects will be explored, such as whether gender of participants has any influence on time spent solving tasks and completion rate.

It will be examined whether there is reason to believe that difficult tasks are rated as more frustrating than easy tasks (i.e. level of frustration versus level of difficulty). Further, it will be examined whether there is a connection between task status (solved/unsolved) and level of frustration, and whether the time spent on solving a problem is related to the subsequent rating of the difficulty and frustration level of that task. An overview of completion time and task status (solved/unsolved) distributed by gender will be given as well.

3.2.1 Analysis of the User Experience Questionnaire Data

The developers of the User Experience Questionnaire have also included an analysis tool for the questionnaire, with the goal of making the analysis of the data as easy as possible (Schrepp, 2015, p. 8).

First the results of the collected answers from the User Experience Questionnaire will be entered into the User Experience Questionnaire Data Analysis Tool Excel worksheet (found in the RawData folder, file name: UEQ_Data_Analysis_SKAT). The analysis tool automatically processes the entered data in a number of ways. Since the scale order in the User Experience Questionnaire is randomised (i.e. low ratings do not necessarily indicate a positive value, and vice versa) the data analysis tool orders the answers before processing the data (DT tab in the worksheet, Figure 36). The Analysis Tool converts the rating values from values of 1-7 to values of -3-3, where low values indicate a negative rating.

When the data has been added to the data analysis tool it is processed in different tabs in the excel worksheet. The main results are shown in the “Results” tab. No overall score is calculated across the six different scales, since that would not produce any meaningful insight.

Figure 36: Tabs in the Data Analysis Tool

The analysis tool calculates the mean correlation of the different items in the six scales of the User Experience Questionnaire (ibid., p. 9). Means are calculated for each item, based on the scale the item belongs to (e.g. item 1 and item 12 both belong to the Attractiveness scale). Means for the six scales are calculated as well, and the means measure the consistency of a scale. The higher the mean value for an item, the more consistent the answers for that item are. Schrepp (ibid.) states that a small value for one scale may indicate, that the scale should be interpreted carefully. It is further suggested that a value of >0.6 or >0.7 may be sufficiently high.

If a high number of participants misinterpret the meaning of an item in a scale, this may lead to small mean values for that item. For instance, in the context of web the item unsecure or secure may be misinterpreted by some participants. The specific item belongs to the Dependability scale, and should measure whether participants find that interaction with the system is safe and controllable. However, in the context of web it may be misinterpreted as measuring how safe the users' data is in the system. Another explanation for low mean values can be that a specific scale may not be relevant in the context of the tested system.

3.3 Ethical Considerations

The participants were not required to log into the system with their NemIDs in order to solve the tasks. This decision was made partly because the interaction would be recorded, and thus personal information would have been recorded as well, and partly because requiring login might complicate the recruitment process. No data that would allow users to be identified after the test was collected during the test, such as names, occupation, or address. However, webcam footage of participants interacting with the system was recorded during the test. The webcam feed will only be accessed by the researcher responsible for conducting the test, in accordance with the oral agreement between researcher and test participant. The intended use of the data collected was disclosed verbally to the participants, and they were asked if they were willing to participate, before the tests were conducted. This was not recorded, since they had not yet agreed to be recorded. The audio recordings of the tests can be found as mp3-files in folder named Audio. The audio recordings contain no personal info. Recordings from Morae, where the webcam recordings have been removed, are included as mp4-files in folder named Video.

Since some test participants were initially expected to be high school students it was decided that all test participants must be at least 18 years old. Recruiting participants under the age of 18 years would require their parents or guardians signature. The space in which the tests were to take place was carefully considered. In the case of the high school students it was decided that tests should take place at their school. However, no high school students were tested in the end.

It was deemed important, that participants should not need to go far in order to participate, since the benefit of participating for them was small. However, it was also decided that the tests should not take place at their actual homes, since that may be perceived as an intrusion into their private lives. These considerations contributed to the decision of setting up tests equipment in places were the participants would be present regardless of participating in the test or not, such as Skjoldhøj Kollegiet.

Most of the tests were conducted at Skjoldhøj Kollegiet, either in the common room or in the apartment of the recruiter. This was considered as places were researcher and participant were on fairly equal terms. Two of the tests took place at other spaces, that had been agreed upon by both parties beforehand.

4 Results of User Experience Test

15 participants were successfully recruited for the test. They were all native Danish speakers. One participant was dyslexic.

A large part of the participants are residents at Skjoldhøj Kollegiet. One participant is an acquaintance of the researcher. Participants had vastly different backgrounds, both in regards to city of origin, as well as educational background, and occupation. Most participants were not asked about their occupation. However, some participants' occupations were discussed outside of the test session. It is assumed that most participants were students, as being a student is a requirement of living at Skjoldhøj Kollegiet. Of the participants were occupation was discussed, some were students at a variety of educational institutions in Aarhus, one is occupied as a web-developer, and another is living on social benefits.

Most of the test participants were tested at Skjoldhøj Kollegiet. One was tested at a mutual friends' apartment, one was tested at the apartment of the researcher, three were tested in the common room at Skjoldhøj Kollegiet, and the rest were tested in the room of the recruiter at Skjoldhøj Kollegiet.

The raw data can be found in the folder named RawData.

4.1 Demography

The gender ratio of the test participants was almost 50/50, with eight women and seven men, as illustrated in Figure 37.

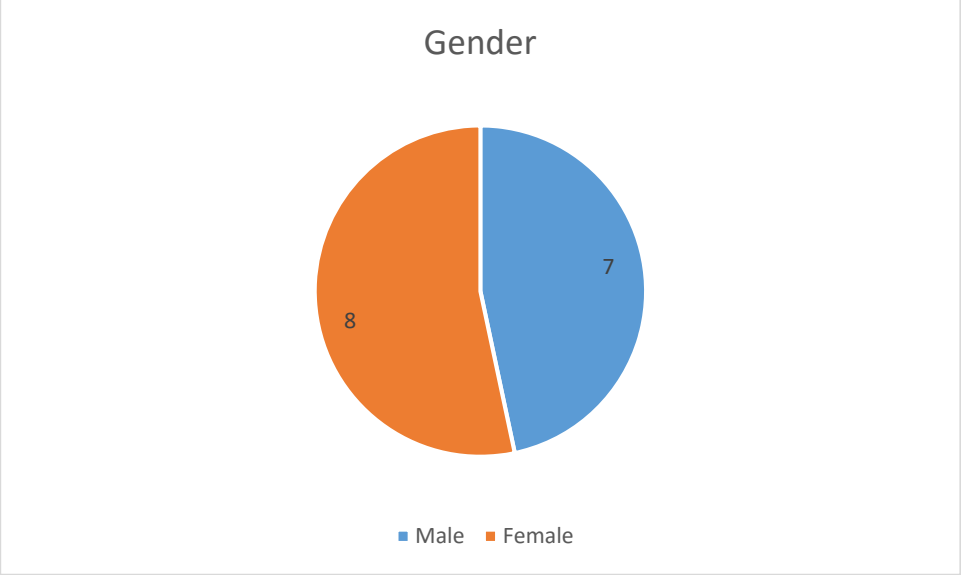


Figure 37: Gender of Participants

Participants were between 20-26 years old. 22 year olds were the largest group represented in the test with five participants (Figure 38). The next largest group were 25 year olds (three participants), followed by 26 year olds and 24 year olds (with two participants). 20, 21 and 23 year olds were each represented by one participant.

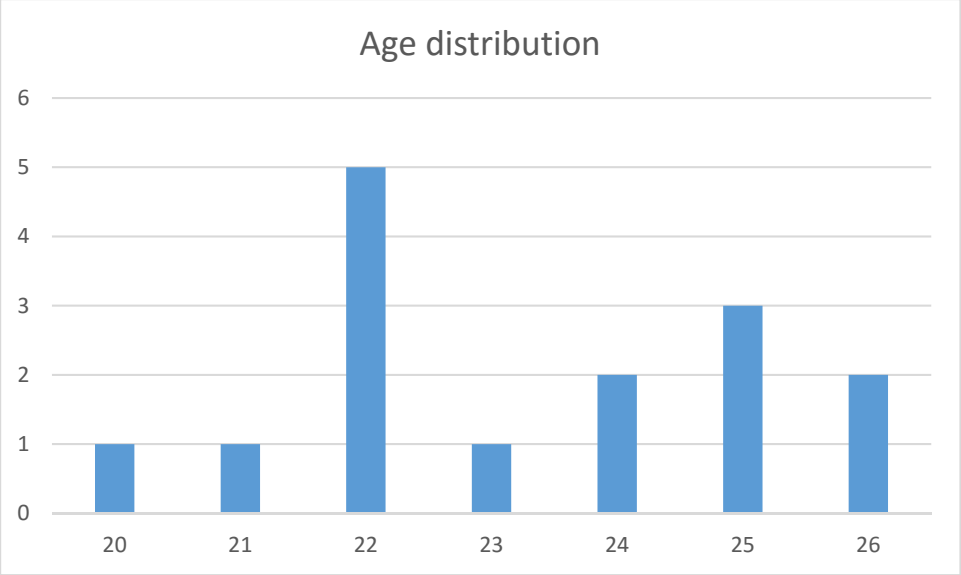


Figure 38: Age Distribution Among Participants

As shown in Table 3 most participants had visited SKAT's website within the last 5 months, while one had visited within the last year, one visited more than a year ago, and one did not remember when the last visit occurred. Since the number of participants, who did not report that they had visited SKAT.dk within the last 5 months, was quite small it is difficult to make relevant conclusions based on this information. However, there does not seem to be any connection between the overall performance of the participants that reported they had visited SKAT.dk recently, and those who did not.

Table 3: Participants' Answers to the Question "Have you ever visited SKAT's website before?"

	Yes, in the last 1-5 months	**Yes, in the last 5-12 months**	**Yes, more than 12 months ago**	**No, never**	**Unknown**
Participant 1	X				
Participant 2	X				
Participant 3	X				
Participant 4	X				
Participant 5			X		
Participant 6	X				
Participant 7	X				
Participant 8	X				
Participant 9	X				
Participant 10		X			
Participant 11	X				
Participant 12	X				
Participant 13					X
Participant 14	X				
Participant 15	X				

4.2 Task Solution

Task 1 was solved by thirteen out of fifteen participants, Task 2 and 3 was solved by fourteen participants each, while Task 4 was solved by nine participants (Figure 39). For Task 1-3 the variation in the number of participants that solved the task is not noteworthy. However, Task 4 was solved by significantly fewer participants than the other tasks. It should be noted that one participant was excluded from the data on Task 4, since the participant ended the task without reading the task description.

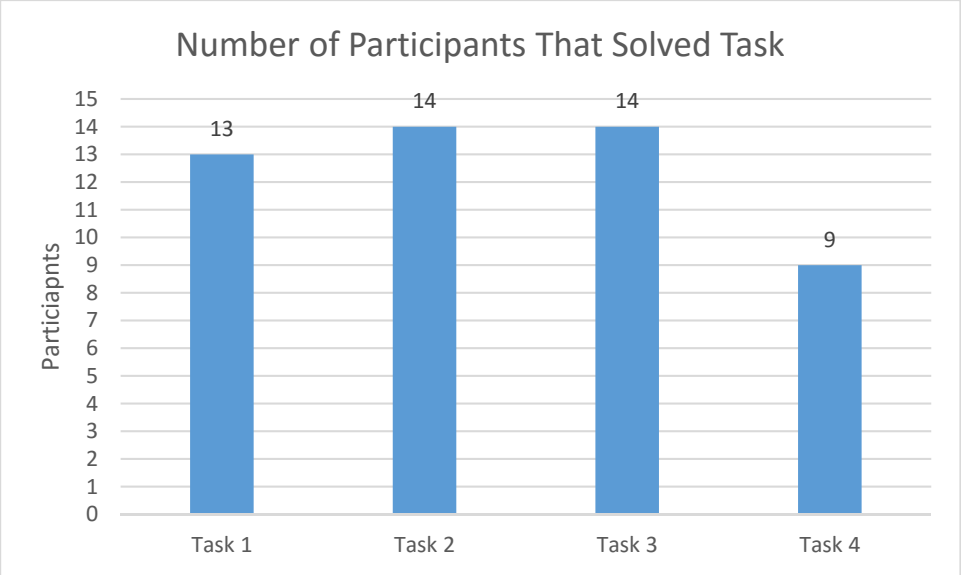


Figure 39: Number of Participants That Solved Task

Figure 40 shows which participants solved which tasks. Two participants did not succeed in solving Task 1. It was noted that both of these participants (9 and 13) were confused by Morae's interface. Participant 13 solved only Task 3. Participant 6 was the only participant that did not solve Task 3. It was noted that participant 6 is dyslexic. Of the fourteen participants nine solved Task 4. Participant 2 did not end the task at the same subpage as the other participants, but the task was categorised as solved, since he did find a date for a seminar in Aarhus. It should be noted that participant 13 ended the task before reading the task description.

Of the fifteen participants eight solved all the tasks, while five participants solved three tasks. One participant solved two tasks, and another participant solved one task. Participant 6 and 13 stand out somewhat from the general results, as they are the only participants to solve fewer than three out of four tasks.

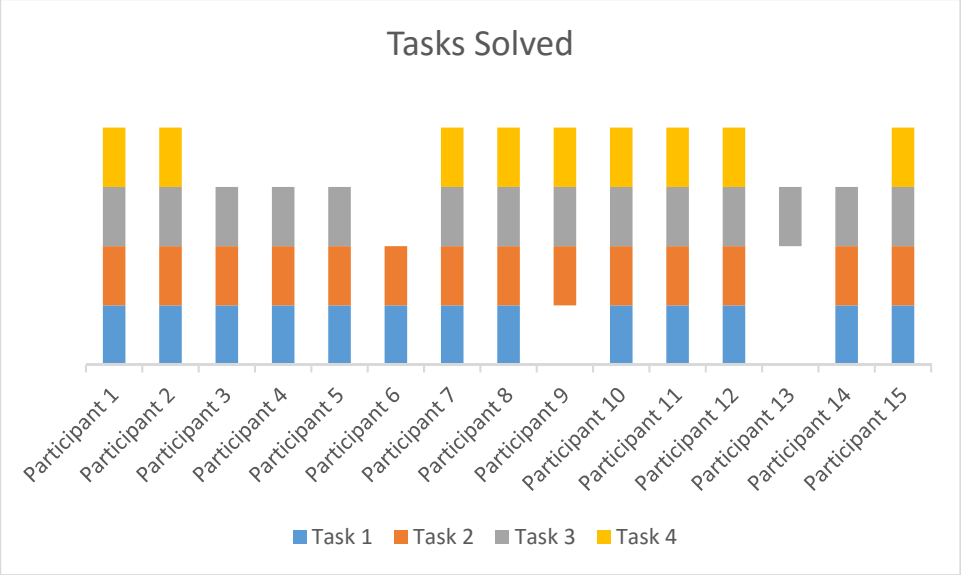


Figure 40: Tasks Solved by Participants

Figure 41 shows how many participants solved the individual tasks, divided by gender. Task 1 was solved by seven female participants and six male participants, Task 2 was solved by eight female and six male participants, Task 3 was solved by the same number of participants (seven) by both genders, while five female participants and four male participants solved Task 4. It should be noted that one male participant solved one out of four tasks, which is fewer than any other participant.

In general, there does not seem to be any reason to believe that there is any connection between the participants' gender and their ability to solve the tasks. It should be noted, that there was one more female participant than the number of male participants.

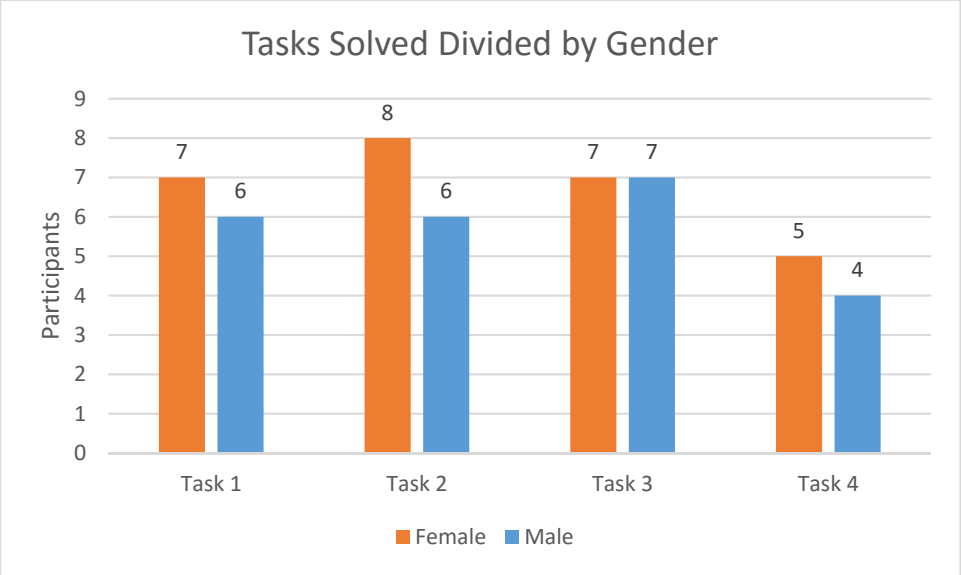


Figure 41: Number of Participants That Solved Task Divided by Gender

Figure 42 shows the average-, minimum-, and maximum time of completion for each task, as well as the difference between minimum time spent, and maximum time spent. If a

participant did not manage to solve a task, that participant has been excluded from the calculations.

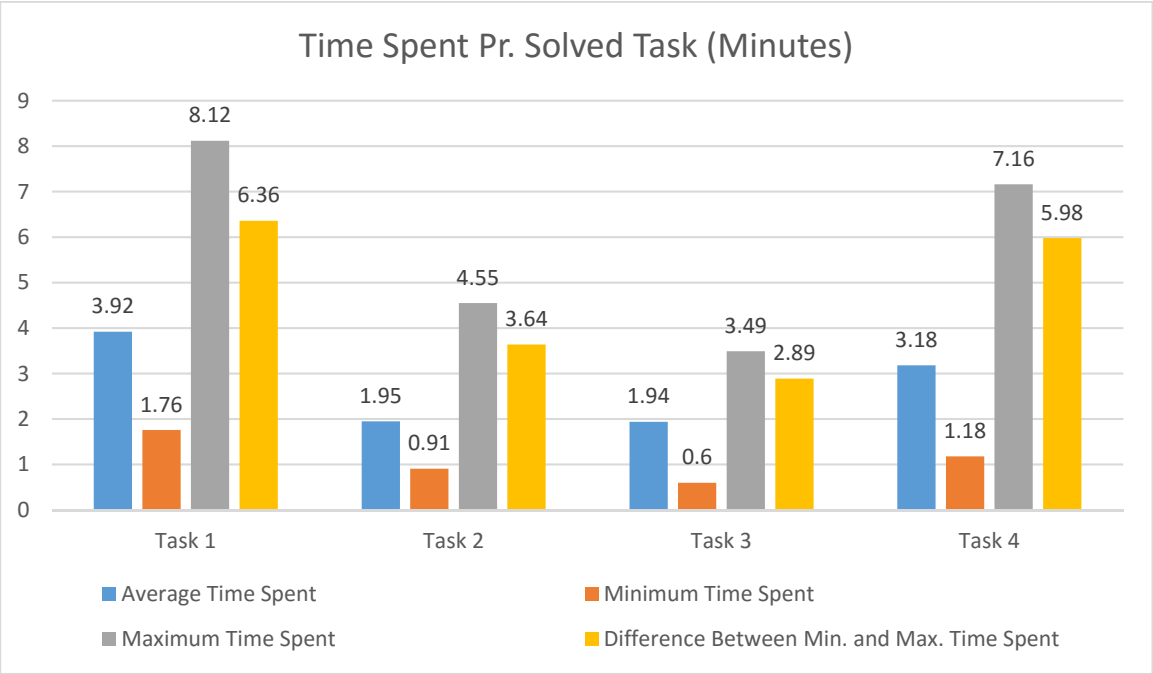


Figure 42: Time Spent Pr. Solved Task (Minutes)

The average time spent by participants for solving Task 1 is 3.92 minutes. Task 2 required on average 1.95 minutes to solve, almost the same as Task 3 with 1.94 minutes. Task 4 required 3.18 minutes on average to solve. The minimum time required by a participant to solve Task 1 is 1.76 minutes. The minimum time required for both Task 2 and 3 was less than a minute (0.91 and 0.6 minutes respectively). For Task 4 the minimum time required was 1.18 minutes. The maximum time required to solve Task 1 was 8.12, for Task 2 it was 4.55 minutes, for Task 3 it was 3.49 minutes, and for Task 4 it was 7.16 minutes.

In Task 1 and 4 the difference between minimum- and maximum time spent is higher (6.36 min. and 5.98 min. respectively), than for Task 2 and 3 (3.64 min. and 2.89 min. respectively). It took participants longer to solve Task 1 and 4 than 2 and 3 (Figure 42). The time spent for each task is listed in Table 4. The maximum time required by participants to solve Task 1 and 4 are also somewhat higher (around 6 min.) than for Task 2 and 3 (less than 5 min.).

On average the time spent on solving Task 1 and 4 was longer than the time used to solve Task 2 and 3. Similar differences are seen regarding the minimal and maximal time use and the difference between them (Table 4).

Table 4: Time Spent Pr. Solved Task (Minutes)

	Task 1	**Task 2**	**Task 3**	**Task 4**
Average Time Spent	3.92	1.95	1.94	3.18
Minimum Time Spent	1.76	0.91	0.60	1.18
Maximum Time Spent	8.12	4.55	3.49	7.16

Difference Between Min. and Max. Time Spent	6.36	3.64	2.89	5.98

Figure 43 shows the time spent pr. task by each participant. Participant 2, 4, 5 and 15 spent more than the average amount of time on solving Task 1. Participant 2, 4, 6, 7, 8, and 12 spent more time than average on solving Task 2. Participant 4, 5, 7, 9, 10 and 13 spent more time than average solving Task 3, and participant 1, 2, 8, and 10 spent more than average time on solving Task 4. Thus, four participants spent longer than average on solving Task 1 and 4, while six participants spent longer than average solving Task 2 and 3.

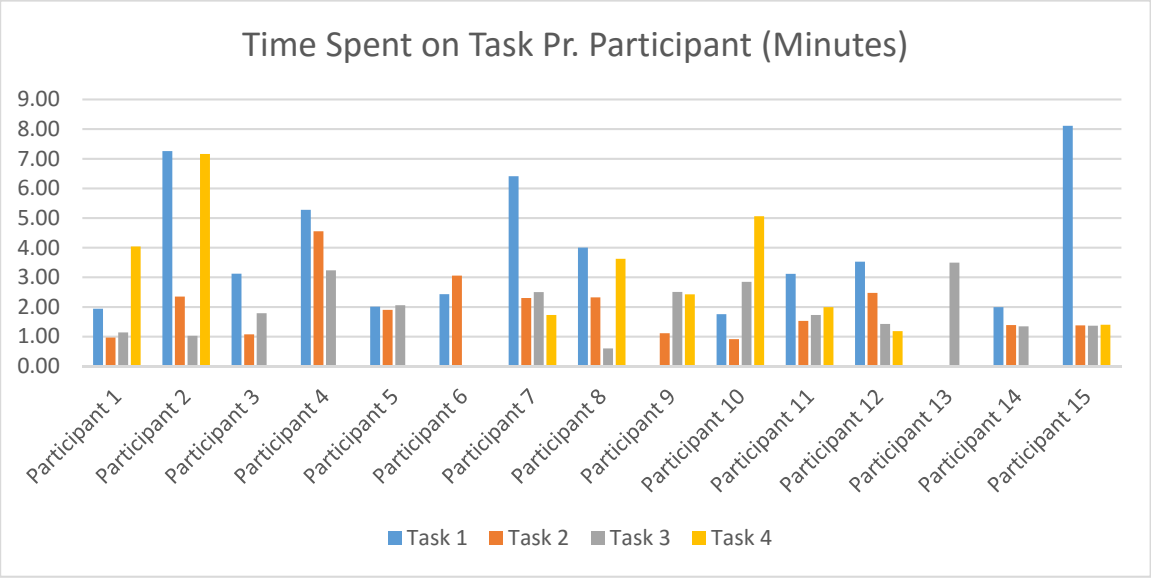


Figure 43: Time Spent on Tasks (in Seconds), Pr. Participant

Table 5 shows the time taken by each participant for solving each task. Completion time written in **Grey** indicates that the participant did not solve the task. Completion time written in **Green** indicates that the task was solved in the minimum amount of time, while completion time written in **Red** indicates that the participant used the maximum amount of all completion times.

Table 5: Time pr. task pr. participant in minutes.

	Task 1	**Task 2**	**Task 3**	**Task 4**
Participant 1	1.94	0.97	1.14	4.04
Participant 2	7.26	2.35	1.03	7.16
Participant 3	3.13	1.07	1.79	5.62
Participant 4	5.28	4.55	3.24	5.59
Participant 5	2.01	1.90	2.06	4.23
Participant 6	2.43	3.06	2.06	3.62
Participant 7	6.41	2.30	2.50	1.73
Participant 8	4.01	2.33	0.60	3.63
Participant 9	4.06	1.12	2.51	2.43
Participant 10	1.76	0.91	2.85	5.06
Participant 11	3.12	1.53	1.73	2.00

Participant 12	3.53	2.47	1.43	1.18
Participant 13	14.54	2.42	3.49	N/A
Participant 14	1.99	1.39	1.35	7.35
Participant 15	8.12	1.38	1.37	1.40

Figure 44-Figure 47 show completion times for male and female participants for each task. The participants that did not solve a specific task will be marked with "unsolved" in the figure. Unsolved tasks will not be included in the comparisons.

Female participants solved Task 1 in between 1.94-8.12 minutes, while male participants solved the task in between 1.76-7.26 minutes. With a gap of 6.18 minutes between the fastest and slowest time for the female participants the completion times vary slightly more, than that of the male participants, which vary by 5.5 minutes. In general, the completion times between male and female participants for Task 1 are relatively similar (Figure 44).

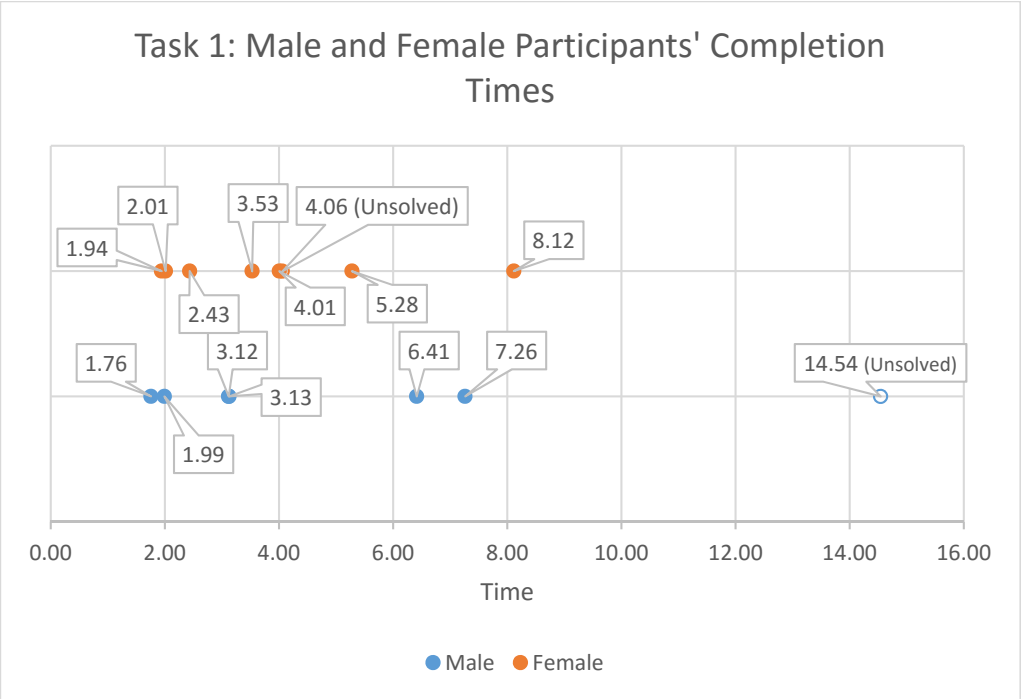


Figure 44: Completion Time for Task 1 by Gender

In Task 2 the completion times between the female participants were between 0.97-4.55 minutes, which means the gap was between 3.58 minutes (Figure 45). For the male participants the completion times were between 0.91-2.35 minutes, a gap of 1.44 minutes. This the gap between the fastest and the slowest of the female participants was larger than that between the fastest and the slowest of the male participants.

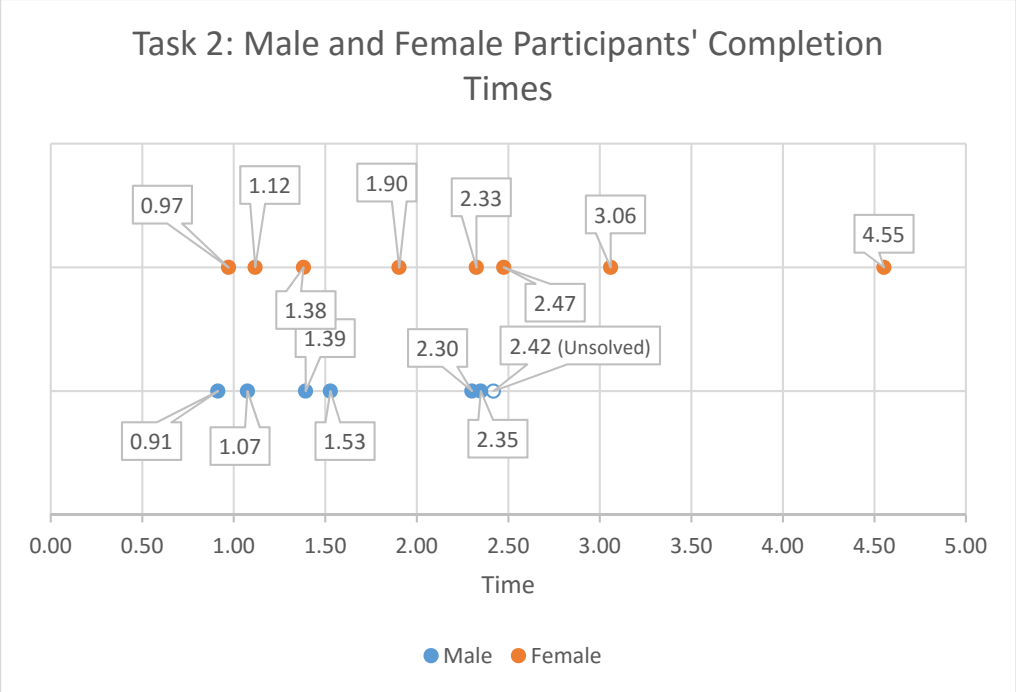


Figure 45: Completion Time for Task 2 by Gender

The completion times for female participants in Task 3 were between 0.60-3.24 minutes, which is a difference of 2.64 (Figure 46). For the male participants the completion times were 1.03-3.49, resulting in a difference of 2.46 minutes. The completion times for male and female participants were somewhat similar for Task 3.

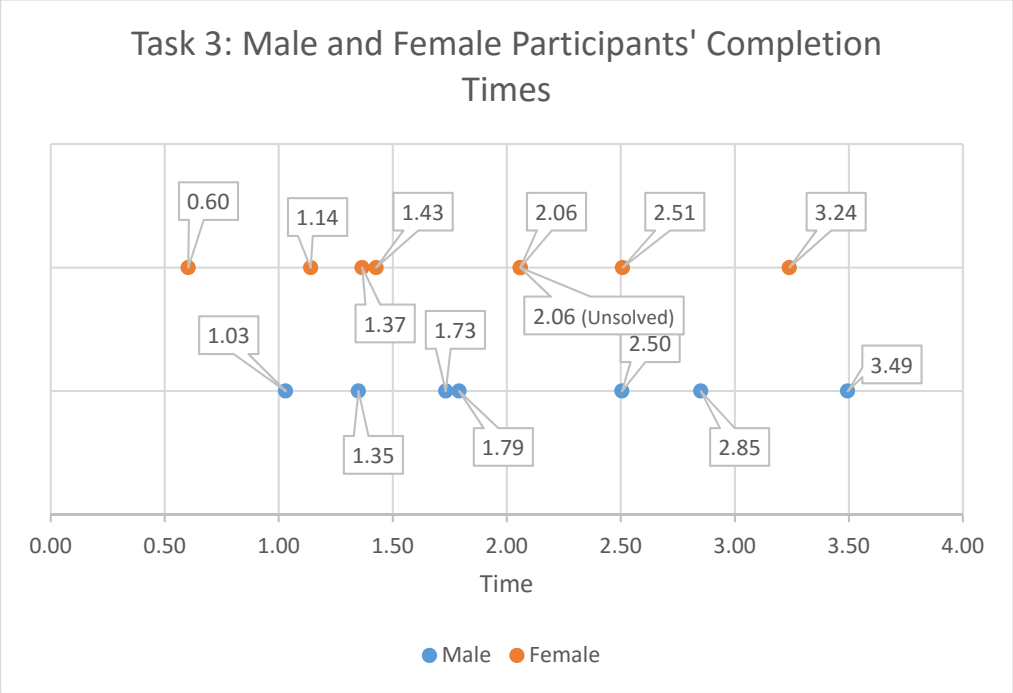


Figure 46: Completion Time for Task 3 by Gender

In Task 4 the completion times for female participants were from 1.18-4.04 minutes, resulting in a gap of 2.86 minutes between the fastest and slowest participant (Figure 47).

For the male participants the fastest time was 1.73 minutes, while the slowest was 7.16 minutes. This means that the gap between fastest and slowest times was 5.43 minutes. In Task 4 the gap between the female participants was thus smaller than that of the male participants.

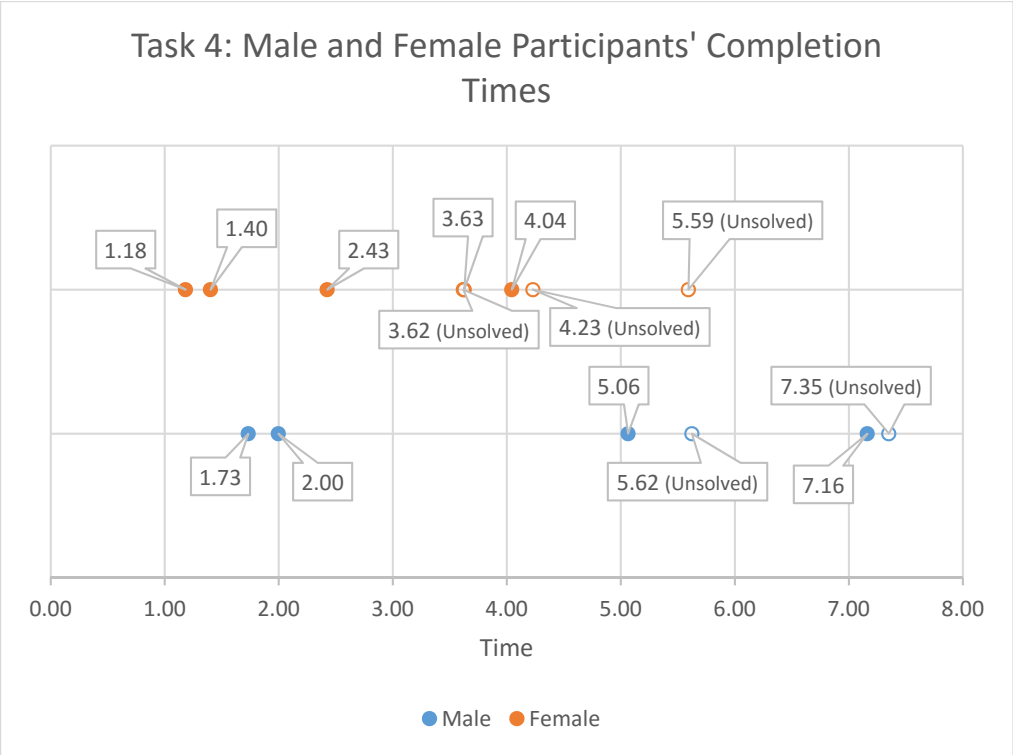


Figure 47: Completion Time for Task 4 by Gender

For Task 1-3 female participants' completion times varied more than the male participants'. For Task 4 the difference between fastest and slowest male participant is greater than that between the female participants. However, there does not seem to be reason to believe that gender has significant influence on completion times.

4.2.1 Participant Rating of Task Difficulty

Figure 49 shows how the participants rated the difficulty of Task 1-4. The rating scale ranges from 1 (easy) to 5 (difficult). Again, it should be noted that one participant ended Task 4 before reading the task description, and has therefore not been asked to rate the difficulty of that task. In Figure 48 a recap of the task descriptions can be found.

Task 1: Participants were asked to imagine that they were about to order a t-shirt online, from a company outside the EU. They were then asked to calculate the customs duty for the item.

Task 2: Participants were asked to find a definition of what a transfrontier worker is.

Task 3: Participants were asked to calculate how much SU after taxes.

Task 4: Participants were asked to find the next date for seminars regarding tax and VAT taking place in either Aarhus or online.

Figure 48: Task 1-4

Figure 49 gives an overview of participants' rating of task difficulty of Task 1-4. As the figure shows Task 1 received mainly positive ratings (i.e. the task was evaluated as being easy), with some neutral ratings as well. Task 2 received mainly positive ratings as well, and the same is true for Task 3, that received no ratings of 5. However, Task 4 received mostly negative ratings stating that the task was difficult to solve by ten out of fourteen participants, with no neutral ratings.

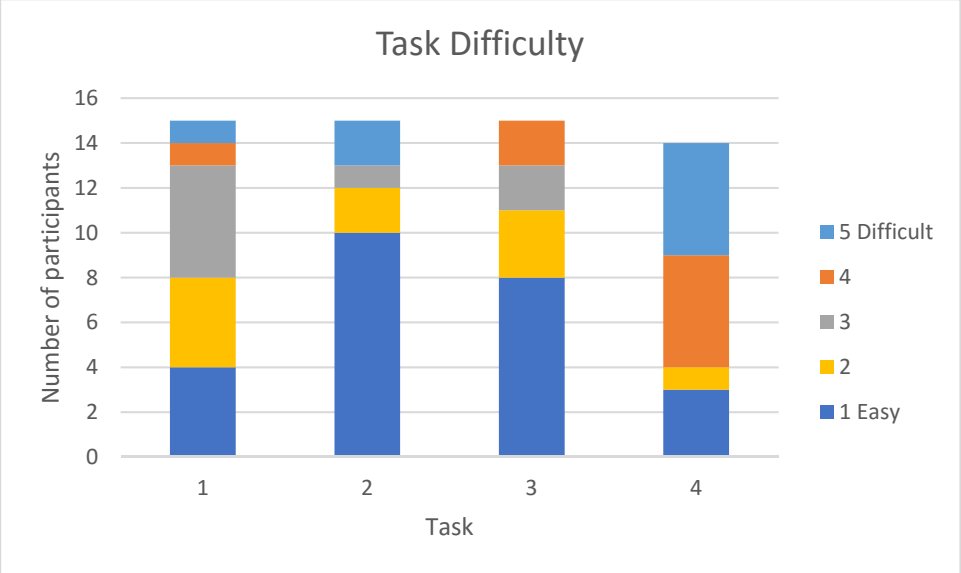


Figure 49: Participants' Rating of Ease/Difficulty of Solving tasks

Figure 50 shows individual participants' ratings of Task 1. Eight participants rated Task 1 as easy or relatively easy to solve. Five participants gave a neutral rating, while two

participants found the task difficult to solve, with the participants giving Task 1 a rating of 4 and 5 respectively. Two participants were unable to complete this task. The participant that gave a rating of 5 was unable to complete the task, whereas the rating of 4 was given by a participant who completed the task faster than the average time. The other participant that did not complete the task gave a neutral rating of 3. This was the participant that was dyslexic.

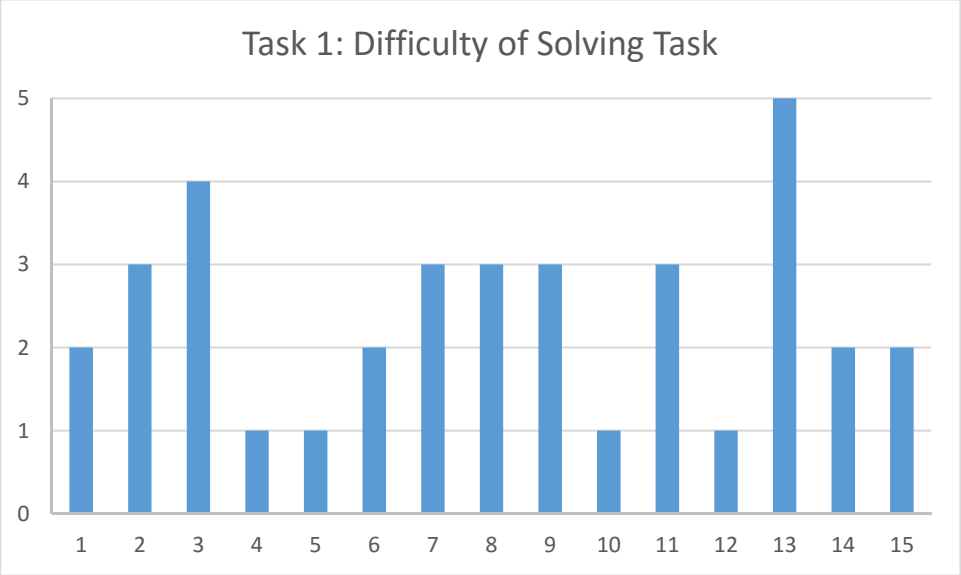


Figure 50: Difficulty Rating of Task 1, Pr. Participant

Figure 51 shows the difficulty ratings of Task 2. Out of the 15 participants, 10 rated Task 2 as easy to solve, while two participants gave a rating of 2. One participant gave a neutral rating, and two participants gave a rating of 5. Of the two 5 ratings one was given by a participant that was unable to complete the task, while the other was given by the participant that was dyslexic.

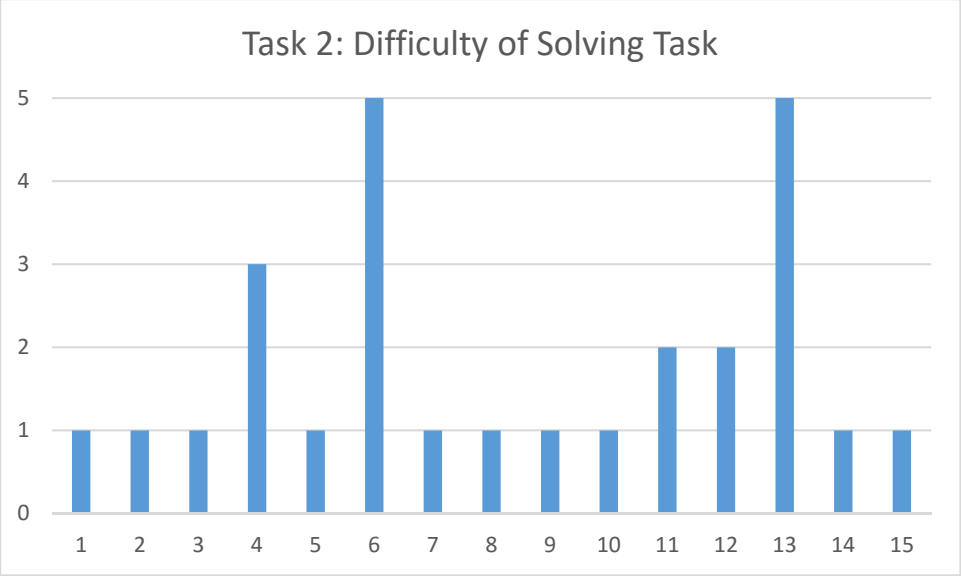


Figure 51: Difficulty Rating of Task 2, Pr. Participant

In Task 3 (Figure 52) eight participants gave a rating of 1, while three gave a rating of 2. Two gave a neutral rating. Two participants rated the difficulty at 4, and none gave a rating of 5. One of the 4 ratings was given by the dyslexic participant, that did not solve Task 3.

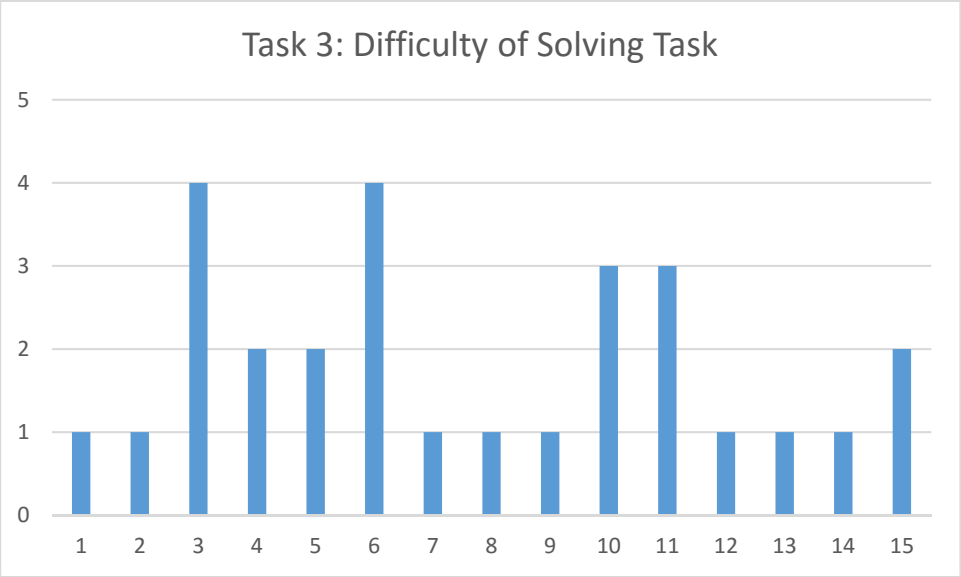


Figure 52: Difficulty Rating of Task 3, Pr. Participant

As seen in Figure 53, Task 4 was rated as a 1 on the scale by three participants, while one gave a rating of 2. None gave task 4 a neutral rating. However, five participants rated task 4 as a 4, and five rated it as 5 on the scale. Of the participants that gave a positive rating (1-2) all succeeded at completing the task. Five participants gave a rating of 4. Four of these participants solved the task.

Of the five participants that gave a rating of 5 four did not manage to complete the task. One of these participants was not familiar with the search term ("Seminar") given in the task description. An attempt at a verbal definition was provided by the researcher. One of the participants that gave a rating of 5 succeeded in completing the task. One participant ended the task before reading the task description, and was not counted in the total of Task 4.

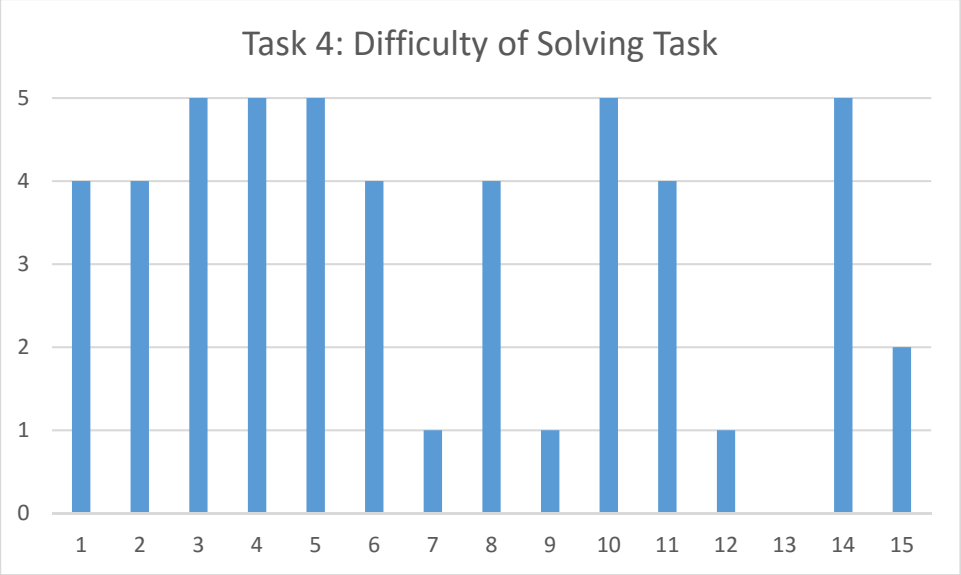


Figure 53: Difficulty Rating of Task 4, Pr. Participant

4.2.2 Participant Rating of Level of Frustration Pr. Task

The participants were asked to rate how simple (1) or frustrating (5) the different tasks were to solve. An overview of the ratings of the individual tasks can be seen in Figure 54. As can be seen when comparing the graphs in Figure 49 and Figure 54 there is a slight difference between how participants rate difficulty and frustration level of the tasks. However, it is clear that task ratings are somewhat consistent, showing that Task 1 and 3 were the least frustrating tasks, while most participants experienced some frustration when solving Task 4. If a task received mainly positive ratings in regards to the difficulty it also received mainly positive ratings regards to frustration level.

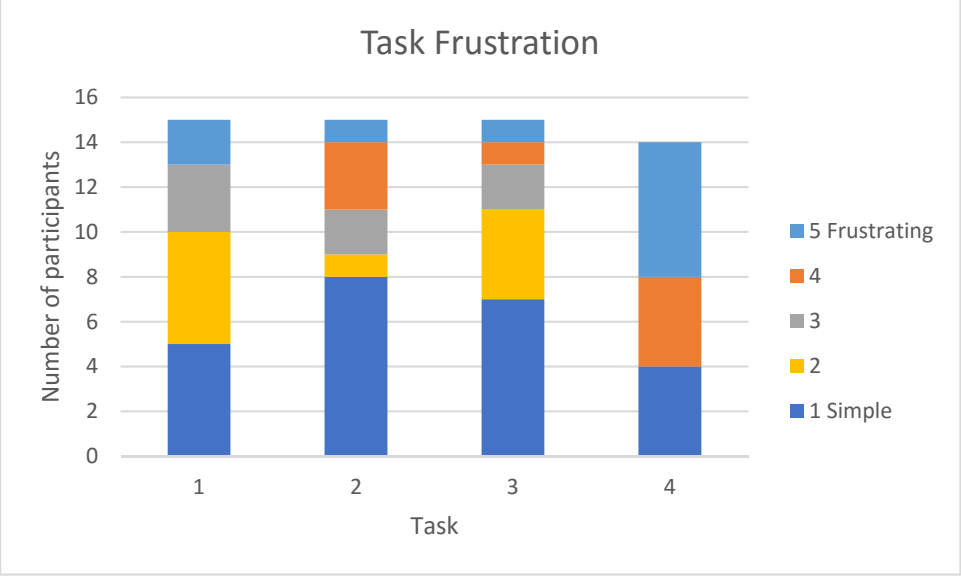


Figure 54: Participants' Rating of how Simple/Frustrating it was to Solve tasks

Figure 55 shows individual participants' ratings of Task 1. Five participants gave Task 1 a rating of 1. A rating of 2 was also given by five participants. The neutral rating of 3 was given by 3 participants. The two participants that gave the task a negative rating of 5 were also

the two participants that did not complete the task. It was noted that these participants were confused by Morae's interface. Ratings of Task 1 were mainly positive.

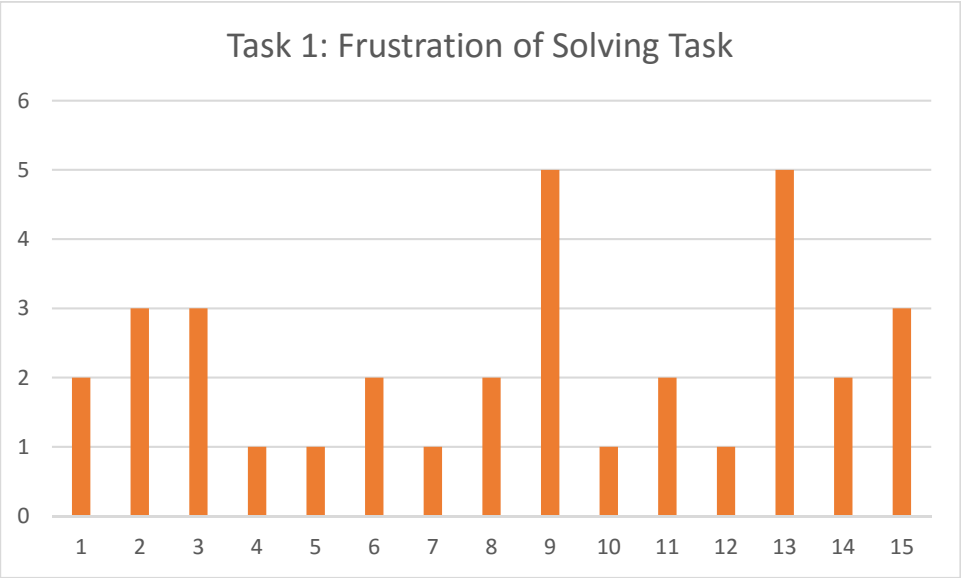


Figure 55: Frustration Rating of Task 1, Pr. Participant

Individual participants' ratings of Task 2 can be seen in Figure 56. Eight participants gave Task 2 a positive rating of 1, while one gave a rating of 2. Two participants gave a neutral rating of 3. Three participants gave a negative rating of 4. A negative rating of 5 was given by the one participant who did not complete the task.

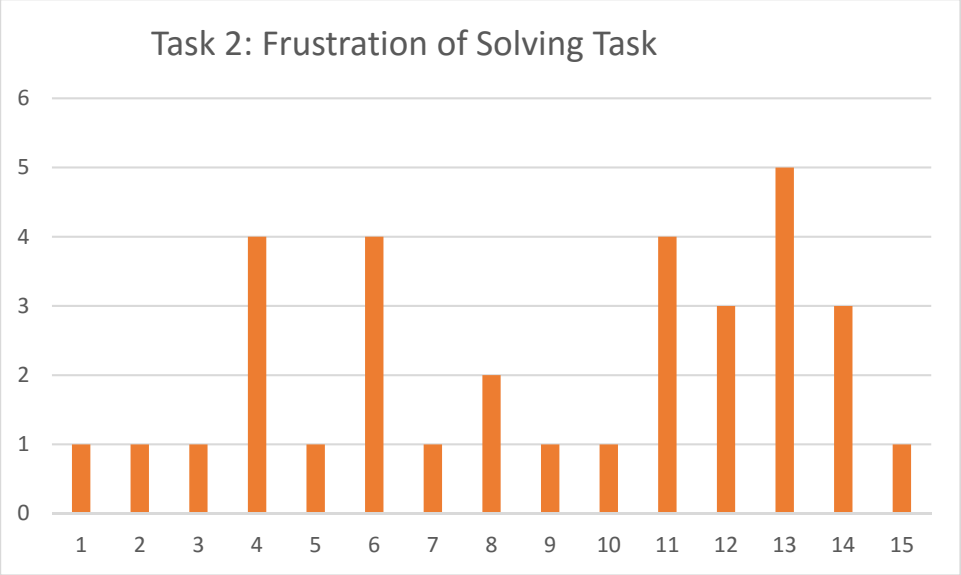


Figure 56: Frustration Rating of Task 2, Pr. Participant

As seen in Figure 57 seven participants rated Task 3 at 1, four gave the task a rating of 2, and two chose to give a neutral rating of 3. One participant gave a rating of 4. The rating of 5 was given by the participant who is dyslexic.

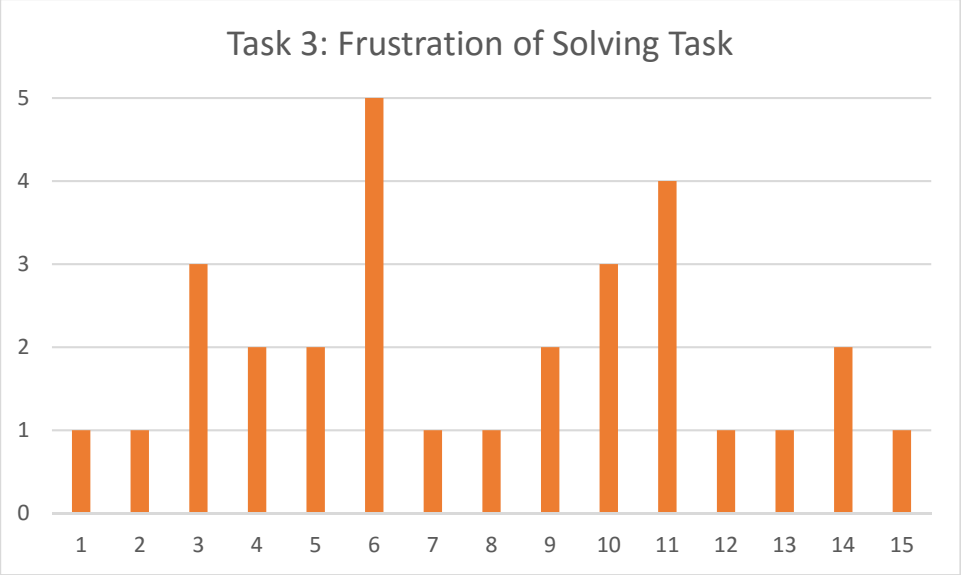


Figure 57: Frustration Rating of Task 3, Pr. Participant

Figure 58 shows individual ratings for Task 4. Four participants rated Task 4 as simple, with a rating of 1. However, four participants gave a rating of 4, one of these participants was not able to solve the task. This was the participant that is dyslexic. Of the six participants that rated Task 4 as a 5 on the scale, four were unable to solve the task. As stated in section 4.2.1 one participant ended the task before reading the task description, and has therefore not been asked to rate Task 4. One participant was unfamiliar with the keyword ("Seminar") provided in the task description.

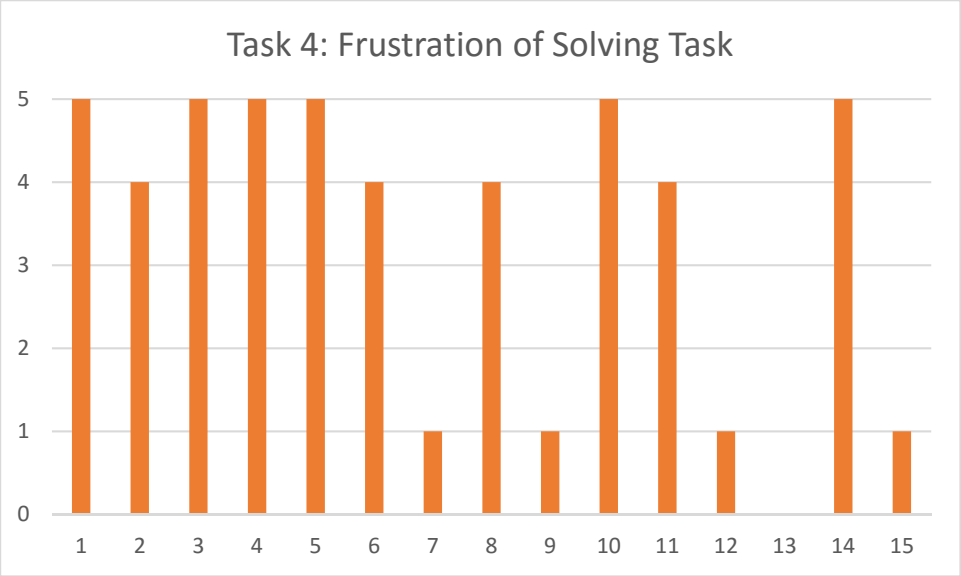


Figure 58: Frustration Rating of Task 4, Pr. Participant

4.2.3 Relation Between Ratings

The following section summarises the connection between ratings of the ease of solving a task, and the level of frustration in solving the tasks, for the individual participants. In general, the ratings are somewhat consistent (i.e. if a positive rating is given on one of the scales, the rating in the other scale is usually positive as well, and vice versa).

In Figure 59 the relation between the ratings given on both 5-point scales of task 1 are shown. Nine participants are shown to give the same rating on both scales, while 4 have given ratings that differ by 1 on the scales. Only two participants gave ratings that differ by 2.

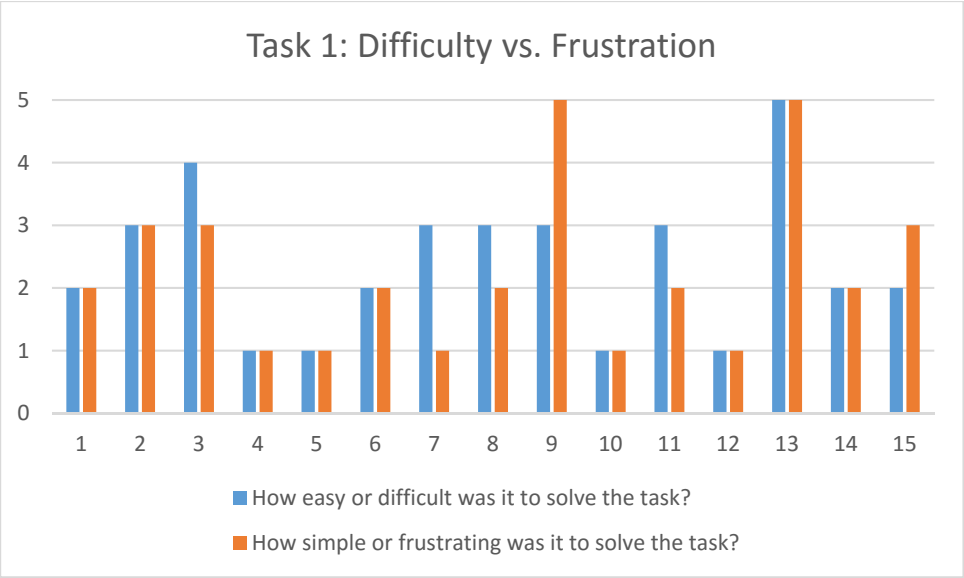


Figure 59: Ratings of Task 1

Figure 60 shows the relation between the ratings for Task 1. There is a clear tendency among participants to give similar ratings of difficulty and frustration, so that we may regard frustration and task difficulty to be related to each other.

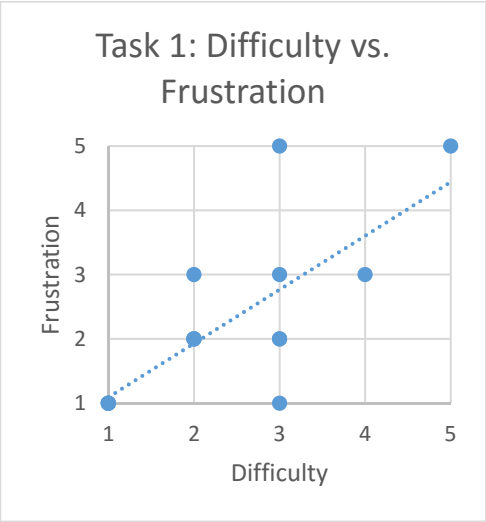


Figure 60: Relation Between Ratings of Task 1

Ratings by individual participants of Task 2 are shown in Figure 61. Four participants gave ratings that differed by 1 on the two scales, while two gave ratings that differed by 2. Participant 11 gave a positive rating on the difficulty scale, but a negative rating on the frustration scale. This indicates that while the participant perceived Task 2 as relatively easy to solve it was still perceived as frustrating.

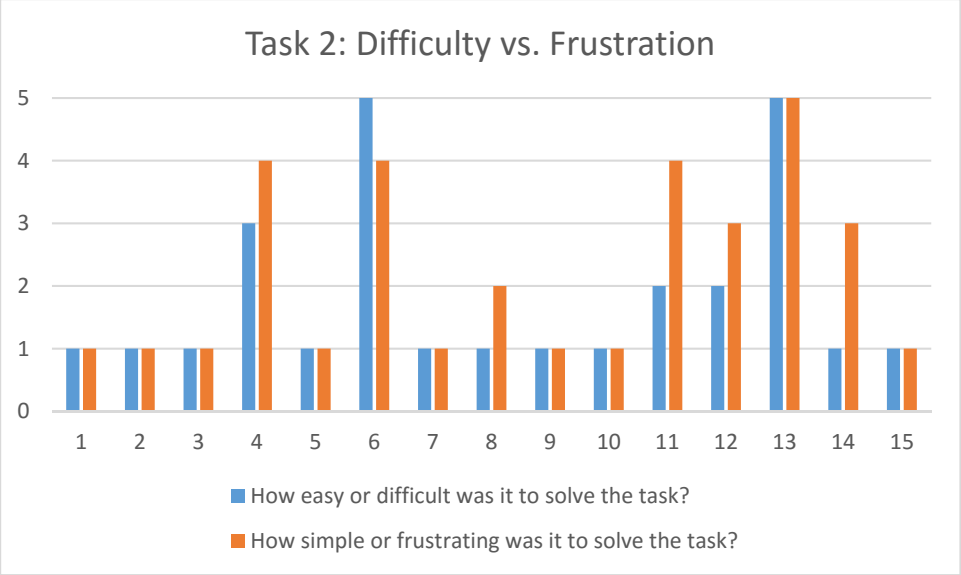


Figure 61: Ratings of Task 2

Figure 62 shows the relation between the ratings for Task 2. Again there seems to be a clear relation between ratings of difficulty and frustration.

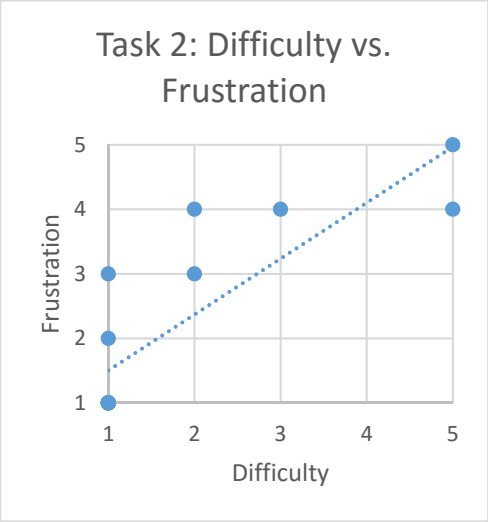


Figure 62: Relation Between Ratings of Task 2

Figure 63 shows individual participants' ratings for Task 3. Six participants gave ratings that differed by 1 on the scales.

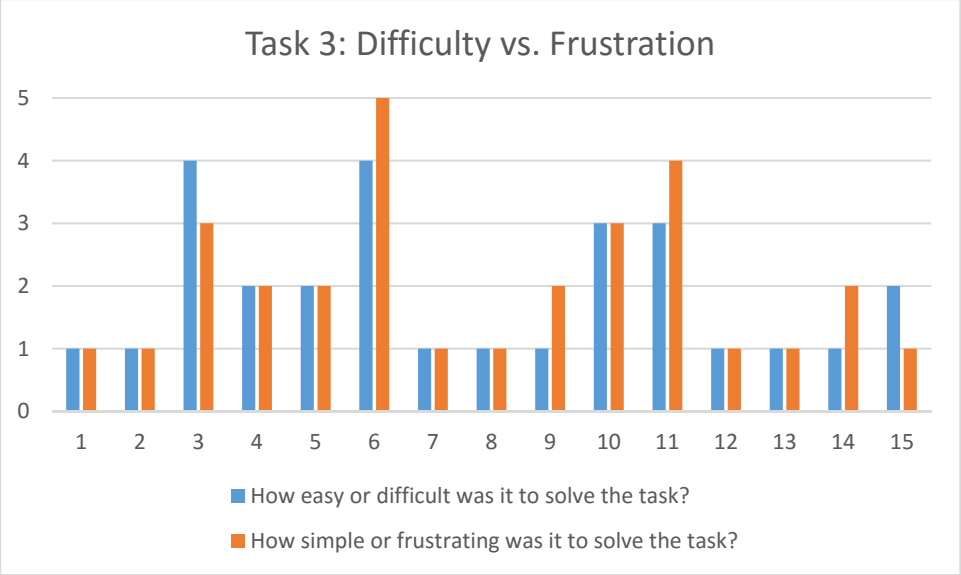


Figure 63: Ratings of Task 3

As indicated in Figure 64 there seems to be a clear relation between the ratings of Task 3. The more difficult participants found the task to be, the more frustrating they found it as well.

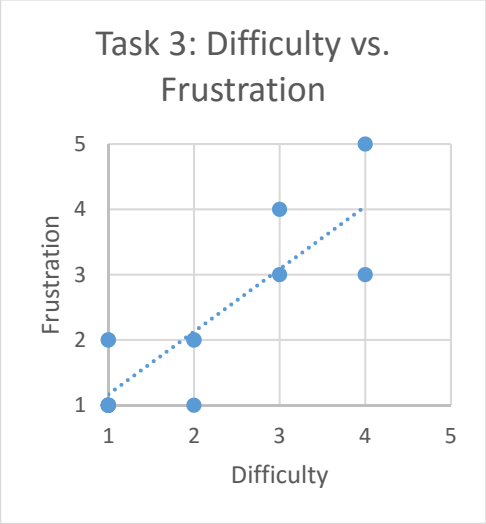


Figure 64: Relation Between Ratings of Task 3

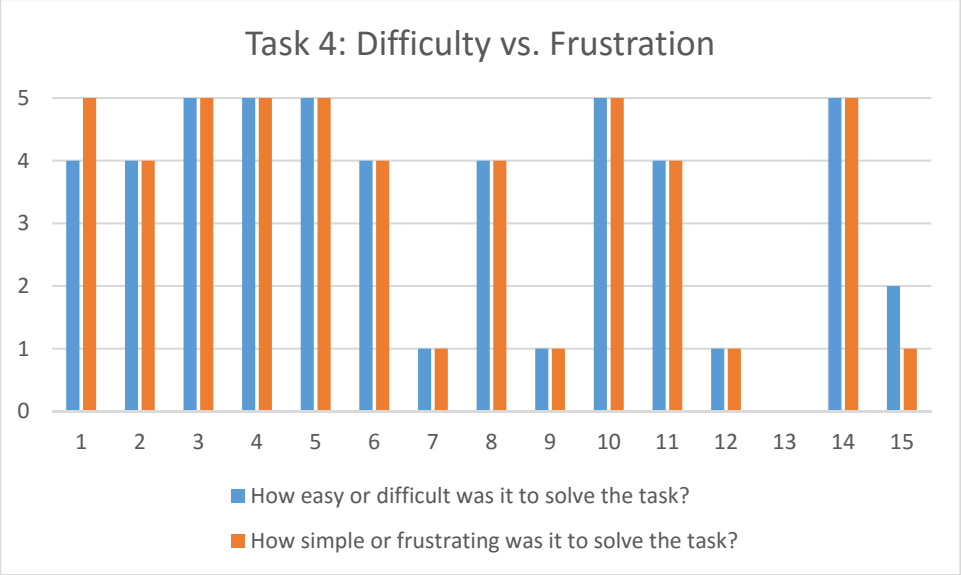


Figure 65: Ratings of Task 4

Participants were generally frustrated by Task 4, and found it difficult to solve. However, not only participants that did not solve the task gave negative ratings. As seen in Figure 66 no neutral ratings were given. Either participants found the task easy to solve and found the task simple, or they found the task difficult and frustrating.

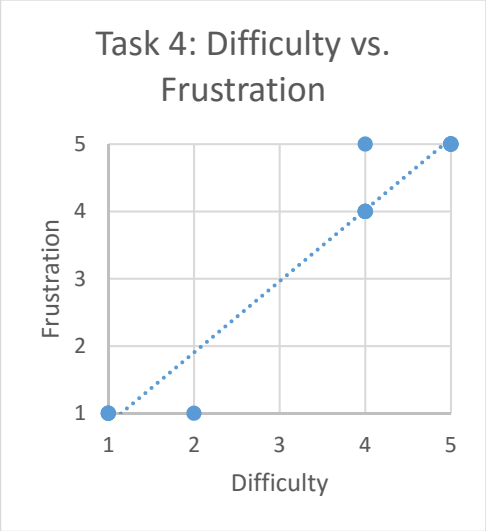


Figure 66: Relation Between Ratings of Task

4.2.4 Relation Between Frustration Rating and Time Spent on Task

Figure 67 shows the relation between the time participants spent solving Task 1 and the level of frustration experienced. There seemed to some relation between how much time participants spent solving a task, and how frustrating they found the task.

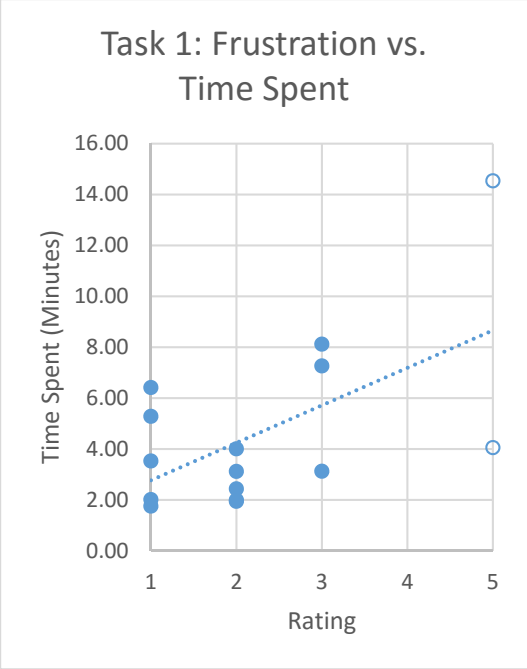


Figure 67: Connection Between Level of Frustration and Time Spent Solving Task 1

For Task 2 there seemed to be some relation between time spent and experienced frustration as well (Figure 68).

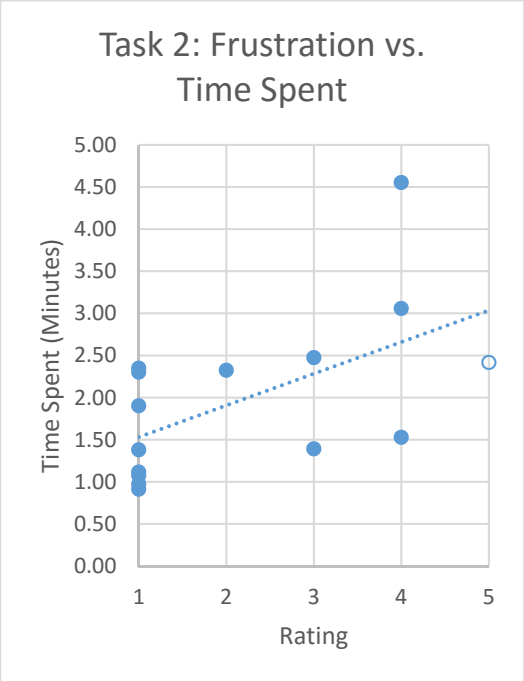


Figure 68: Connection Between Level of Frustration and Time Spent Solving Task 2

Figure 69 shows time spent and frustration level for Task 3. In this task there does not seem to be an obvious connection between time taken to solve the task and level of frustration.

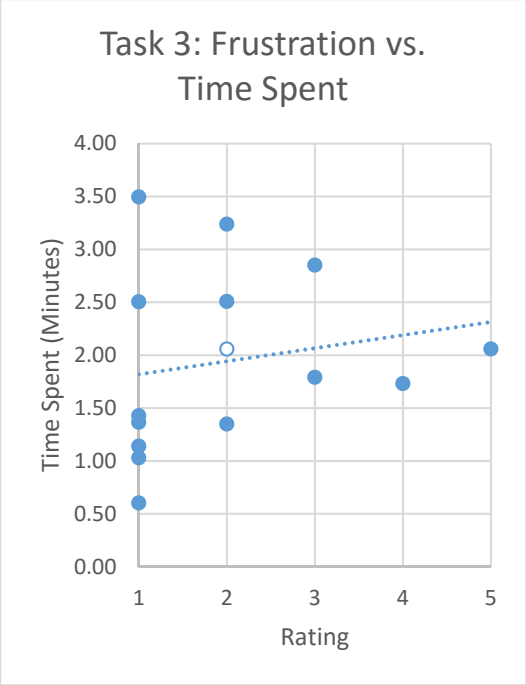


Figure 69: Connection Between Level of Frustration and Time Spent Solving Task 3

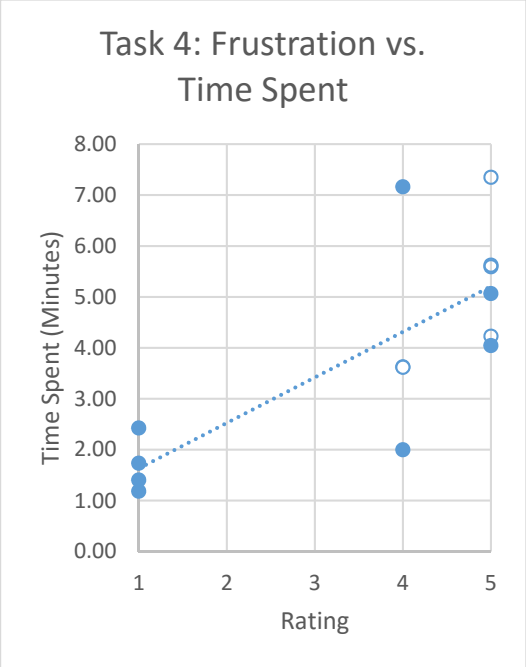


Figure 70: Connection Between Level of Frustration and Time Spent Solving Task 4

For Task 4 there were no neutral ratings. However, there does seem to be a connection between time spent and level of frustration (Figure 70).

The analysis of task difficulty vs. use of time and evaluation of the level of frustration reveal these to be related to each other. We may conclude that the more difficult a task is experienced to be, the longer time it takes to solve it and the more frustration it generates in general.

4.2.5 Search Behaviour for Task 4

Ten participants used the search function of the website in connection with Task 4. Of these participants six solved the task. However, it should be noted that most of the participants that solved the task did so, not as a result of using the search function, but despite of it.

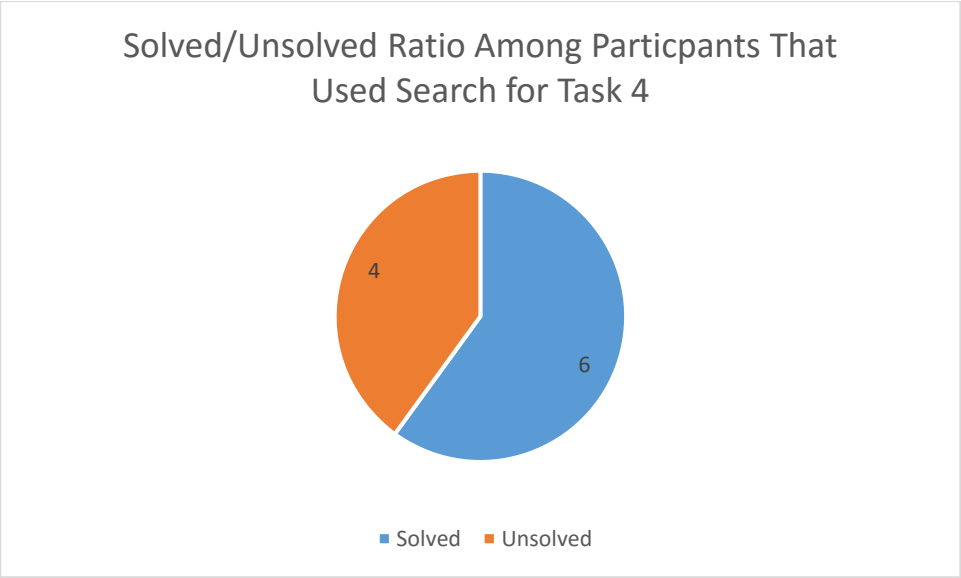


Figure 71: Ratio of Solved/Unsolved Tasks Among Participants the Used Search in Task 4

Figure 72 gives an overview of the search queries made by participants in Task 4. Almost all of these queries involve the term "seminar", which was part of the task description. Only the query "skat og moms i århus" did not include "seminar". Interestingly this was the only search query that led directly to the user finding the desired result.

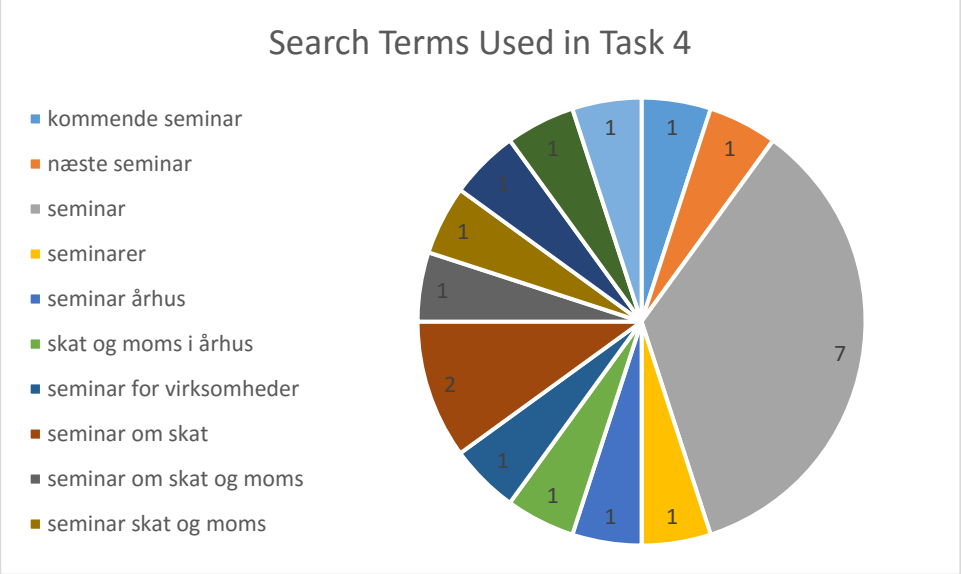


Figure 72: Search Queries Made by Participants in Task 4

The search queries may point to a tendency among the participants to use only a limited set of search terms. Although many participants rephrased their queries one or two times, the rephrased queries almost always included "seminar", with the result that the search was

unsuccessful. Interestingly, Task 2 was designed to require participants to use the search function. Only one participant did not solve Task 2, presumably because the term used in the task description ("grænsegænger") provided the participants with relevant results.

4.2.6 Prior Knowledge of Problem

For each task the participants were asked whether they knew about the problem presented in the task description before solving it.

Figure 73 shows how many participants answered "Yes" to the question of whether they knew about the problem presented in Task 1 before trying to solve the task. Nine participants answered that they did not know the problem beforehand, while six did. Participants 3, 7-10, and 12-15 did not know problem beforehand.

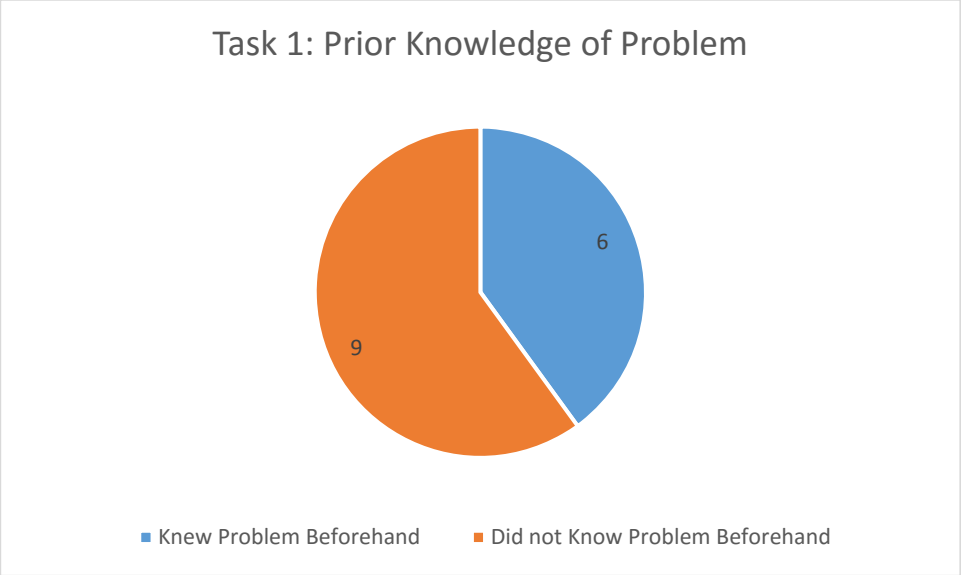


Figure 73: Prior Knowledge of Problem for Task 1

In Task 2 all participants answered "no" to the question of whether they knew the problem beforehand (Figure 74). However, two participants stated that they knew of the concept of working in one country, while paying taxes in another, but did not knowing there was a term for that particular situation.

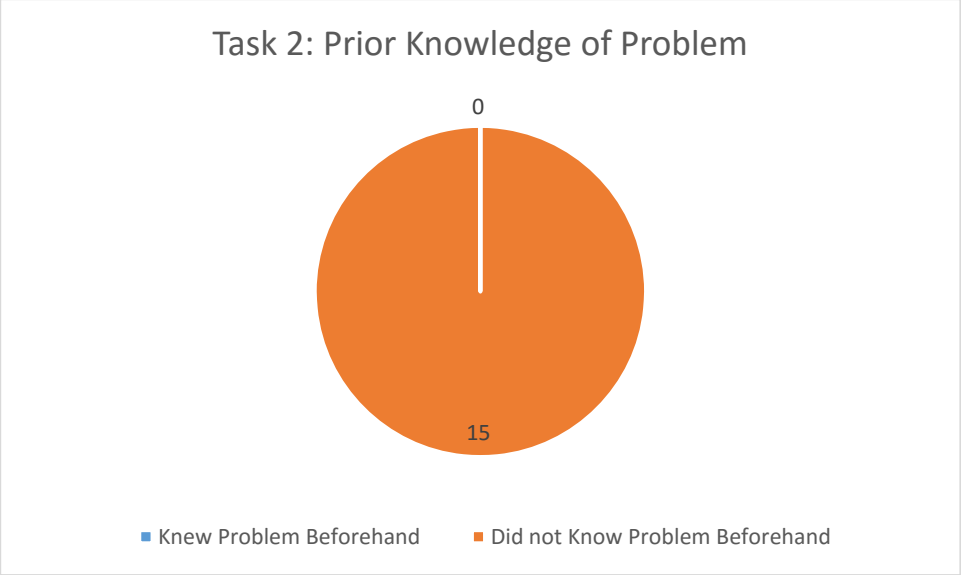


Figure 74: Prior Knowledge of Problem for Task 2

Ten participants answered "Yes" to knowing the problem for Task 3 beforehand (Figure 75). These were participants 1-3, 6, 8, 11, 12, 14, and 15. Five answered that they did not know.

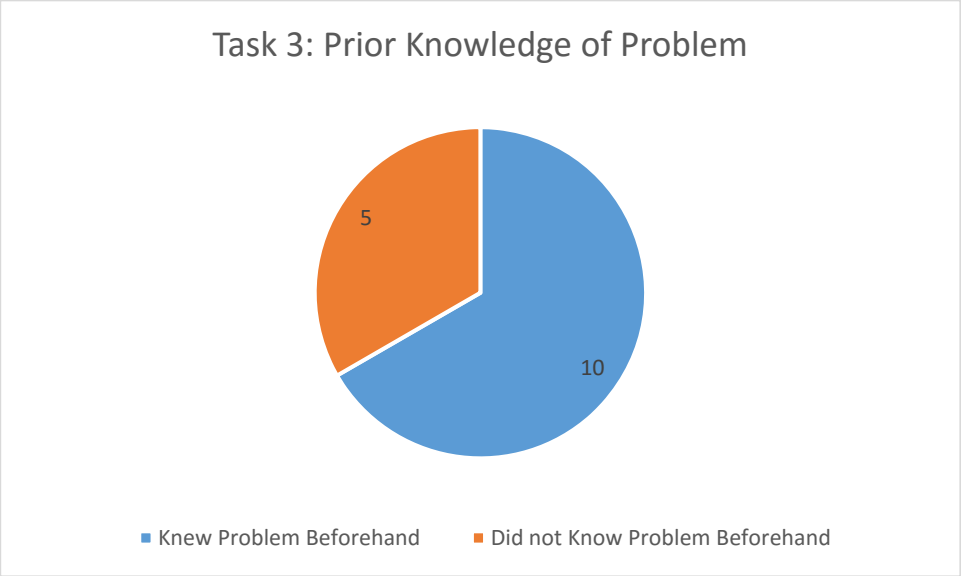


Figure 75: Prior Knowledge of Problem for Task 3

Of the fourteen participants, that were asked to rate Task 4, 13 answered that they did not know about the problem beforehand (Figure 76). Only participant 12 answered "yes" to knowing the problem.

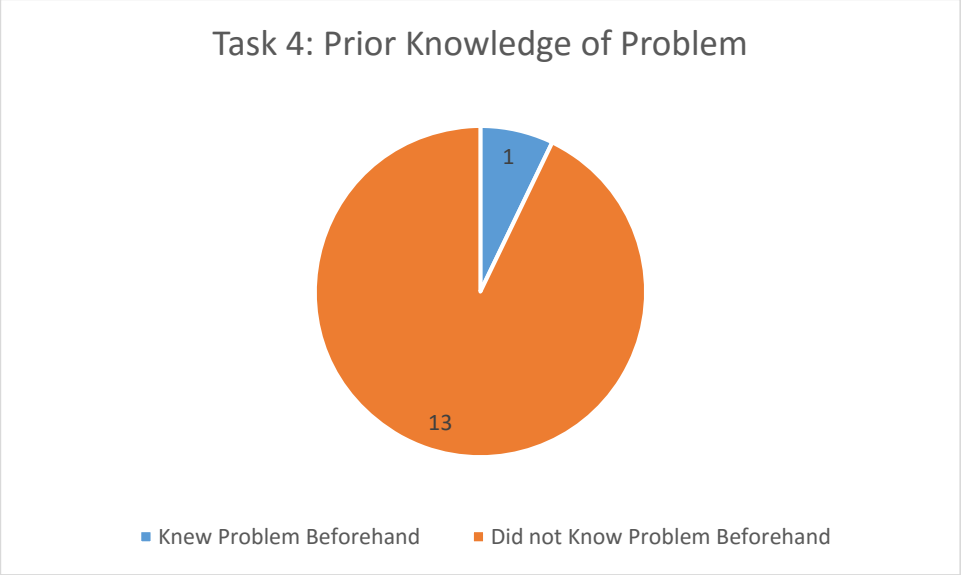


Figure 76: Prior Knowledge of Problem for Task 4

Nine participants stated that they did not know about the problem of Task 1 before trying to solve the task. Task 1 was rated by participants to be easy or of neutral difficulty. No participants knew the problem of Task 2 beforehand, however, this was the task that received the largest number of positive ratings on the task difficulty, meaning that this task was in general perceived to be easy. In Task 3 5 participants stated that they did not know the problem beforehand. This task received generally positive ratings of the easiness of the task. Of the fourteen participants that were asked whether they knew the problem of Task 4 beforehand, 13 answered that they did not. This task was rated mostly as being difficult to solve. Based on these results there does not seem to be a relation between whether participants knew of the problem beforehand, and whether they found the tasks difficult to solve or not.

4.3 User Experience Questionnaire

The results of the User Experience Questionnaire will be presented in three separate sections. As explained in chapter 3, section 3.1.7 of this thesis, “items” are the statements participants were asked to rate in the User Experience Questionnaire (e.g. item 3 measures whether the system is “creative or dull”), while “scales” refer to which aspect the individual item measures (e.g. item 3 belongs to the “Novelty” scale).

First the overall results of the User Experience Questionnaire will be presented (i.e. how the system scores on the different scales, the mean values of item ratings, the variance, and standard deviations for each item). Hereafter, a presentation is made of the confidence intervals (i.e. how precise the mean values of the different items are). Lastly, the level of consistency of the answers for the different scales will be discussed.

4.3.1 Results

The results from the User Experience Questionnaire indicate that the User Experience of SKAT.dk is fairly neutral overall (Figure 77). The score on the Attractiveness scale is 0.511 (neutral), Perspicuity scale is 0.833 (positive), Efficiency scale is 0.750 (neutral),

Dependability scale is 1.233 (positive), Stimulation scale is (neutral), and the score for the Novelty scale is -0.900 (Negative).

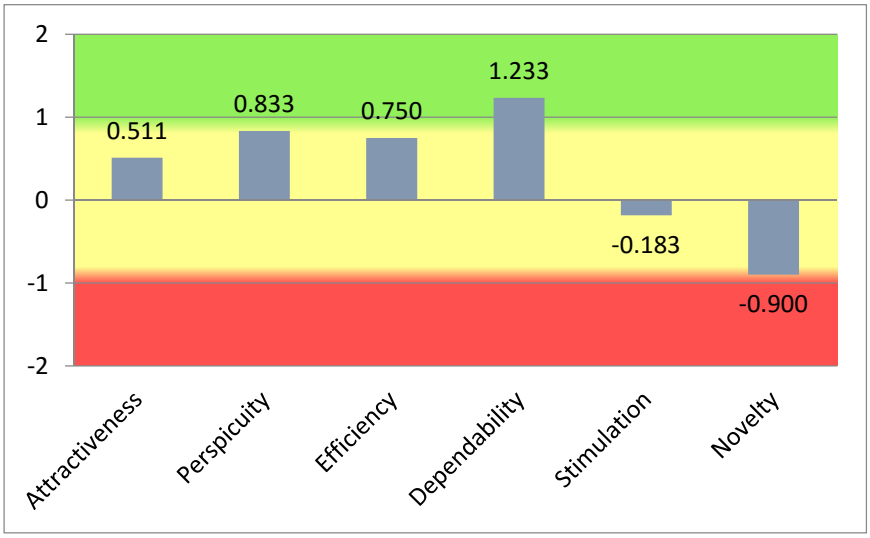


Figure 77: SKAT's Score on UEQ Scales, All Participants

Figure 78 shows the score for SKAT.dk's Attractiveness, as well as Pragmatic (which consists of Perspicuity, Efficiency and Dependability, i.e. goal directed) and Hedonic (which consists of Stimulation and Novelty, i.e. not goal directed) Quality, as measured in the User Experience Questionnaire. The score for the Pragmatic Quality is 0.94. and Hedonic Quality is -0.54. This indicates that SKAT.dk may offer a slightly positive User Experience when it comes to goal directed aspects, whereas the non-goal directed aspects should be improved. Overall, the result of the User Experience Questionnaire suggests that there is reason to consider working on improving the User Experience on the website.

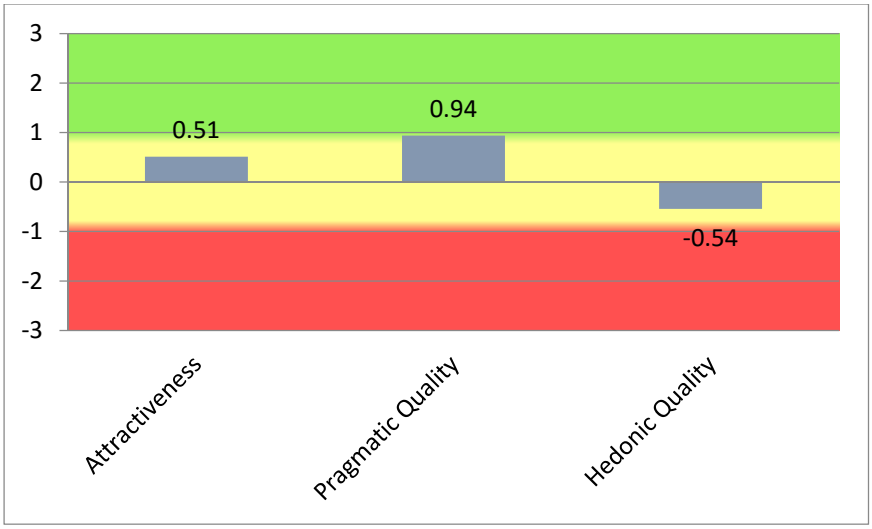


Figure 78: Attractiveness, Pragmatic and Hedonic Quality, All participants

Table 6 shows the mean values, the variance and the standard deviation of each item, as well as which scale the item belongs to. A positive mean value is indicated by green, neutral by yellow, and negative by red. The colours in the Scale column show which scales each item belongs to.

Table 6: Mean, Variance and Standard Deviation Pr. Item.

Item	Mean	Variance	Std. Dev.	Left	Right	Scale
1	0.1	1.8	1.4	annoying	enjoyable	Attractiveness
2	1.2	2.3	1.5	not understandable	understandable	Perspicuity
3	-0.2	2.0	1.4	creative	dull	Novelty
4	0.7	3.1	1.8	easy to learn	difficult to learn	Perspicuity
5	0.8	3.0	1.7	valuable	inferior	Stimulation
6	-1.1	1.7	1.3	boring	exciting	Stimulation
7	-0.3	1.8	1.3	not interesting	interesting	Stimulation
8	0.5	2.7	1.6	unpredictable	predictable	Dependability
9	0.0	3.6	1.9	fast	slow	Efficiency
10	-1.2	1.6	1.3	inventive	conventional	Novelty
11	0.3	1.8	1.3	obstructive	supportive	Dependability
12	1.1	1.6	1.3	good	bad	Attractiveness
13	0.7	4.1	2.0	complicated	easy	Perspicuity
14	0.1	3.8	1.9	unlikable	pleasing	Attractiveness
15	-1.6	2.8	1.7	usual	leading edge	Novelty
16	0.8	1.7	1.3	unpleasant	pleasant	Attractiveness
17	2.5	0.6	0.7	secure	not secure	Dependability
18	-0.1	2.1	1.5	motivating	demotivating	Stimulation
19	1.6	2.5	1.6	meets expectations	does not meet expectations	Dependability
20	0.5	4.1	2.0	inefficient	efficient	Efficiency
21	0.7	3.8	2.0	clear	confusing	Perspicuity
22	1.1	3.6	1.9	impractical	practical	Efficiency
23	1.4	2.0	1.4	organized	cluttered	Efficiency
24	-0.2	3.0	1.7	attractive	unattractive	Attractiveness
25	1.2	1.7	1.3	friendly	unfriendly	Attractiveness
26	-0.6	1.3	1.1	conservative	innovative	Novelty

The range of the scale is -3 (negative) to +3 (positive). Values -0.8 between 0.8 indicate a neutral rating, while values above 0.8 indicate a positive rating. Values below -0.8 indicate a negative rating.

Nine items have a positive rating; one item has a rating above 2 indicating that participants perceived the site as secure. Three items received negative ratings of -1 or below. They were: boring/exciting, inventive/conventional, and usual/leading edge. Of the neutral ratings six are at +, while five are at -.

Figure 79 is a graphic representation of the mean value of each item. Mean value for items on the Novelty, Stimulation and Attractiveness scales shows some inconsistencies (Table 6, Figure 79). On the Stimulation scale, item 5 (valuable/inferior) has a mean value of 0.8, item 6 (boring/exciting) has a value of -1.1, 7 (not interesting/interesting) has a value of -0.3, and item 18 (motivating/demotivating) has a mean value of -0.1.

On the Attractiveness scale SKAT.dk was rated as mainly positive and neutral by the participants, but there is a slight inconsistency with one negative rating in item 24 (attractive/unattractive).

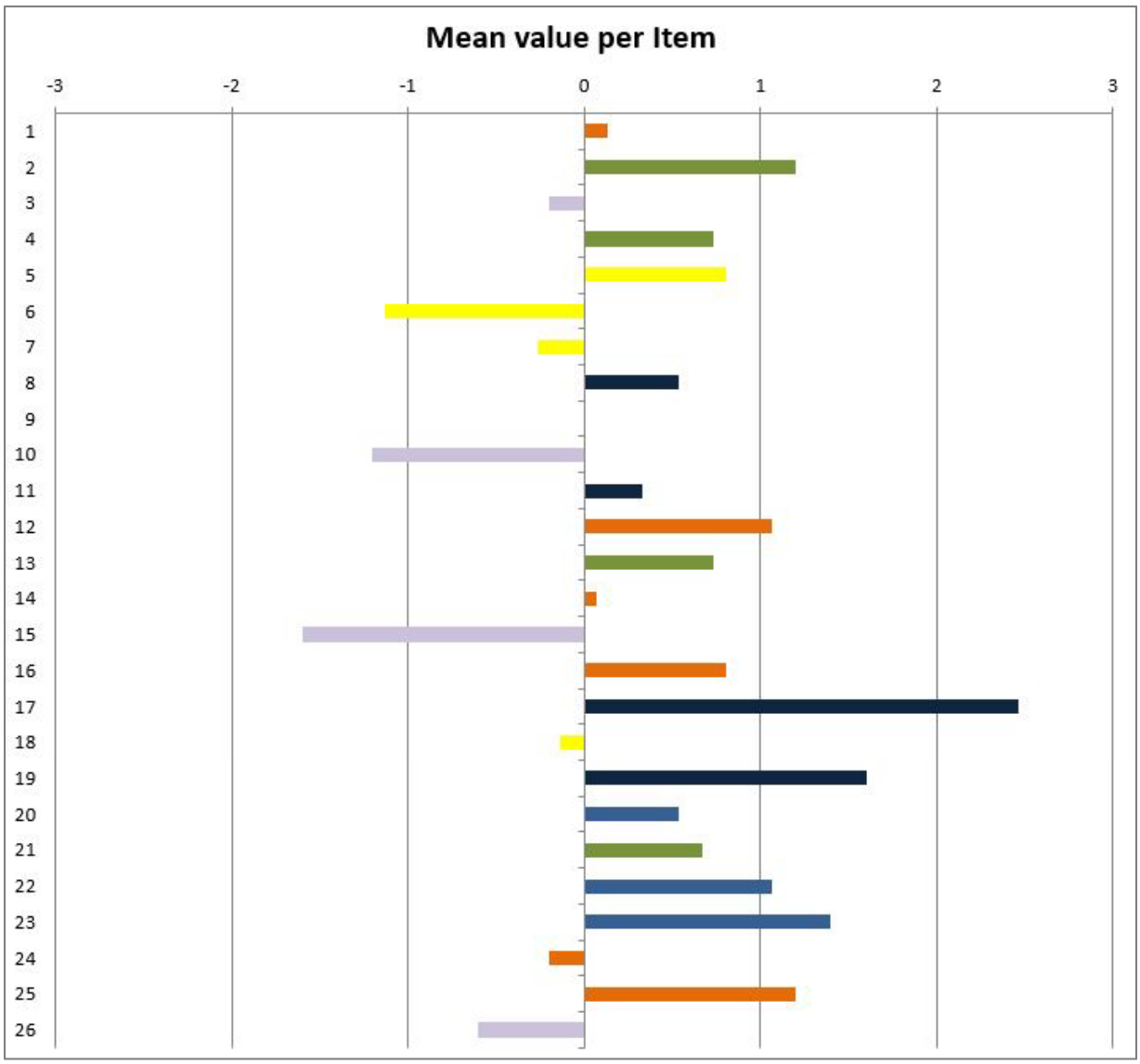


Figure 79: Mean value pr. item

4.3.2 Confidence Intervals

The confidence intervals give an indication of the accuracy of the mean values of the scales (Table 7). The lower the confidence intervals for the individual items are, the more likely it is, that the estimation of the mean values are precise.

Table 7: Confidence Interval

Confidence intervals (p=0.05) per scale						
Scale	**Mean**	**Std. Dev.**	**N**	**Confidence**	**Confidence interval**	
Attractiveness	0.511	1.222	15	0.619	-0.107	1.130
Perspicuity	0.833	1.505	15	0.762	0.072	1.595
Efficiency	0.750	1.225	15	0.620	0.130	1.370

Dependability	1.233	0.961	15	0.486	0.747	1.720
Stimulation	-0.183	0.923	15	0.467	-0.651	0.284
Novelty	-0.900	1.085	15	0.549	-1.449	-0.351

The confidence values for the different scales are relatively high, which indicates that the estimation of the mean values are not very precise.

4.4.3 Inconsistencies

If a participant gives conflicting answers to items belonging to the same scale that participant may be considered inconsistent. Out of the fifteen participants seven gave inconsistent ratings on more than two scales. E.g. participant 2 gave inconsistent answers on the Efficiency scale, by rating the website as both impractical (item 22), and organized (item 23).

Table 8: Scales with Inconsistent Answers

Scales with inconsistent answers							
Parti-cipant	**Attractive-ness**	**Perspicuity**	**Efficiency**	**Depend-ability**	**Stimulation**	**Novelty**	**Critical?**
1			1				1
2			1	1	1		3
3							0
4				1		1	2
5							0
6	1		1	1			3
7	1		1		1		3
8							0
9		1	1	1			3
10	1						1
11							0
12				1			1
13		1	1	1	1		4
14			1		1		2
15							0

Two participants (4, and 14) gave inconsistent ratings on two scales. Four participants (2, 6, 7, and 9) gave inconsistent ratings on three scales. One participant (13) gave inconsistent ratings on four scales.

The scales these participants gave inconsistent ratings on were:

- **Participant 4**: Dependability, Novelty.
- **Participant 14:** Efficiency, Stimulation.
- **Participant 2:** Efficiency, Dependability, Stimulation.
- **Participant 6:** Attractiveness, Efficiency, Dependability.
- **Participant 7:** Attractiveness, Efficiency, Stimulation.
- **Participant 9:** Perspicuity, Efficiency, Dependability.

- **Participant 13:** Perspicuity, Efficiency, Dependability, Stimulation.

On the Attractiveness scale three participants gave inconsistent ratings. Two gave inconsistent ratings on the Perspicuity scale. Seven participants rated items on the Efficiency scale contradictorily. Items on the Dependability scale were rated inconsistently by six participants. Four participants rated items on the Stimulation scale inconsistently, while items on the Novelty scale was rated inconsistently by one participant.

4.4.4 Separating Consistent and Inconsistent Participants

Since a relatively high number of participants gave inconsistent answers on two or more scales new calculations were made based on only the consistent participants' answers. As shown in Figure 80 this resulted in an increase in the websites overall measured performance on a number of the scales. The score for Attractiveness increased from neutral to slightly positive. Perspicuity and Efficiency increased from neutral to positive. Dependability became slightly more positive. Stimulation stayed neutral (although improving slightly). However, the Novelty score became slightly more negative.

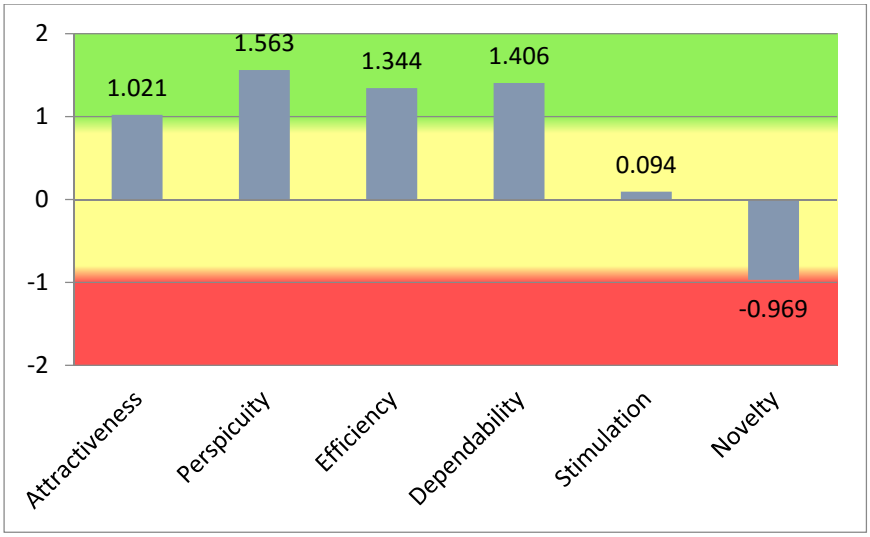


Figure 80: SKAT's Score on Scales, After Inconsistent Participants Were Removed

Figure 81 shows the Attractiveness score, as well as Pragmatic Quality, and Hedonic Quality. After removing the inconsistent answers, the Attractiveness score for SKAT.dk increased, and the Pragmatic Quality score improved, while the Hedonic Quality improved slightly.

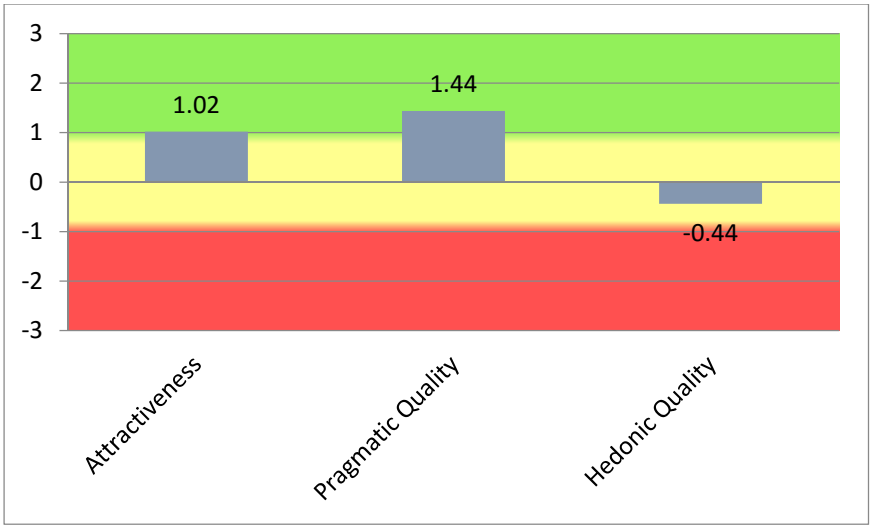


Figure 81: SKAT's Score on Quality After Inconsistent Participants Were Removed

When ratings are calculated on the basis of the inconsistent participants' ratings on all scales become more negative (Figure 82 and Figure 83).

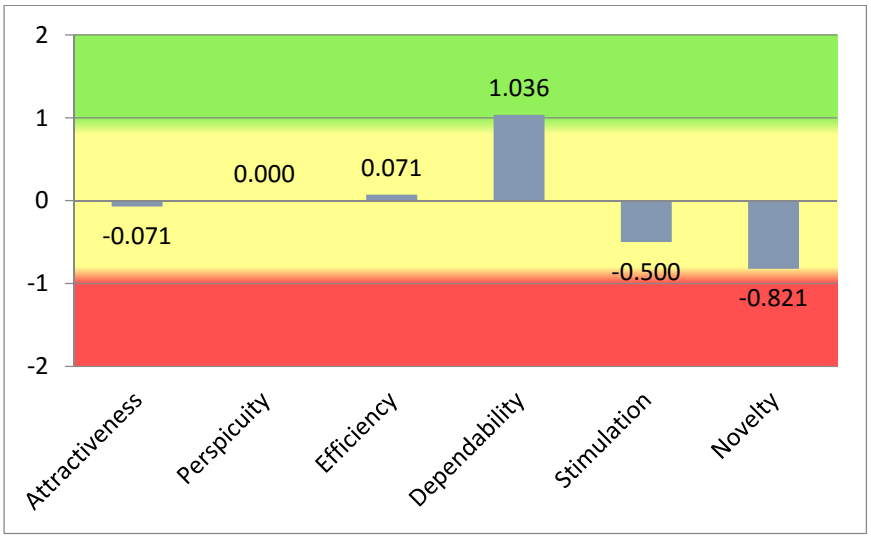


Figure 82: SKAT's Score on Scales, After Consistent Participants Were Removed

The data from the inconsistent participants should be interpreted with care, since the answer behaviour for these participants was somewhat unpredictable. However, the data indicates that these participants may have had a more negative experience than the participants that gave consistent answers on the scales of the User Experience Questionnaire.

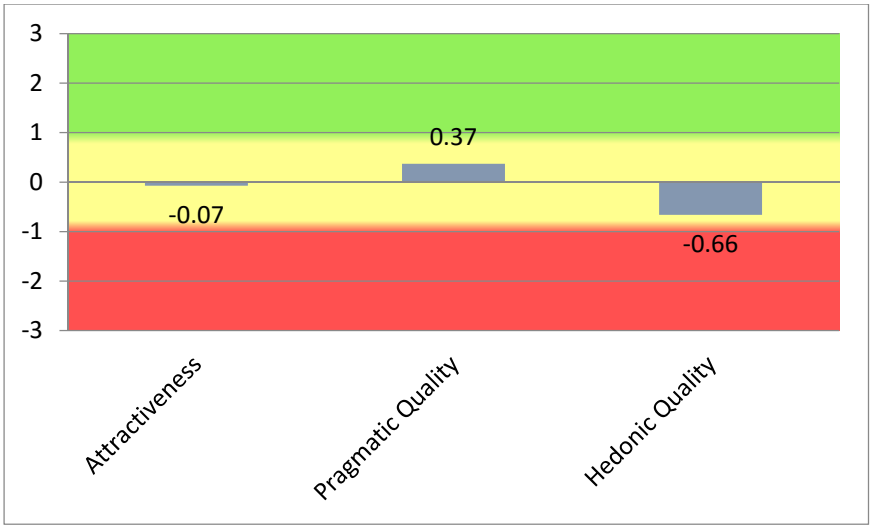


Figure 83: SKAT's Score on Quality After Consistent Participants Were Removed

Since removing the participants that gave inconsistent answers to the User Experience Questionnaire caused changes in the overall rating of the system, differences between the ratings and time spent pr. participant pr. task between the consistent and inconsistent participants were examined.

For Task 1 the participants that gave consistent answers spent between 1.76 and 8.12 minutes solving the task (Figure 84). All participants solved the task successfully. They gave mainly positive or neutral ratings of the difficulty and frustration level experienced during the task. One participant rated the difficulty as a 4, indicating that the participant found the task to be somewhat difficult.

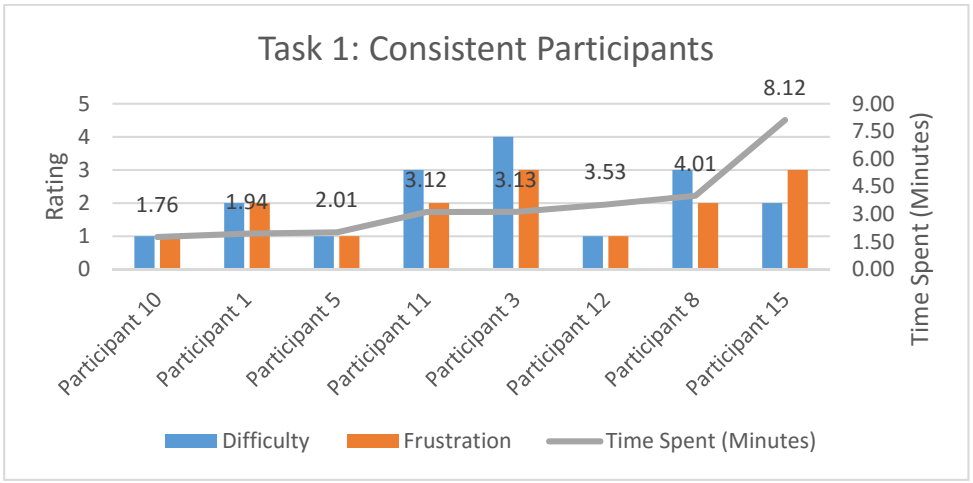


Figure 84: Consistent Participants' Time Use and Ratings for Task 1

In the group of participants that gave inconsistent ratings participants spent between 1.99 and 7.26 minutes on solving the task (Figure 85). Two participants did not solve the task. Ratings were neutral or positive, with the exception of the two participants that did not solve the task, who gave ratings of 3/5 (neutral easiness/frustrating) and 5/5 (difficult and frustrating).

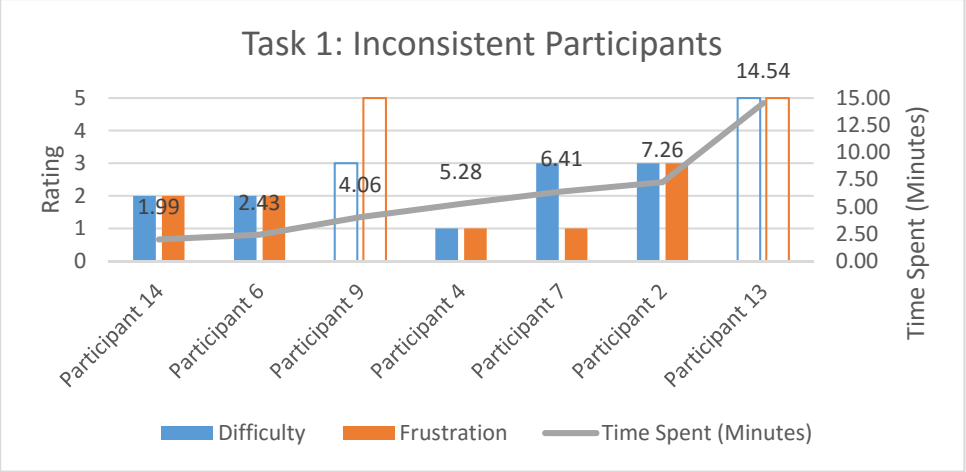


Figure 85: Inconsistent Participants' Time Use and Ratings for Task 1

In the group of consistent participants everybody solved Task 2 successfully. They did so in between 0.91 and 2.47 minutes (Figure 86). The task received one neutral rating and one negative rating, indicating that consistent participants found the task to be easy to solve, and in general they were not frustrated by the task.

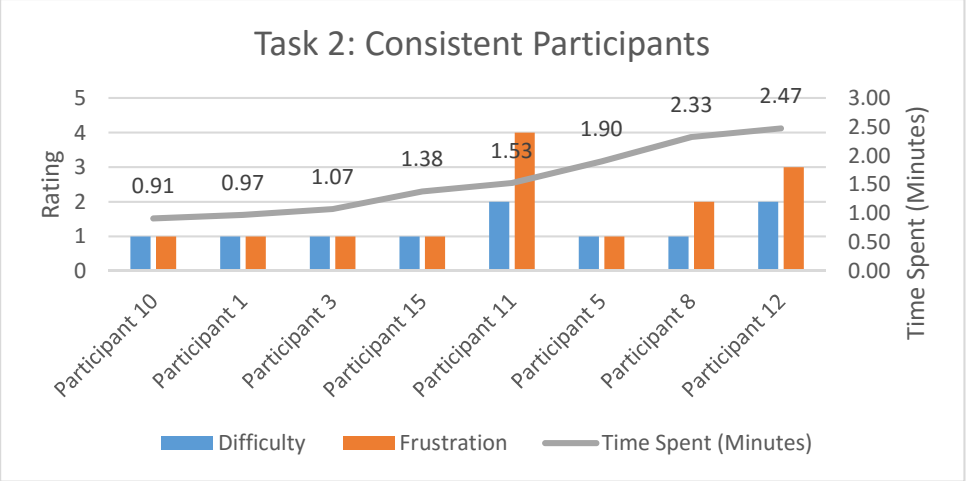


Figure 86: Consistent Participants' Time Use and Ratings for Task 2

However, among the inconsistent participants the time spent solving Task 2 was between 1.12 and 4.55 minutes (Figure 87). One participant in this group did not solve the task. Three of the participants gave a negative rating on at least one of the aspects of difficulty and frustration.

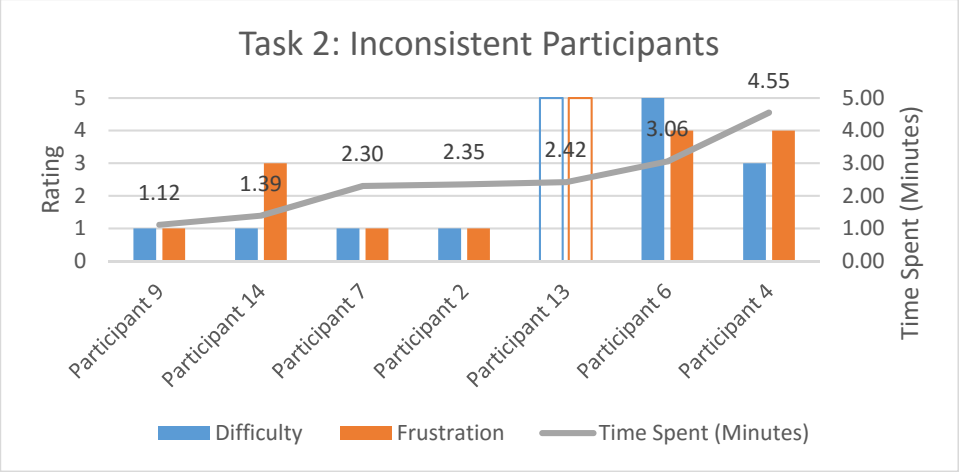


Figure 87: Inconsistent Participants' Time Use and Ratings for Task 2

Task 3 was completed in 0.60 to 2.85 minutes by consistent participants (Figure 88). Ratings were generally positive, although one participant gave neutral ratings on both aspects, while on rated the task as neutral in terms of difficulty and frustrating, and third participant rated the task as difficult, but neutral in terms of level of frustration.

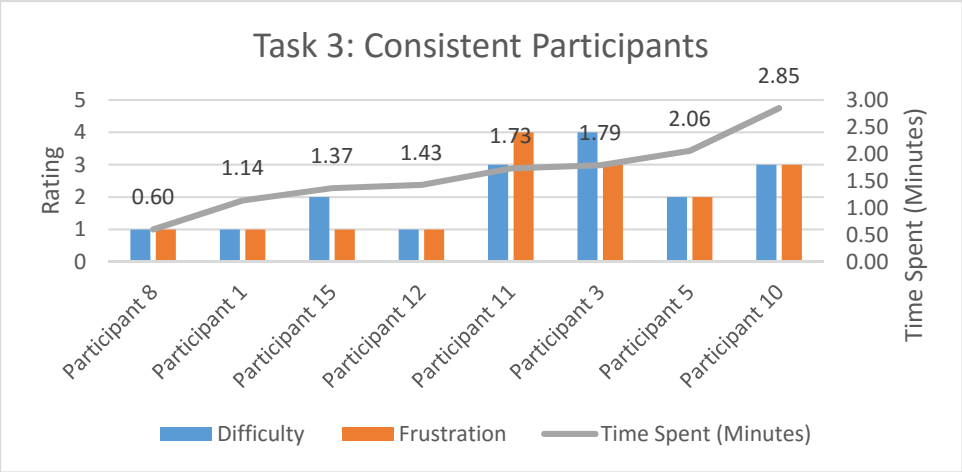


Figure 88: Consistent Participants' Time Use and Ratings for Task 3

The group of inconsistent participants completed the third task in between 1.03 and 3.49 minutes (Figure 89). One participant did not complete the task. The participants rated the task positively, save for the participant that did not solve the task.

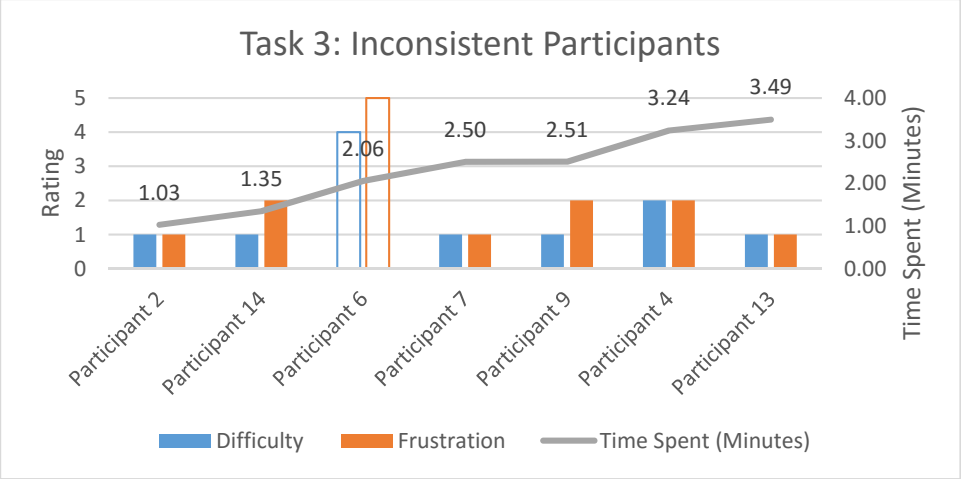


Figure 89: Inconsistent Participants' Time Use and Ratings for Task 3

Task 4 was completed in 1.18 to 5.06 minutes by the consistent participants (Figure 90). Ratings of the task were mainly negative, with only two participants giving the task a positive rating. Two participants failed to complete the task.

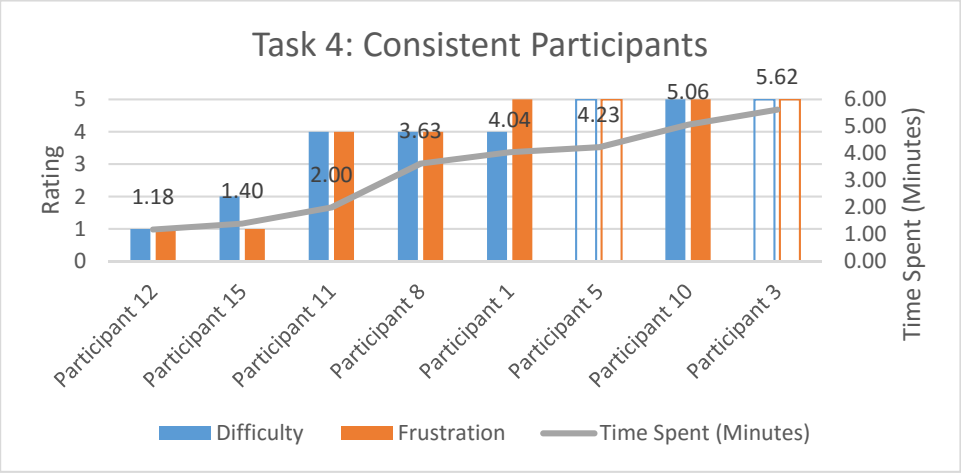


Figure 90: Consistent Participants' Time Use and Ratings for Task 4

The inconsistent participants solved the task in between 1.73 and 7.16 minutes (Figure 91). Ratings were mainly negative, and three participants failed to solve the task.

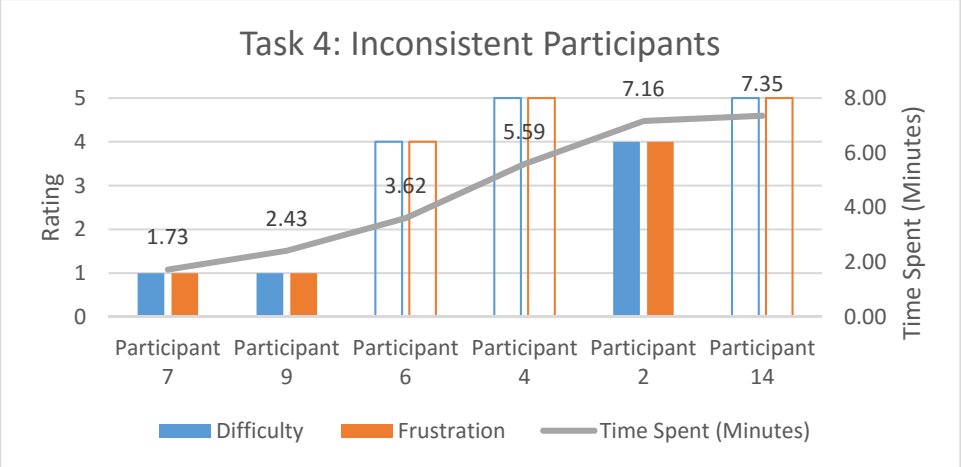


Figure 91: Inconsistent Participants' Time Use and Ratings for Task 4

When comparing the results for the group of consistent participants with the group of inconsistent participants it is seen that inconsistent participants as a group were overall slower at completing three out of four tasks. Participants in this group also failed at completing more tasks than participants in the group of consistent participants.

4.4 Data Analysis

Most participants had visited the website within the last 5 months. There is no obvious connection between when a participant last visited the website and their ability to solve tasks, or their rating of the tasks. Only two participants reported that they did not visit SKAT's website within the last 5 months. They did not seem to perform better or worse with the website than other participants. One participant (13) did not remember when he last visited the website. This participant performed significantly worse than the other participants, solving only one task (Task 3), but required the most time of all participants to do so.

The number of participants that solved Task 1-3 did not differ significantly. However, fewer participants solved Task 4 than the other tasks.

Out of fifteen participants, eight participants solved all the tasks. Five participants solved three tasks, while one participant solved only two tasks, and one solved only one task.

The time needed to solve Task 1 and 4 was higher than the time needed to solve Task 2 and 3. Four participants spent more than the average amount of time on solving Task 1 and 4, while six participants spent more than the average amount of time solving Task 2 and 3.

There does not seem to be any significant difference between male and female participants in regards to the time spent solving task, success rate in completing tasks, or the ratings given. However, for Task 1-3 the female participants' time spent solving tasks varied slightly more, than male participants'. For Task 4 the variation between the fastest and slowest participant was greater among the male participants than that among the female participants. Still, the results indicate no significant difference on completion times between the genders.

The ratings of the task difficulty for Task 1 were primarily positive (easy) or neutral. For Task 2 ratings were mostly positive (easy), with a few rating the task as difficult to solve. Task 3 received mostly positive ratings as well, with a few rating the task as slightly difficult or neutral. Task 4 was primarily rated as difficult, with no neutral ratings, and only relatively few positive ratings as easy. This indicates that Task 1 was perceived as relatively easy, Task 2 and 3 were perceived as easy to solve, while Task 4 was perceived as difficult to solve. Furthermore, the ratings for Task 4 were either positive or negative, and received negative ratings even from participants that solved the task.

Ratings of the level of frustration of Task 1 were mainly positive, indicating that participants were not generally frustrated by solving this task. Task 2 received many positive ratings as well, although some participants rated this task as frustrating or slightly frustrating. Participants generally rated Task 3 as simple as well, with only a few participants rating the task as frustrating. Task 4 was generally rated as frustrating, although some participants gave the task a positive rating. No Neutral ratings were given for Task 4.

There seems to be a tendency that participants give similar ratings on the two scales (i.e. if the task was given a low rating on one scale, it generally receives a low rating on the other scale as well). This may indicate that making it easier for the user to achieve their goal will

also result in less frustration. Furthermore, there seems to be a relation between the time spent by the participant on solving the task, and the level of frustration experienced by the participant.

Task 2 was designed to make participants use the search function, and the task description of this task included a term known to bring up relevant search results. The opposite was true for Task 4, i.e. the term “seminar” used in the task description was known not to provide useful results if used in a search query. Of the fourteen participants that tried to solve Task 4 ten used the search function in their attempt. Almost all of the search queries made by participants for the task included the term “seminar”. Only one query completely excluded the term, and this was incidentally the only search that led to the correct answer.

For Task 1-2 and 4 the majority of participants did not know of the problem presented in the task description before trying to solve the task. For Task 3 ten participants out of fifteen knew the problem beforehand. However, the data collected for this thesis does not indicate any significant relation between prior knowledge and perceived level of difficulty or frustration, or task solution times.

The overall results of the data gathered through the User Experience Questionnaire indicate the User Experience provided by SKAT’s website is relatively neutral, with neutral ratings on the Attractiveness, Perspicuity, Efficiency and Stimulation scales. The Dependability score for the site is fairly positive, while the Novelty score is fairly negative. In regards to the Pragmatic Quality of the website the results are relatively positive, although the score is not overwhelmingly high. However, the Hedonic Quality of the site is evaluated as neutral by participants.

The confidence intervals for the data collected through the User Experience Questionnaire are relatively large. This indicates that the estimation for the mean values are not necessarily very precise. Given the small dataset this is not surprising.

Almost half of the participants gave inconsistent ratings on the User Experience Questionnaire. The scales that received the highest number of inconsistent answers were: Efficiency, Dependability, and Stimulation. Conversely, items on the Novelty scale and the Perspicuity scale were rated in a more consistent fashion.

Because of the high number of participants that gave inconsistent answers an examination of the difference between the participants that gave consistent and inconsistent ratings was made. When using only the data from the consistent participants an increase in the overall User Experience score for the site was seen. The ratings for all scales, with the exception of Novelty, became more positive. However, the score for the Novelty scale became less positive. The score for the Pragmatic Quality increased, and the Hedonic Quality increased slightly. However, when entering only data from the inconsistent participants, the overall score for SKAT.dk decreased on all scales except Novelty. The novelty score improved slightly, but remained on the negative side of the diagram. It is important to note, that due to the answer patterns of the inconsistent participants, their answers should be interpreted with much caution.

Having identified significant differences between the rating of the system by the consistent and inconsistent participants', different aspects of their behaviour were examined. It was found that the consistent participants in general were more likely to solve the tasks than inconsistent participants. In fact, the only task in which consistent participants failed to solve the task was Task 4, whereas at least one inconsistent participant would fail to solve any of the tasks.

5 Discussion

As most participants reported that they had visited the page within the same time period, there was not enough data on participants that had not visited SKAT.dk within the last 5 months, to indicate whether last visited has influenced participants' performance with the site.

Through comparing task solving times with users' ratings of difficulty and frustration of solving the tasks it was found that there seems to be a relation between these aspects. If solving a task takes the user longer than the average time spent by all participants the data indicates that they are more likely to find the task difficult as well as frustrating. When asked to rate the overall experience of interacting with the website in item 1 (annoying/enjoyable) in the User Experience Questionnaire Participant 1 stated: "So then it's pulled in the direction of annoying because of that one experience, which is interesting.". This indicates that the user's overall experience of the whole site may be affected by one frustrating task. This raises the possibility of measuring User Experience objectively by researching the time users need to solve tasks. This would require the users' time to be compared to a set of standard metrics of task time. These standard time metrics would have to be measured first.

Although the ratings for the first task were generally positive or neutral the task did uncover a possible problem with the site. When solving Task 1 some participants came to a wrong conclusion based on the information presented to them by the site, before eventually solving the task. They thought they would not have to pay additional duty when ordering the specified clothing from outside the EU. This may be problematic, since in a real life situation this could potentially lead to misunderstandings.

In the test it was found that the task in which a useful search term was provided in the task description participants spent a shorter time solving the task and rated the task as easier and less frustrating, than in the task were a term known not to be useful in the search was provided. When using the websites' search function, the participants had a tendency to use terms presented in the task description in their search queries. In addition, they would use the provided term even when the term did not provide useful results through the search. If users are likely to stick to the first search term that occurs to them, it may be difficult to use the search function successfully if that term does not prove to be useful. The search results from Task 4 may indicate, that users that are domain novices, and thus do not know the proper terminology, may find it difficult to use the search function at SKAT.dk. Improving the algorithms used for the search function on SKAT's website may result in an improved overall experience for the user.

Participants generally found Task 4 frustrating, and many did not solve that particular task. This may well be a reflection on the task design, more than a reflection on the design of SKAT.dk. However, it does indicate, that in cases where the user does not know the right search terms, they may be unable to find certain information. One particular participant did not know the meaning of the word "seminar", which made it particularly difficult for this participant to complete Task 4.

In regards to the process of construction of the User Experience Questionnaire the translated items would have benefitted from having been tested more thoroughly before being implemented in the User Experience test. At least one mistake was mentioned by test participants, namely that the word “Behagelig” (translated from item 1: “enjoyable” and item 16: “pleasant”) was used on two different items. Some participants were confused by the word occurring more than once, and even went back to check whether they remembered correctly. Furthermore, the relevance of some items in the User Experience Questionnaire may be limited in the context of this particular test. Some participants commented on this as well. In addition, some items caused confusion for some participants. For instance, some participants had difficulty understanding the meaning of item 5 (“Valuable/inferior” or “Værdifuld/inferiør” in the Danish translation) of the Stimulation scale. Some participants expressed a sense of wonder at being asked about certain aspects.

The results of the User Experience Questionnaire should be interpreted with caution, since the confidence intervals are quite large. However, this is not surprising, given the small dataset. The overall User Experience of SKAT.dk is indicated to be neutral in terms of Attractiveness, Stimulation, Efficiency, and Perspicuity. However, both the Stimulation and Efficiency scales received a high number of inconsistent rating, and the results may therefore be unprecise. The Dependability was rated as positive, but had a high number of inconsistent answers, meaning that this score is possibly unprecise as well. The Novelty score was negative, but received a high number of consistent answers, indicating that many participants gave consistent, negative, ratings on the novelty of the site. It is possible that participants did not interpret all items the same way. For instance, the “secure/not secure” item seemed to be interpreted by some participants as a term for data-security, while others interpreted is as whether they felt secure in their interactions with the site. The “meets expectations/does not meet expectations” item was interpreted differently by some participants as well. Some interpreted it as meaning “does the site live up to positive expectations?”, while others had negative expectations of the site and answered the item from that perspective instead.

Although there was a relatively high number of inconsistencies it is unlikely that the participants rated the items of the UEQ randomly, or in a joking manner, since the researcher was present, and they were asked to express their reflections on their answers verbally. It is far more likely that the inconsistencies are the result of misinterpretation of the items.

When separating the participants that gave consistent answers from the inconsistent participants, it was found that consistent participants seemed to have had a better experience when using the site. The consistent participants were more likely to complete the tasks, and all of the fastest participants were in this group. However, the inconsistent participants were more likely not to solve the tasks. In fact, the two participants that solved the fewest tasks were in the group of inconsistent participants. Because of the inconsistent answers given by the inconsistent group the results of the User Experience Questionnaire are likely not precise for this group. However, the drop in scores for many of the scales, when viewing only inconsistent answers, may indicate, that users that experience more

trouble solving their tasks in the system are also likely to have a poorer experience overall. For this reason, the Usability and User Experience design of e-government websites like SKAT.dk is important.

The site generally receives the best scores on the Pragmatic Quality, both when counting only consistent or inconsistent participants, and when counting all participants. This may be because users perceive the site as a tool to solving tasks, more than anything. During the test participant 1 stated that she regards SKAT.dk as a tool: “It should preferably be a tool, shouldn’t it?”.

The behaviour of participant 13 stood out somewhat among the participants. When presented with the first task he seemed apathetic, and stated that he was unsure of what to do to solve the task. He solved the least tasks among all participants, and was the slowest to solve the task he did solve. Although participant 13 fit in the age group of Prensky’s Digital Native, he did not behave as such. In fact, he solved only one of the four tasks, in one case because he was confused by the user interface of Morae.

6 Conclusion

With the contact between citizens and the public sector becoming increasingly digitized, it is gradually becoming mandatory for Danish citizens to use the provided e-government websites. As the number of users that report that they can easily find the information they need on the e-government sites is increasing, a logical next object of study may be the User Experience offered at e-government websites like SKAT.dk.

This thesis sought to answer the questions of which kind of User Experience is offered at SKAT's homepage, where the User Experience on the existing page can be improved, and if there is reason to study the User Experience of SKAT's homepage further. Answers were sought through designing and conducting a User Experience Test among users between the ages of 20-26 years.

Results from the test indicate that there is a connection between the time spent by users on solving a task, and the perceived easiness of the task, and the level of frustration experienced by the user. Furthermore, these factors seemed to influence the overall User Experience of the website as well. Thus, users that can easily find the information they are looking for on the website are likely to spend less time on completing their task and be less frustrated by the process. There is reason to suspect, that one frustrating may influence the users experience of the entire website negatively. The relation between time spent solving a task, the perceived difficulty of the task, and the level of frustration experienced when solving the task could be studied further. If the connection between time, difficulty and frustration proves to be general it might be possible to develop a way of measuring the User Experience objectively through comparing the time spent by the user to a set of standard temporal metrics.

During the test participants that used the search function of the website tended to stick to the first search term that occurred to them, and make only slight changes to their search queries when no relevant search results were found. It is proposed, that the search algorithms used on SKAT's website should be studied in order to improve the search function for users.

The test results indicate that SKAT's website offers a somewhat positive experience when it comes to the Pragmatic Quality (goal directed) aspects, whereas the experience of the Hedonic Quality (not goal directed) and Attractiveness were found to be neutral. Thus the test results indicate that the User Experience could be improved on Pragmatic Quality, Hedonic Quality, and Attractiveness, although the last two were perceived by participants to be the least positive. Especially the aspects covered by the Stimulation and Novelty scales of the User Experience Questionnaire leave room for improvement.

The test results are not generalizable to the wider population. However, they do point to subjects that could be further studied. For instance, while the site does not perform badly on Pragmatic Quality, both the Hedonic Quality and Attractiveness scores are neutral. These aspects could be studied, first to see if the findings presented in this thesis apply to a more

generalizable sample of the population, and second to develop more detailed suggestions for improvements.

7 Reference list

Baumgarten, J., Chui, M. (2009) E-government 2.0. McKinsey on Government, Number 4.

Britten, N. (1995). Qualitative Research: Qualitative Interviews in Medical Research. BMJ, 311:251, DOI: http://dx.doi.org/10.1136/bmj.311.6999.251

Buxton, B. (2007). Sketching User Experiences: Getting the Design Right and the Right Design. San Francisco, CA: Morgan Kaufmann. Available at: https://www-dawsonera-com.zorac.aub.aau.dk/abstract/9780080552903

Danish Agency for Digitisation, The [Digitaliseringsstyrelsen] (2014). Wave 2 on Mandatory Self-Service. [Website] http://www.digst.dk/Servicemenu/English/Policy-and-Strategy/Mandatory-digital-self-service/Wave-2-on-mandatory-self-service

Fasick, F. A. (1977-1978). Some Uses of Untranscribed Tape Recordings in Survey Research. The Public Opiniion Quarterly, 41:4, pp. 549-552. Available at: http://www.jstor.org/stable/2748533

Halcomb, E. J., & Davidson, P. M. (2006). Is Verbatim Transcription of Interview Data Always Necessary? Applied Nursing Research 19, 38-42, DOI: 10.1016/j.apnr.2005.06.001

Hassenzahl, M. (2002). User Experience and Experience Design. In: Soegaard, M. & Dam, R. F. (eds.). The Encyclopedia of Human-Computer Interaction, 2nd Ed.. Aarhus, Denmark: The Interaction Design Foundation. Available at: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/user-experience-and-experience-design

Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons. Morgan & Claypool. [Ebook] DOI 10.2200/S00261ED1V01Y201003HCI008

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda, Behaviour & Information Technology, 25:2, 91-97, DOI: 10.1080/01449290500330331

Helsper, J. E., & Eynon, R. (2010). Digital Natives: Where is the Evidence? British Educational Research Journal, 36(3), pp. 503-530.

Hung, S.-Y., Chang, C.-M., Yu, T.-J. (2006). Determinants of User Acceptance of the E-government Services: The Case of Online Tax Filing and Payment System. Government Information Quarterly 23, 97-122.

Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users, Foundations and Trends in Information Retrieval, 3:1–2, 1–224, DOI: 10.1561/1500000012

Kuniavsky, M. (2003). Observing the User Experience: A Practitioner's Guide to User Research. San Francisco, CA: Morgan Kaufmann Publishers.

Kuniavsky, M. (2010). Smart Things: Ubiquitous Computing User Experience Design. Morgan Kaufmann. [Online]

Kvale, S. (1996). InterViews: An Introduction to Qualitative Research Interviewing. London, UK: SAGE Publications, Inc.

Kvale, S. (Ed.). (2007). Doing Interviews. London, England: SAGE Publications, Ltd. DOI: http://dx.doi.org.zorac.aub.aau.dk/10.4135/9781849208963

Laugwitz, B., Held, T., Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. A. Holzinger (ed.) http://link.springer.com/chapter/10.1007%2F978-3-540-89350-9_6

Lauterbach, T. (2015a). Ni ud af ti kommunikerer digital med det offentlige [Nine Out of Ten Communicate Digitally with the Public Sector]. Nyt fra Danmarks Statistik [News from Statistics Denmark], Nr. 601. Available at: www.dst.dk/nyt/25396

Lauterbach, T. (2015b). It-anvendelse I befolkningen [IT-Use in the Population]. Copenhagen Ø: Danmarks Statistik [Statistics Denmark]. Retrieved from: www.dst.dk/Publ/ItBefolkning

Madsen, C. Ø., Kræmmergaard, P. (2015). Channel Choice: A Literature Review. In: Tambouris, E., et al (Eds.). EGOV 2015, LNCS 9248, 3-18, DOI: 10.1007/978-3-319-22479-4_1

Morville, P., & Rosenfeld, L. (2007). Information Architecture for the World Wide Web, 3rd Ed.. Sebastopol, CA: O'Reilly.

Poland, B. D. (1995). Transcription Quality as an Aspect of Rigor in Qualitative Research. Qualitative Inquiry, 1:3, 290-310.

Prensky, M. (2001). Digital Natives, Digital Immigrants Part 1. On the Horizon, 9(5), 1-6, DOI: http://dx.doi.org/10.1108/10748120110424816

Rubin, J., & Chisnell, D. (2008). Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, 2nd Ed.. Indianapolis, IN: Wiley Publishing, Inc.

Russell-Rose, T., & Tate, T. (2013). Designing the Search Experience: The Information Architecture of Discovery. Waltham, MA: Morgan Kaufmann.

Sanders, L. (2008). An Evolving Map of Design Practice and Design Research. Interaction – Designing Games: Why and How, 15:6, 13-17, DOI: 10.1145/1409040.1409043

Scholl, H. J. (2014). The EGOV Research Community: An Update on Where We Stand. In M. Janssen et al. (Eds.): EGOV 2014, LNCS 8653, pp. 1–16, 2014. IFIP International Federation for Information Processing 2014

Schrepp, M. (2015). User Experience Questionnaire Handbook: All You Need to Know to Apply the UEQ Successfully in Your Projects.

Wellard, S., & McKenna, L. (2001). Turning Tapes into Text: Issues Surrounding the Transcription of Interviews. Contemporary Nurse, 11:2-3, 180-186, DOI: 10.5172/conu.11.2-3.180

Appendix 1 - Transcription

Participant 1

Task 1: Direkte til “Told” > direkte til “Beregn”

“Okay, den er bare langsom”

“Jeg synes det gik relativt nemt for mig, men jeg tror hvis ikke man ved at det er told, altså hvis ikke du kender fagterminologien, så har du problemer. For mig var det nemt nok, fordi at jeg ved at det er told, og fordi jeg har købt ting på nettet før, men men det forudsætter at folk de ved hvad det er for nogle ord, der bliver brugt og de i forvejen har en idé om hvad det er de skal lede efter.”

”Jeg synes det var ret enkelt. Jeg var ikke vildt frustreret. Der var ikke noget der sådan umiddelbart generede mig.”

Task 2: Søger med det samme på ”grænsegænger”, bruger et af de foreslåede søgeord ”grænsegængere”.

Vælger søgeresultat 4, ”SKAT: Grænsegængere”

”Der har de faktisk en fin lille forklaring når man går ind og søger på det.”

”Det var nemt. Der er et søgefelt. Det er nemt at finde søgefeltet. Og det er nemt at bruge. Altså, jeg behøver faktisk bare at skrive grænsegænger, og så kommer der endda forslag op med hvordan jeg skal gøre det. Så det gik nemt. Men hvis ikke man vidste at man kunne få den information derinde, så igen, kunne det være kompliceret.”

”Jeg synes det var meget enkelt. Teksten er også lettilgængelig, når først du kommer ind og læser den. Det er ikke svært for dig at forstå det. Så det var en rar oplevelse.”

Task 3: Straight to ”Forskudsopgørelse” > ”Regn ud hvad du får udbetalt”

”Eller, det kunne være jeg skulle gå ind på SU&SKAT.”

Fills in form.

”Det var nemt. Jeg var lidt forvirret om hvad for et af stederne jeg skulle ind til at starte med. Men det viste sig at jeg kunne gøre det det første sted hvor jeg tænkte det gav mening inde på beregneren, sjovt nok. Så det var nemt og overskueligt.

”Det var også rimeligt let og overskueligt.”

Task 4: ”Erhverv tænker jeg giver mening”

Opens ”Erhverv” > ”Egen virksomhed” menu

”Skat af egen virksomhed”

”Mere” > ”Hjælp til tast selv erhverv”

”Det kunne godt være at det seminarie var en del af deres hjælpepakke”

”Det var et vældig godt spørgsmål, hvor fanden finder man lige det.”

”Kontakt” > ”Moms”

”Okay, nede i bunden der står noget med informationsmøder. Men hvordan fanden man så kom ind til det gennem sådan deres standard … heroppe i toppen, det er jo ikke til at sige.”

Footer > ”Informationsmøder” > ”Aktuelle informationsmøder”

”Den var træls og finde rundt i. Jeg skulle igennem mange steps, hvor jeg tænkte her kunne det ligge, hvor det ikke lå. Og der var ikke åbenlyst et sted, hvor der stod seminarer eller informationsmøder, oppe i den bar de havde oppe i toppen. Jeg skulle hele vejen ind gennem en masse steder og så helt ned i bunden, hvor der faktisk stod noget om seminarer, sådan nærmest gemt væk.[AS1] Det synes jeg ikke er god kommunikation lige der. Det gør det forvirrende, og det gør det bøvlet for folk, hvis de rent faktisk gerne vil forstå hvordan man gør tingene.”

”Det her var frustrerende”

UEQ

”Så bliver den jo trukket i den irriterende retning pga. den ene oplevelse, hvilket er interessant”

”Vi giver den en neutral, fordi de første var ok”

Giver neutrale ratings.

”Skal SKAT’s hjemmeside være kreativ?”

”Det skal helst være værktøj, skal det ikke?”

”Overodnet let nok at bruge siden, hvis man kender terminologien.”

”Den skal heller ikke være spændende, den har et behov for at være funktionel.”

”Det er ikke sådan et sted man lige går ind for at finde ud af et eller andet bare lige fordi man lige sidder og keder sig.”

”Den er hverken utilfredsstillende eller tilfredsstillende. Man finder ud af de ting man skal finde ud af, det meste af vejen.”

”Den er ikke synderligt nyskabende, af en hjemmeside, det må vi nok erkende. Men det er vel også meget, altså det handler vel meget om at det skal være et trygt sted for folk går jeg ud fra.”

”Den vil gerne tage én i hånden.”

”Jeg synes den er sikker nok, man er tryg, men det er fordi den er statsstøttet.”

”Jeg synes ikke at den er hverken motiverende eller demotiverende. Den er som den skal være, men den giver mig ikke mere lyst til og aflevere mine skatter, eller mere lyst til og undersøge hvordan jeg skal, sådan, fortolde mine ting. Den er mere sådan et værktøj jeg bruger når jeg har brug for det. Den giver mig ikke som udgangspunkt lyst til at gøre mere i det. Men det er fordi det der med skatter generelt ikke er særligt sjovt for nogen og den behøver måske ikke være motiverende. Jeg tænker også at det umiddelbart er et svært emne at motivere til.”

”Praktisk, ja det må man sige, det er den. Den har et formål, og det er et praktisk formål, og det bruger man den til.”

”Når man går ind og læser i teksterne så er de konkrete, de er tydelige, men man fornemmer også at de i hvert fald er forsøgt og være, sådan, ikke positive nødvendigvis, men de er forsøgt at være letlæselige, hvilket giver et indtryk af tryghed og venlighed.”

”De forventninger man umiddelbart har er at den er træls, ikke? Man forventer at man går ind på SKAT’s hjemmeside, så er det noget rod og man kan ikke finde noget og det er svært og dumt. Det levede den op til med den sidste opgave.”

Participant 2

Task 1: Hurtigt til den rigtige underside, men finder frem til det falske svar (at man ikke skal betale moms på varen). Går derpå til underside jeg ikke har set før, "Toldtariffen". Klikker ufortrødent rundt på undersiden, selvom den ikke umiddelbart ser ud til at have svaret. Bliver på et tidspunkt bedt om at vende tilbage til forsiden.

"Når jeg ser at det er deroppe, så er det jo nemt nok"

Task 2: Forside > læser menupunkter I top navigationen > åbner "Mere" drop-down menu > læser menupunkter I dropdown.

Går til "Jura" underside.

Går til "Borger" underside > søger på "Grænsegænger" > vælger resultat nr. 1 "Grænsegængerreglen"

Task 3: Tilbage til forsiden.

Borger > Forskudsopgørelse > Regn ud.

Task 4: Meget fokuseret på top-navigationsmenuen.

Søger til sidst på "Seminar" > vælger resultat 1 "SKAT: Offentligt arrangement reklameudgift"

Vender tilbage til søgeresulatater.

Vælger "Anmodning om tilbagebetaling moms – momsregistreret – seminar"

Tilbage.

Vælger "Undervisning"

Ændrer søgestreng "Seminar for virksomheden"

Vælger "Uge 47: Rift om SKAT og Webinars".

UEQ:

Ja, altså, hjemmesiden er irriterende. Den er på ingen måde behagelig. Men den er let at forstå, når man først lige kommer ind I det. Så virker det. Men så er den kreativ, fordi den er virkelig ikke logisk placeret. Deres ting. Som sagt, let at lære. Man finder hurtigt ud af det. […]

Hjemmesiden er kedelig, og på ingen made spændende at se på […] Det lidt interessant at lære tingene, men … lige pt. skal jeg ikke starte virksomhed, så det er ikke noget jeg skal tænke på nu. […]

Og som sagt opfindsom, det bliver en 2'er, fordi at, igen den er kreativ med måden de har placeret tingene på. […] Det hindrer lidt mere end det støtter, at det er sådan. […]

Det er demotiverende. Og skulle søge rundt på alle tingene, istedet for at det bare skal være der med det samme.

Det lever helt klart op til forventningerne. (Og hvad var forventningerne?) Pissebesværligt at finde.

Klart ineffektivt, da det ikke er nemt nok at finde. Som sagt, der er, hvad er det, nihundrede-et-eller-andet-tusinde mennesker, der I Danmark ikke, eller der betaler alt for meget I skat, fordi de ikke ved hvordan de skal gå ind og ændre det.

Participant 3

Task 1: Kommer til det forkerte svar, at man ikke skal betale told. Bliver bedt om at finde beregneren. Finder derefter beregneren.

”Det var ikke meget tydeligt hvad man skulle og hvor man skulle finde det henne”

”Det var frustrerende at lige umiddelbart at finde ud af hvor jeg skulle hvor og hvordan og hvorledes og hvad det helt præcis var jeg skulle, men at løse opgaven var jo egentligt ikke særligt svært, så den må ligge et eller andet sted i midten. […] Selve opgaveløsningen var meget let, men og finde frem til, hvor jeg skulle løse opgaven, var ikke særligt let […]

Task 2: Søger fra undersiden på ”Grænsegænger”, vælger ”Grænsegængere” i autocomplete-funktionen. Vælger søgeresultat 4 ”SKAT: Grænsegængere”.

”Det var sgu egentlig meget let […] Deres søgefunktion fungerer.”

Task 3: Kigger på undermenuerne under Borger ”Forskudsopgørelse”, ”Årsopgørelse” > Betal skat for 2015> Betal din skat for 2015 (underside med NemID login) > Tilbage til ”Betal skat for 2015” > Beregn skatten 2015.

Tilbage til ”Forskudsopgørelse > ”Regn ud hvad du får udbetalt”

”Det var en lille smule bøvlet, egentlig. Det havde været nemmere bare og åbne lommeregneren på computeren og regne de 37% af indkomsten ud […] Det var lidt mere bøvlet end jeg syntes det burde være.”

Task 4: Går ind under ”Erhverv” fordi det er for virksomheder. Kigger på virksomhedskalenderen. Går til ”Tast-selv erhverv”.

Tilbage til ”Erhverv” > vælger ”Selskaber” i menuen på Erhverv-undersiden > læser opgavebeskrivelsen > Søger på ”seminar” i søgebaren i top navigationen, kigger søgeresultaterne igennem > går til ”Erhverv” > ”Jura” > ”Kontakt” > ”Erhverv”, åbner alle dropdowns > går til ”Borger” og åbner alle dropdowns > ”Kontakt” og bruger udersidens specifikke navigation …

Ender med at give op efter at have undersøgt de fleste undersider, der er umiddelbart tilgængelige fra forsiden.

UEQ:

Når man finder de ting man vil finde derinde, jamen så virker det egentligt som om det fungerer udmærket, men det er det der med lige at finde rundt og finde ud af hvor man skal finde det, det kan godt være lidt bøvlet.

Den virker ikke sådan særligt kreativ, den er sådan lidt bare … plain.

Altså, jeg tror. Den er. Når først man finder ud af hvor tingene er, jamen så tror jeg egentligt den let nok at finde rundt i, deres hjemmeside. Når man kommer lidt i den retning.

Den er ikke særlig spændende. Og egentlig heller ikke særligt interessant.

Mange af tingene er rimeligt forudsigelige og til at finde ud af og til at finde, og så er der nogle af tingene, som åbenbart slet ikke er til at finde.

Den er meget konventionel.

Lidt over i nærheden af det hindrende fordi, altså. Ja, det ved jeg ikke. Jeg synes ikke den hjalp mig ret meget. Jeg ved ikke om det er dens opgave at gøre det, men … det synes jeg ikke den gjorde.

[…]

Men det ved jeg ikke, altså. Revisorer og så videre de skal vel også have noget at tjene deres penge på, ikke, altså? Det duer jo ikke hvis vi bare selv kan finde ud af at finde det hele, ikke.

Jamen, hjemmesiden som helhed er jo egentligt rimelig enkel, ikke.

Det vækker ikke forfærdelig mange følelser i det, i mig, den her hjemmeside.

Meget demotiverende fordi det er kedeligt, og træls. Ikke noget vi gider at beskæftige os med.

Ja, og den lever op til mine forventninger, fordi den er kedelig og træls, og det er skat's hjemmeside, så det er lidt det jeg forventede af den.

Den må være forholdsvist ineffektiv, fordi den er ikke sådan nødvendigvis skidenem at finde rundt på.

Den kan være forvirrende. Alt efter hvad du skal finde. Og er ikke nødvendigvis skide praktisk.

Tingene er godt organiseret og ikke nødvendigvis overfyldt, men det er så også problematisk nok, fordi den kunne godt være overfyldt med nogle af de ting som man gerne ville finde [...] der mangler nogle ting, altså, der kunne godt være flere muligheder, synes jeg. Som eventuelt kunne gøre det nemmere at finde nogle af de ting, man leder efter. Men ja, både og.

Konservativ, det er den i hvert fald.