Usability research on three video-based telemedicine software used in healthcare

Timo Hantunen

Master's thesis
February 2022
Information and Communication Technologies
Full Stack Software Development

Jyväskylän ammattikorkeakoulu
University of Applied Sciences

Description

Hantunen, Timo

Usability research on three video-based telemedicine software used in healthcare

Jyväskylä: JAMK University of Applied Sciences, February 2022, 70 pages

Information and communication technologies. Master's Degree Programme in Information Technology, Full Stack Software Development. Master's Thesis.

Permission for web publication: Yes

Language of publication: English

Abstract

The purpose of the thesis was to study the state of usability and workload of video-based telemedicine software when subjects use three different software utilized in healthcare and gather information about what makes usability good or worse in telemedicine software. The thesis was a part of DigiSote and OSSI projects funded by the ELY Center of South Savonia from the European Social Fund and the Ministry of Social Affairs and Health.

The method of the thesis was to act a made-up video-based telemedicine visit for control visit for high blood pressure where the subject would perform as a patient. Every participant conducted the same telemedicine visit with three different video-based telemedicine software and after using each software the questionnaires were answered. The questionnaires were Telehealth Usability Questionnaire (TUQ) and NASA Task Load Index Test (NASA-TLX). Also, in the end of the test qualitative comments were gathered from the participants. There were 21 participants taking part in the study, and they all were teachers or students at South-Eastern Finland University of Applied Sciences (XAMK)

The state of usability was found good. The tested software received average results 4.35 ± 0.80, 3.39 ± 1.16 and 4.13 ± 0.84 from TUQ. The modified Telehealth Usability questionnaire used in the thesis evaluated the usability with Likert-scale from 1 to 5, 5 being the most positive option. Also, the workload while performing different tasks in made-up telemedicine visit was low. In NASA-TLX questionnaire the average results were 1.30 ± 0.54, 2.24 ± 1.10 and 1.62 ± 0.84 with scale from 1 to 5, 1 being the least workload. Still, there were some differences in the usability of the software. The most significant differences were in the Ease of use and learnability section of TUQ. The cause of this may be in login-phases of software. For example, software 2's login-phase received bad qualitative comments and software 2 received the lowest average score in Ease of use and learnability section. The lowest difference was in the Interaction quality-section, but this was mainly because the communication with the doctor was made-up.

The results of the thesis contain some limitations; all users had higher education background and therefore the results cannot be generalized to make assumptions on whole population-level concerning the usability of video-based telemedicine software. The results also indicate the state of usability during the first-time when using the software tested in this study.

Keywords/tags (subjects)

Usability, User experience, Video based telemedicine, Digital services, healthcare

Miscellaneous (Confidential information)

n/a

Kuvailulehti

Hantunen, Timo

Usability research on three video-based telemedicine software used in healthcare

Jyväskylä: Jyväskylän ammattikorkeakoulu, Helmikuu 2022, 70 sivua.

Tietojenkäsittely ja tietoliikenne. Tieto- ja viestintätekniikan tutkinto-ohjelma, Full Stack Software Development. Opinnäytetyö YAMK.

Verkkojulkaisulupa myönnetty: Kyllä

Julkaisun kieli: englanti

Tiivistelmä

Opinnäytetyön tarkoituksena oli tutkia kolmen eri terveydenhuollossa videotapaamiseen käytettävien ohjelmistojen käytettävyyttä ja kuormittavuutta. Tarkoituksena oli myös kerätä tietoa ohjelmistojen käytöstä ja selvittää, mikä tekee videotapaamisien käytettävien ohjelmistojen käytettävyydestä hyviä tai huonoja. Opinnäytetyö tehtiin osana DigiSote- ja OSSI-hankkeita, joita rahoitti Etelä-Savon Ely-keskus Euroopan sosiaalirahastota ja Sosiaali- ja terveysministeriö.

Opinnäytetyön menetelmänä oli järjestää keksitty lääkärin etävastaanotto, jossa kommunikaatio tapahtuu videon kautta. Vastaanoton aiheena oli korkean verenpaineen kontrollikäynti ja koehenkilö toimii potilaana. Jokainen koehenkikö suorittaa saman etäkäynnin kolmella eri vdeo-ohjelmistolla ja jokaisen ohjelmiston käytön jälkeen täyttää kyselyn, joka koostui Telehealth Usability Questionnairesta (TUQ) ja NASA Task Load Index'istä (NASA-TLX). Kun koehenkilö oli suorittanut etäkäynnin kaikilla kolmella ohjelmistolla, koehenkilöltä pyydettiin laadulliset kommentit ohjelmistojen käytöstä. Tutkimukseen osallistui yhteensä 21 koehenkilöä ja koehenkilöinä olivat Kaakkois-Suomen ammattikorkeakoulun (XAMK) opettajat ja opiskelijat.

Käytettävyys arvioitiin hyväksi tutkimuksessa. TUQ-osiossa ohjelmistot saivat arvosanat 4.35 ± 0.80, 3.39 ± 1.16 and 4.13 ± 0.84. Muokatussa TUQ-kyselyssä käytettiin Likert-asteikkoa 1–5, jossa 5 oli positiivisin vaihtoehto. Ohjelmistojen käyttö huomattiin tutkimuksessa kuormittavuudeltaan alhaiseksi. NASA-TLX-osiossa ohjelmistot saivat arvosanat 1.30 ± 0.54, 2.24 ± 1.10 and 1.62 ± 0.84 asteikolla 1–5, jossa 1 oli vaihtoehdoista kaikkein vähiten kuormittavin. Kuitenkin ohjelmistojen välillä huomattiin eroja. Huomattavin ero oli TUQ-osiossa Käytön helppoudessa ja opittavuudessa. Tämä ero saattoi johtua ohjelmistojen erilaisista sisäänkirjautumisista. Esimerkiksi ohjelmisto 2:n sisäänkirjautumisvaihe sai huonoja laadullisia kommentteja sekä kyseinen ohjelmisto sai myös huonoimmat arvosanat käytön helppous- ja opittavuuskysymyksissä. Pienin eri ohjelmistojen välillä huomattiin vuorovaikutuksen laatukysymyksissä. Tämä johtui mahdollisesti siitä, että kommunikaatio keksityssä etävastaanotossa oli suunniteltu etukäteen.

Opinnäytetyön tulokset sisältävät jonkin verran rajoituksia. Koehenkilöillä oli korkeakoulutaustaa, joten tuloksia ei voi verrata koko väestön osalta. Tulokset kertovat myös ainoastaan käytettävyyden ja kuormittavuuden, kun käyttäjä käyttää kyseisiä ohjelmistojen ensimmäisen kerran.

Avainsanat (asiasanat)

Käytettävyys, käyttäjäkokemus, käyttäjäkeskeinen suunnittelu, etäpalvelut, videovälitteinen vastaanotto, terveydenhuolto

Muut tiedot (salassa pidettävät liitteet)

n/a

Contents

Figures

Tables

1. Introduction

There is a clear need for digital services in health and social care sector. According to Hyppönen et al. (2018), the citizens see that multiple benefits can be achieved with digital services and those can be more efficient care, for example money and time savings compared to on-site visits. However, the biggest obstacles for digital services are that on-site visits can be replaced with digital services, the terms of use and manuals are too difficult, and users lack ability to use digital services. Hyppönen & Ilmarinen (2016) have also found out usability is also one obstacle. There is also a clear lack of knowledge about the usability and accessibility of digital services used in health and social care.

The thesis brings more in-depth knowledge about the state of usability of video-based telemedicine software used in healthcare. The research was a part of DigiSote and OSSI projects and the projects were funded by the ELY Center of South Savonia from the European Social Fund and the Ministry of Social Affairs and Health. More information about these projects can be found in chapter 2.2.

1.1 eHealth and telemedicine

eHealth is the top concept of digital systems and services used in healthcare. eHealth can mean different things such as electronic health records (EHR's), electronic prescriptions, remote consulting or electronic decision support and it contains different sub-concepts, one of which is telemedicine. Telemedicine means treatment and care that is given remotely (Reponen, 2015). Telemedicine can mean communicating via audio and video and it can be synchronous or asynchronous. Asynchronous telemedicine is mostly used when direct communication is not required between, for example a patient or a physician. The early telemedicine was mostly asynchronous but as technology has evolved the use of synchronous communication has increased. Synchronous communication can be educationally based, physician consulting another physician or patient visiting physician at a distance. Also, remote monitoring of blood pressure or blood sugar levels can be seen as a form of telemedicine (McGowan, 2008).

1.2 Video-based telemedicine appointments

Video-based telemedicine appointments are executed so that the physician communicates via video and the patient is alone or with another healthcare professional. Video-based telemedicine allows both physician and patient to effectively communicate with geographical separation using devices such as smartphones, tablets or webcam enabled computers. (Reponen et al., 2015; Dhahri et al., 2020). According to Rodriquez et al. (2021), the coronavirus disease pandemic has increased the use of telemedicine, both video and telephone visits. Comparing video and telephone visits shows that video visits allow a physical examination and a more personal connection between a physician and a patient. On the other hand, video visits require a video-enabled device, digital literacy and broadband internet access which may be an issue for some. (Rodriguez et al., 2021) For effective video visits the telemedicine system must provide adequate quality video and audio in real-time so that visits are much like onsite face-to-face consultation experience (Malindi, 2011).

Telemedicine visits have been shown to have many benefits. Qiang & Marras (2015) conducted a survey of 34 patients with Parkinson's disease who had telemedicine experience. They found out that patients reported saving an average $200 (CAN) and 209 minutes in travel time due to telemedicine (Qiang & Marras, 2015). In the research of Nord et al. (2019) 650 patients were surveyed after receiving care from an on-demand telemedicine program with an emergency physician. Nord et al. (2019) inquired patients about alternative care to telemedicine visits that they had and found out that 70 patients would have done nothing, and they would not have seen an emergency physician. Nord et al. (2019) determined the cost impact of the on-demand telemedicine visit and outcome of that was net cost savings per telemedicine visit was calculated to vary from $19 - $121 (USA) per visit. Results of the research of Nord et al. (2019) follows the results of the research of Qjang & Marras (2015) on cost-savings and raises the possibility that telemedicine video visits increase the patients' access to healthcare services.

2 Research

2.1 Research question

Despite several research in the world agreeing that patients feel generally satisfied with telemedicine visits via video (Agnisarman et al., 2017; Layfield et al., 2020; Patel et al., 2021; Thelen-Perry et al., 2018), there is still a lack of Finnish research in that area. There is no clear vision if the video-based telemedicine software in Finland are usable for meeting healthcare personnel in the distance. Because of a lack of vision there also is not any clear rules or guidelines on how to design usable software for video-based healthcare meetings and what could be potential stumbling blocks. Those studies that have been made about the usability of video-based telemedicine software have been made outside Finland and usability being a cultural issue makes it hard to reliably apply those studies directly to Finnish circumstances. That is why it is necessary to have direct feedback from patients about the usability of software used in healthcare video-visits in Finland.

The aim of this research is to study the usability of video-based telemedicine software when patients use three different software used in healthcare and gather information about what makes usability good or worse in telemedicine software. This study was part of two different projects that are introduced in chapter 2.2. The software tested in this research were used in these projects, which was the reason they were part of the study. As a result of this study healthcare providers can see if patients are able and feel comfortable to use video-based software in Finland and software providers are able to get guidelines for designing usable software for video-based meetings to healthcare appointments. The aim of this research is to purely focus on the usability aspect of video-based telemedicine software and not for example on information security.

Research question 1: What is the state of usability of tested video-based telemedicine software from a patient's point of view?

Hypothesis 0: The patients feel comfortable using telemedicine software and the perceived usability is good. This has been the case in different studies outside Finland (Agnisarman et al., 2017; Patel et al., 2021; Layfield et al., 2020; Thelen-Perry et al., 2018)

Hypothesis 1: The state of usability is bad, and patients do not feel comfortable using telemedicine software.

Research question 2: What affects usability negatively in video-based telemedicine software from a patient's point of view?

Hypothesis 0: The negative effects have been in initiation phase or with bad connection that reflects audio and video that has shown in the research of Agnisarman et al (2017) and Thelen-Perry et al. (2018).

Hypothesis 1: The negative effects cannot be targeted.

2.2 Projects of this research

2.2.1 DigiSote-project

DigiSote-project was a social and healthcare digitalization project in South Savonia, which aim was to develop skills and competences of social and health sector personnel as users of digital applications and as developers of digital services and work processes. The project also carried out for example experimenting and modeling video-based telemedicine visits in health and social care, developing children's and youth's self-services with digital applications and created a model for the introduction of digital technology for professionals. The completion time of the project was between 2016 and 2018 and the partners were South-Eastern Finland University of Applied Sciences which was the administrator of the project, Diaconia University of Applied Sciences, The South Savonia social and health care authority (Essote) and East Savonia healthcare district (Sosteri). DigiSote-project was funded by the ELY Center of South Savonia from the European Social Fund (Eura2014, n.d),

2.2.2 OSSI-project

In the OSSI-project (in Finnish: Etelä-Savon asiakaslähtöinen palveluohjausverkosto ja osaamiskeskus omais- ja perhehoitoon) the aim was to develop more equal, coordinated and rise of expense restrained family care. As a result of the OSSI-project an OSSI-center was created to South Savonia, where family caregivers can seek counselling and guidance. Also, digital solutions in family care were developed in this project. There was a total of 13 partners in this project such as South Savonia social and health care authority (Essote) which was the administrator of the project, South-Eastern Finland University of Applied Sciences, Diaconia University of Applied Sciences and East Savonia healthcare district (Sosteri). The project was funded by the Ministry of Social Affairs and Health (XAMK, n.d).

3 Human-Computer Interaction

Human-computer interaction (HCI) is a research field that studies designing, evaluating and execution of interactive computer systems for the use of humans and phenomena that are associated with those. The mission of HCI is to identify features in information technology and usage situations that must be taken care of when designing computer products. Human-Computer interaction is a field that is a combination of different research fields, such as psychology, ergonomics and sociology. (See Figure 1). (Majrashi & Hamilton 2014; Oulasvirta et al. 2011, pp. 15).

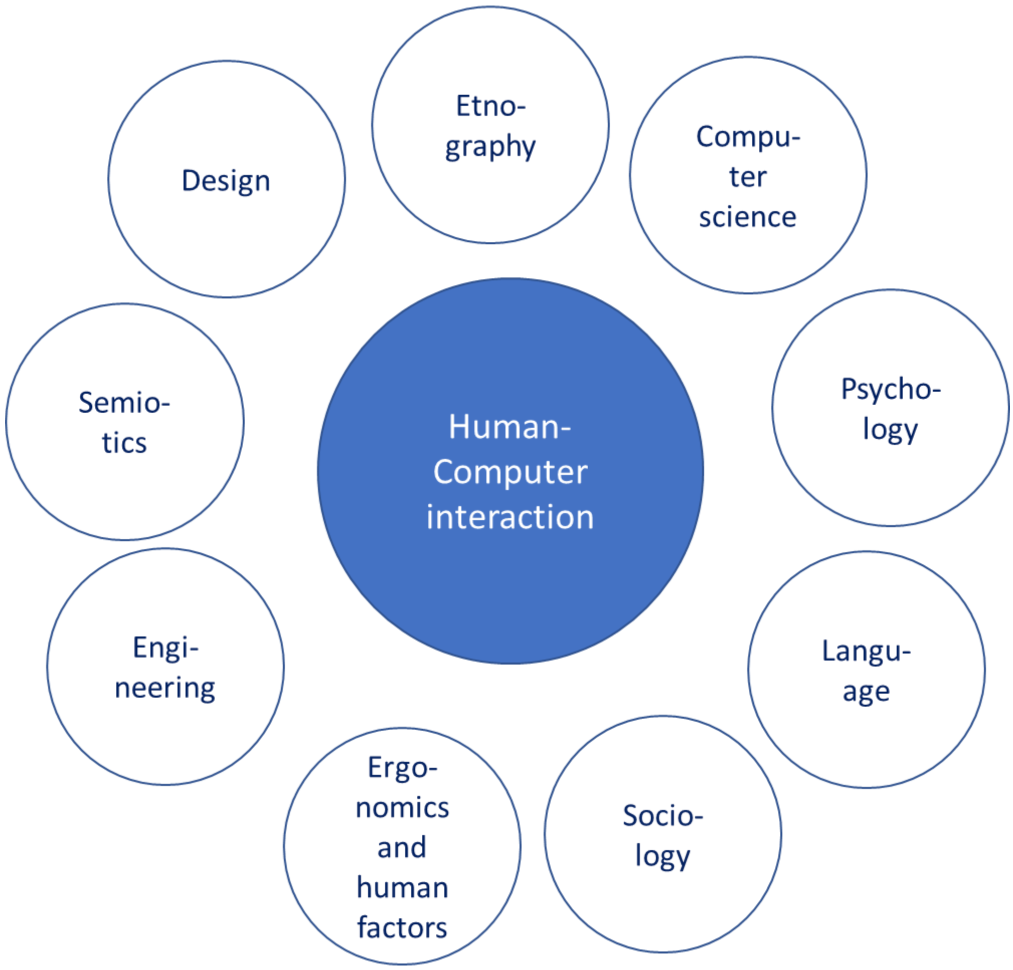

Figure 1: Human-computer interaction and related research fields

What makes HCI more interesting is that though it is a combination of different sciences it cannot be directly seen as just a combination of different fields. Meaning HCI is an own adaptation of science that has evolved and delivered its own rules and norms. This is caused by the diversity of usage phenomenon. The other reason for HCI ‘s adaptation to its own science is the meaning of IT to humans nowadays and how IT has started to change the essence of being a human thinking how many hours a day humans use different applications and computers. (Oulasvirta et al., 2011, pp. 17)

There are many views on what happens when a human interacts with a computer and how it should be described. Sonderegger et al. (2019) uses a term User Experience (UX) as an umbrella construct which encompasses the whole experimental space of human interacting with computer which is shown in Figure 2. This thesis focuses on usability which is a part of UX. According to Sonderegger et al. (2019) user experience also contains components such as affect, trust, value and aesthetics.

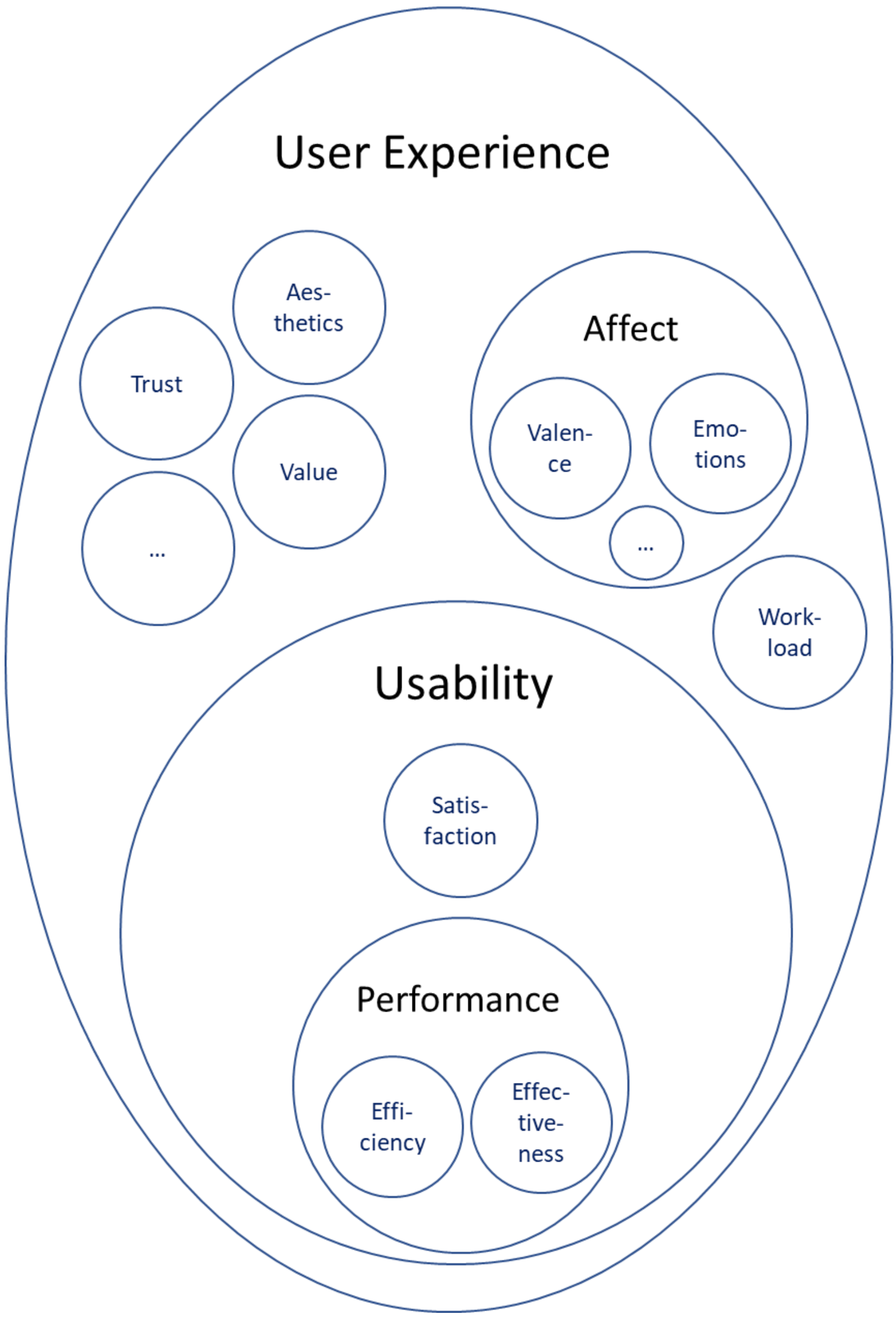

Figure 2: The user experience construct and its components

4 Usability

There are many ways to define usability. Usability can be seen as an essential goal of a product when the product design is based on users' needs and requirements. This means that a usable product meets the users' requirements, and the product supports the users' actions (Oulasvirta et al., 2011, pp. 103-104). Another way of defining usability is to quote Jakob Nielsen who said that usability is a combination of five different attributes of a product or a system. The attributes are learnability, efficiency, memorability, errors, and satisfaction (See Figure 3) (Nielsen, 2010, pp. 26).

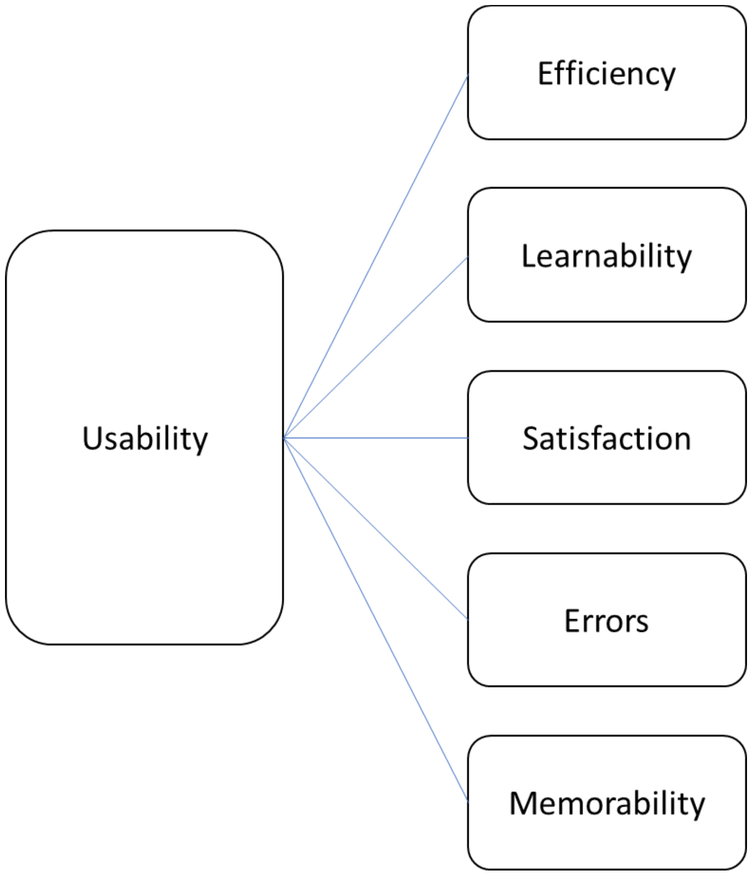

Figure 3: Nielsen’s usability model

Efficiency references the user's ability to gain high productivity using a product. Learnability means how easily a user can learn to use a product or a software. User satisfaction is a key factor in usability especially when satisfaction using a product is targeted. Errors and how you recover from those are of course a big factor in usability. Memorability signifies that a user can after a long pause use a product and the user remembers the product easily (Nielsen, 2010, pp. 26).

Nielsen also has another way to define usability. You can design or evaluate a product’s or software’s usability following 10 different heuristics. Heuristics means guidelines and rules that interface, and software should follow to obtain good usability (Nielsen, 2010, pp. 115.). Those heuristics are listed in Table 1 below according to Nielsen (2020):

Table 1: 10 usability heuristics according to Nielsen

Heuristic	**What does it mean?**
#1 Visibility of system status	The interface should always keep users informed about what is going on and what is the current system status. This information should be presented as quickly as possible.
#2 Match between system and the real world	The software should use the same language as the users' do. This means using words and phrases that are familiar to the user and icons and concepts that are logical to the user.
#3 User control and freedom	Users sometimes make mistakes so the software should help to recover from those. This is easily done by supporting Undo- and Redo-commands and offering a clear way to exit current interaction like a Cancel-button.
#4 Consistency and standards	The interface should use the same words, icons and actions that mean the same thing all over the interface. Good tip is to follow platform and industry conventions.
#5 Error prevention	As told before users make mistakes but the best solution is to design software so that the user cannot make mistakes. So, the error-prone conditions in the software should be removed.

#6 Recognition rather than recall	User's memory load should be minimized so the elements, actions and options should be well visible, and the needed information should be available in the interface.
#7 Flexibility and efficiency of use	Shortcuts for expert users should be provided and there should be even a possibility for users to tailor some actions.
#8 Aesthetic and minimalist design	Interfaces should only contain the information needed and all the irrelevant and even rarely needed information should be removed. Every extra unit diminishes the relative visibility of an important unit.
#9 Help users recognize, diagnose, and recover from errors	When errors occur, good and informative error messages are important. The error messages should for example be expressed in plain language with no error codes. And the error messages should be traditionally visualized such as bold and red text.
#10 Help and documentation	In optimal circumstances there should not be any additional help or documentation but however those may be necessary to help users understand how to complete their tasks.

4.1 What affects usability?

Usability, not being easy to define perfectly, it is also a bit trivial to measure reliably. The reason for that is usability is not constant because usability changes for example during longitudinal usage of the same product. Karapanos et al. (2009) has identified different phases of product use and how phases affect usability and what users appreciate in different phases. For example, during a user's first experience with a product the most appreciated sectors of usability by the user are learnability and aesthetic stimulation. When the product starts to be part of the user's daily life, long-term usability and usefulness starts to become more important to the user.

The same effect was found when Sonderegger et al. (2012) conducted a multiple-session usability test with 60 participants. The participants used two different mobile phones for two weeks each, the other phone being more appealing than the other, and the participants were given different tasks to complete daily with the mobile phone (See Figure 4). Then participants completed questionnaires consisting of usability and aesthetic in the beginning of mobile phone usage and after seven and 14 days of use. The biggest finding was that the positive effect of the more appealing phone on usability started to vanish when the exposure time increased. The study of Sonderegger et al. (2012) indicates that product aesthetics influences usability and this should be considered when conducting one-off usability tests.

Figure 4: Two versions of mobile phone: (a) aesthetically unappealing design and (b) aesthetically appealing design used in the study of Sonderegger et al. (2012)

The concept of usability is also connected to the user’s context which makes usability also a cultural issue. This means it is hard to create a universal approach to usability because external factors play an important role in how users use different products. According to Amant (2017) these different external factors should be taken care of when designing products especially in healthcare settings (Amant, 2017).

4.2 Effects of usability

Davis (1993) developed a model called Technology Acceptance Model (TAM) that predicts user acceptance of computers and IT. According to Davis (1993) there are two specific variables, perceived usefulness and perceived ease of use, that are crucial for technology acceptance (See Figure 5).

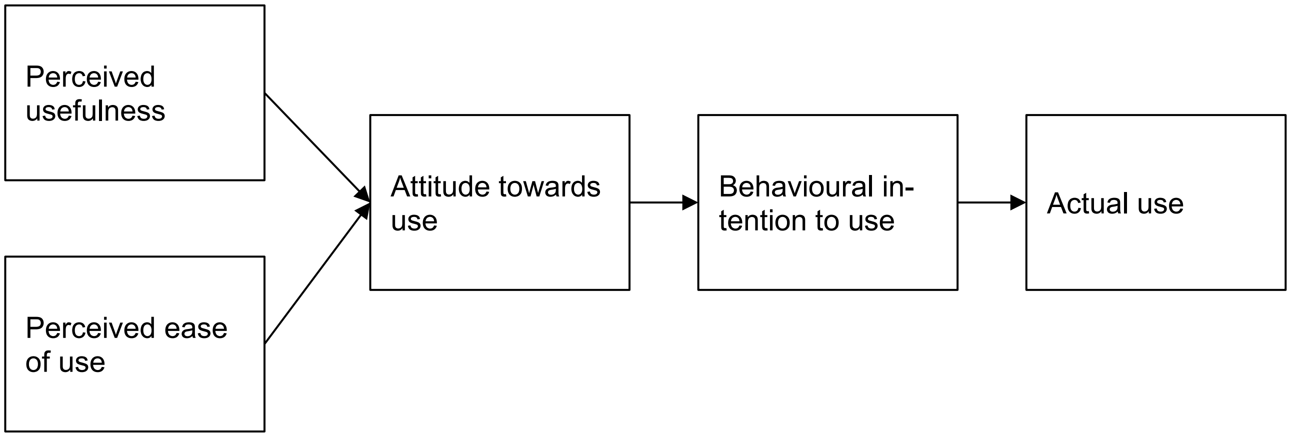

Figure 5: Technology acceptance model (TAM) adapted from Davis (1993)

Despite the Technology Acceptance Model being over 40 years old and that there has been developed new models that predict technology acceptance, TAM is still used widely in many areas for example in healthcare (Klaic & Galea, 2020; Knox et al., 2020; Nguyen et al., 2020).

Lah et al. (2020) wanted to study further the Technology Acceptance Model and how usability affects it. In this research the subjects used three different software products and after using each product the subjects rated their actual experience with three different surveys. Two of these surveys were usability surveys, System Usability Survey (SUS) and Usability Metric for User Experience (UMUX-LITE), and the third survey was a modified survey of Technology Acceptance Model. As a result of this study, they found out correlation between perceived ease of use and usability. So, according to the research of Lah et al. (2020) it seems that usability has an effect also on technology acceptance.

Also, Jakob Nielsen has defined the connection between usability and technology acceptance. Instead of Davis's Technology Acceptance Model Nielsen's models are applied more to IT systems and Nielsen prefers to use term system acceptability which according to Nielsen means if the system is good enough to satisfy the needs and requirements of the user. Nielsen divides system acceptability into practical acceptability and social acceptability. Practical acceptability has further been divided into different categories. Those categories are traditional like cost and reliability but there is also a category called usefulness. It means whether the system can be used to achieve goals that the user is desiring. Usefulness also can be divided into smaller categories such as utility and usability. According to Nielsen, a system's utility means that the functionality of the system is designed in a way that the user's desired goals can achieve with the system. Usability then can be seen by how well the user is able to use that functionality (See Figure 6) (Nielsen, 2010, pp.25).

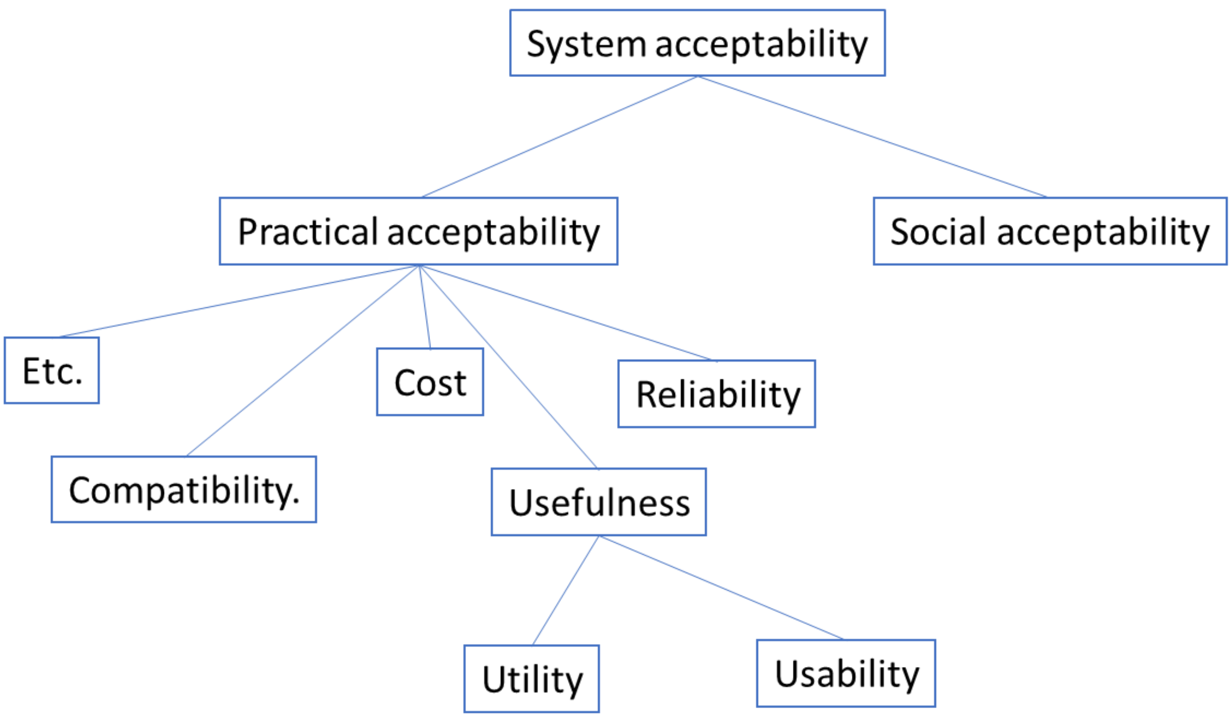

Figure 6: Model of the attributes of system acceptability adapted from Nielsen (2010, pp.25)

It has been proved that usability also influences more concrete things than user's technology acceptance. In 2008 Finnish municipal elections they tried electronic voting devices but because of bad usability over 200 votes were left out. The cause of this was the users had not noticed pressing the OK-button and the device did not notify users to press it (Enlund, 2018). Usability was also an issue in Florida, USA in the presidential election in 2000. They used a so -called butterfly-ballot where you must push the correct pin through the hole that lines the right candidate. This caused a problem when there were candidates on both sides of the page. If you wanted to vote for the candidate listed second on the left page you had to push the third hole (See Figure 7). There were thousands of voters that voted for the wrong candidate (Chisnell, 2016).

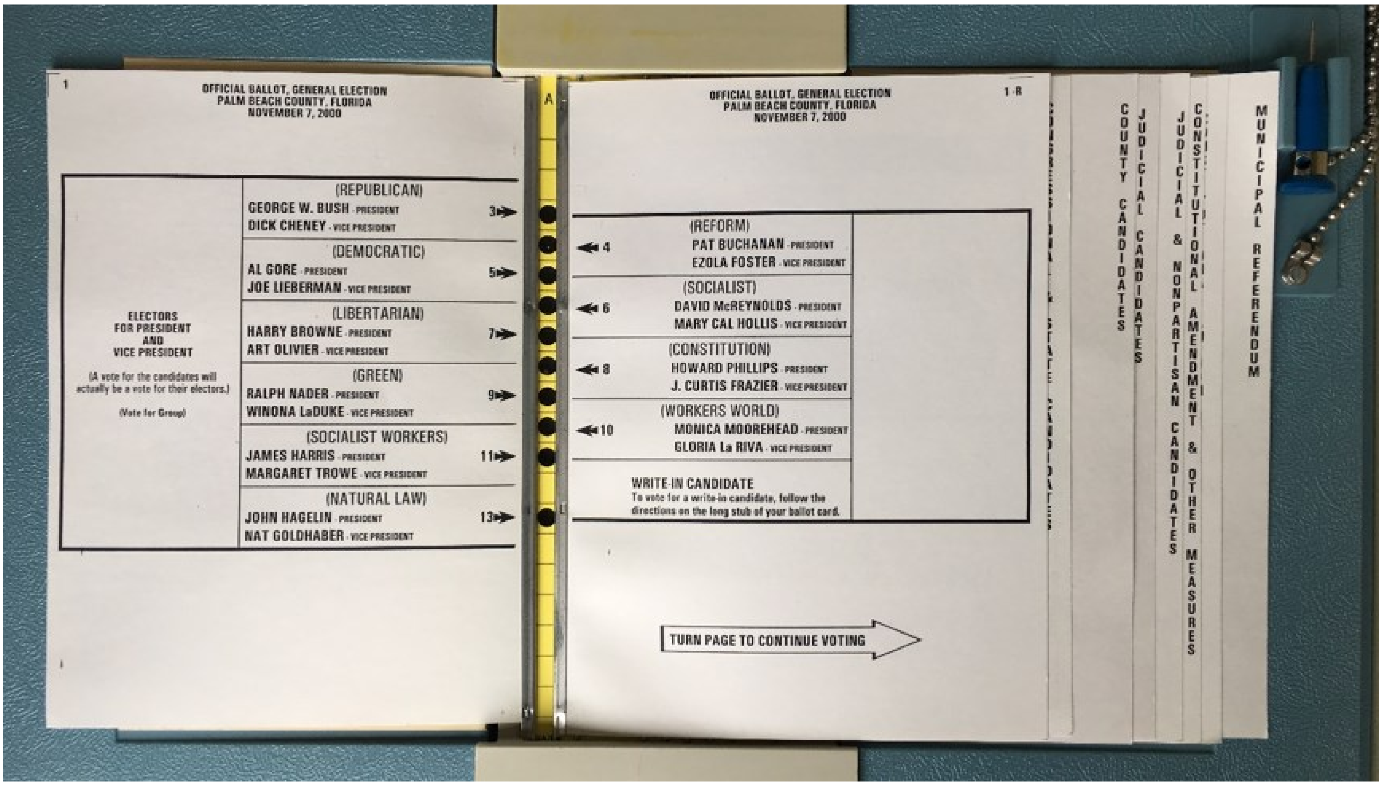

Figure 7: Butterfly-ballot in the presidential election in 2000

In healthcare, bad usability has also been an issue. For example, usability has been associated with patient safety issues, productivity loss and user frustration (Staggers et al., 2010). In Finland Kaipio et al. (2017) conduct a survey to physicians of usability of electronic health records (EHR). On a scale from 1 (fail) to 7 (excellent) used EHR systems got the average rating from 3.2 to 4.4. The results indicate that there are serious problems and usability issues in EHR systems that makes EHR's difficult to use and disturbs physicians' routine work. According to Kaipio et al. (2017) when developing more usable IT systems for healthcare the clinicians' point of view as a user should be considered and it should not be underestimated.

According to the theoretical framework of Davis (1993) and Nielsen (2010, pp.25) and findings of Kaipio et al. (2017), Chisnell 2016 and Staggers et al. (2010) the usability plays a vital role in ensuring that systems used are effective, efficient and by a wide range of users. The studies of Staggers et al. (2010) and Kaipio et al. (2017) also indicate that this is the case also in healthcare and without understanding usability and the relationship between usability and user experience and technology acceptance the whole potential of the benefits of IT in healthcare cannot be achieved. In fact, Jha et al. (2009) indicated that the adoption of IT in healthcare is lower than in other industries and one of the main causes may be poor usability.

4.3 Complexity on designing usable software

As we have noticed there are many ways to define usability and it means various things. In the end it is, despite all the definitions and heuristics that had been discussed earlier, a difficult task to design usable software according to Don Norman. Don Norman is a professor emeritus of Cognitive Science at the University of California and for example in the 1990s he was a User Experience Architect at Apple Computer. According to Norman (2010) technology and software have become more important in our daily life and we are more dependent on technology in our work. This means that software and other digital products and how we use them are affected by organizational structures, social and political issues. As a result of this, designers and developers have become more and more applied behavioral scientists. Designing usable software and products means understanding a human, including human cognition, sensory and motor systems, and a human as a part of different social networks and organizational structures (Norman, 2010).

Norman (2010) also points out that designers and developers should also have knowledge on scientific methods and statistics so they can perform valid and legitimate tests to their designed products. Without this knowledge of these fields the designing will be based on merely assumptions based on the designer’s own history and view of life and not on facts and science which may have a poor effect on usability (Norman, 2010).

According to another usability guru, Steve Krug, who wrote the usability state-of-art book ”Don’t make me think”, every software development team should use a usability expert to build usability into their products. The problem with that is the vast majority cannot afford hiring a dedicated usability professional. And the problem in question has become larger because nowadays many visual designers and developers do interaction design and information architecture design which have a major effect on usability (Krug, 2014, pp. 5).

5 Usability Questionnaires

Questionnaires are a usable way to measure usability of a software and to find out what features users like or dislike. In questionnaires the questions are presented to the respondent and the respondent fills the correct answer. Questionnaires are indirect ways to measure usability because they do not measure the software itself but only users' opinion about the software. Still questionnaires are direct methods when they measure user satisfaction (Nielsen, 2010, pp. 209 - 210).

Rating scales of questionnaires are used especially when asked users how well they liked certain aspects or how useful they find various components. There are for example different semantic differential and Likert scale questions that can be used in questionnaires, but it is good to keep in mind that only a few different types of questions should be added in the same questionnaire (Nielsen, 2010, pp. 213). According to Nielsen there needs to be at least 30 respondents in the usability questionnaire to get reliable results, and to prevent misunderstandings there should be done pilot work before starting the actual research with questionnaire (Nielsen, 2010, pp. 224). In this chapter there are introduced usability questionnaires that have been used to study usability and interaction between human and computer in healthcare.

5.1 Telehealth Usability Questionnaire (TUQ)

Parmanto et al. (2016) realized that current telehealth usability questionnaires were designed mostly for older technologies such as video conferencing systems. Therefore, there was a need for a questionnaire that addresses the changes in telehealth service delivery and technology.

Telehealth Usability Questionnaire contains a total of 21 different questions, and it uses 6 different components for measuring the usability of telehealth (Figure 8). TUQ was designed combining three different usability questionnaires. The usability factors of telehealth come from Telemedicine Satisfaction Questionnaire (Yip et al. 2003). The questions of usefulness and ease of use were derived from the Technology Acceptance Model (Davis 1993) and the questions of ease of use, interface quality, reliability, and satisfaction were derived from the IBM Post-Study System Usability Questionnaire (Lewis 1995). Parmanto et al (2016) indicates that Telehealth Usability Questionnaire is a solid, robust, and versatile measure. It is useful to measure the quality of the computer-based user interface and the quality of telehealth interaction and services.

Components	Questionnaire Items
Usefulness	
1	Telehealth improves my access to healthcare services
2	Telehealth saves me time traveling to a hospital or specialist clinic
3	Telehealth provides for my healthcare needs
Ease of Use and Learnability	
1	It was simple to use this system
2	It was easy to learn to use the system
3	I believe I could become productive quickly using this system
Interface Quality	
1	The way I interact with this system is pleasant
2	I like using the system
3	The system is simple and easy to understand
4	This system is able to do everything I would want it to be able to do
Interaction Quality	
1	I could easily talk to the clinician using the telehealth system
2	I could hear the clinician clearly using the telehealth system
3	I felt I was able to express myself effectively
4	Using the telehealth system, I could see the clinician as well as if we met in person
Reliability	
1	I think the visits provided over the telehealth system are the same as in-person visits
2	Whenever I made a mistake using the system, I could recover easily and quickly
3	The system gave error messages that clearly told me how to fix problems
Satisfaction and Future Use	
1	I feel comfortable communicating with the clinician using the telehealth system
2	Telehealth is an acceptable way to receive healthcare services
3	I would use telehealth services again
4	Overall, I am satisfied with this telehealth system

Figure 8: Telehealth Usability Questionnaire

Hajesmaeel-Gohari & Bahaadinbejy (2021) compared different questionnaires used in telemedicine service studies. They analyzed different telehealth studies and found 53 articles that met their criteria. For example, they ruled out articles that did not have full-text available or the questionnaires used lacked verified validity and reliability. In those 53 articles Telehealth Usability Questionnaire was the most common tool and it was used in 10 articles. The second most common tool was Telemedicine Satisfaction Questionnaire (Yip et al., 2003) and it was used in 7 articles. Hajesmaeel-Gohari & Bahaadinbejy (2021) notes that TUQ is usable for evaluating different types of telehealth systems from videoconferencing systems to computer and mobile based systems. TUQ can also be used to collect opinions of both patients and physicians (Hajesmaeel-Gohari & Bahaadinbejy, 2021).

5.2 Nasa Task Load Index Test (NASA-TLX)

Nasa Task Load Index Test is a questionnaire that measures workload while performing a task or immediately afterwards. NASA-TLX uses 6 different components to measure workload: mental demand, physical demand, temporal demands, frustration, effort and performance. Subject rates each component from low to high and combined result of these components gives the mental workload of the measured task (See Table 2). At first NASA-TLX was designed to measure workload in aviation, but it has been turned out to be useful also in other fields when studying interaction between humans and computers, for example in healthcare (Agnisarman et al., 2017: Hart, 2006; Colligan et al., 2015). Agnisarman et al. (2017) conducted a study which used NASA-TLX when they researched the workload of using different video-based telemedicine software. The research of Agnisarman is explained more deeply in chapter 6.

Table 2: NASA Task Load Index-questionnaire adapted from Hart (2006)

TITLE	ENDPOINTS	DESCRIPTIONS
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the task or tasks elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the tasks set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

6 Usability of video-based telemedicine visits

Agnisarman et al. (2017) did a study where they evaluated the usability of four home-based telemedicine software: Doxy.me, Vidyo, VSee and Polycom. They used a within-subjects experimental design where every participant was asked to complete several tasks in each platform. Tasks were such as downloading software if needed, accessing the meeting, and communicating via video. After completing the tasks for each platform participants answered the IBM Computer System Usability Questionnaire (CSUQ) and the Nasa Task Load Index Test. 19 participants completed the study. According to the results of CSUQ the participants were fairly satisfied with the usability of tested platforms. In Likert-scale from 1 to 5 all platforms received score higher than 3. The lowest score was 3.28 and the highest score was 4.15. There were some statistically significant differences between platforms. The differences were found in areas such as task completion time, mental demand, frustration and computer system usability scores. There were some usability issues, mainly during the initiation phase when installing software and creating an account. These usability issues led to high mental demand and task completion time. Agnisarman et al. (2017) noted that the findings of this study would be useful for software developers to develop user-friendly telemedicine systems.

Patel et al. (2021) conducted a survey to patients who received medical care via video to understand the technical challenges faced by patients. The survey featured questions about technical challenges, patient satisfaction and motivation and there were 180 patients that responded the survey. The software used in visits coupled with Vidyo and patients participated in meetings mainly via a smartphone or tablet. The patients were also given some instructions on how to complete preparations before the appointment. Patients also had a possibility to test the connection via mobile app in advance. Patel et al. (2021) found out that the overall patient satisfaction was high and 90% of respondents agreed that the experience of video meeting was like face-to-face meeting. Even though the overall satisfaction was high 36 respondents had faced technical issues during the video meeting. These issues were mostly because of audio, video or connection. The research of Patel et al. (2021) shows that even if a customer faces technical issues, it does not immediately affect the overall satisfaction of video meetings in healthcare. Still a meaningful number of respondents had technical issues in video meetings which should be taken into notice by healthcare providers, and this should be further researched.

Layfield et al. (2020) researched the usability and patient satisfaction of head and neck ambulatory telemedicine patients. This study used Telehealth Usability Questionnaire (TUQ) and there were 100 respondents. In this research patients used a commercial software in meetings and 78% of patients used smartphones. Other devices used were tablet, laptop and desktop. Before the meeting the patient had received instructions for accessing the video meeting. The patients were mostly pleased with telemedicine visits and the total average usability score was 6.01 in Likert-scale from 1 to 7. The highest score was given when asking the satisfaction of the telemedicine visits, 6.29. The respondents noted that telemedicine visits increased the access to health care services and saved time. Interesting part of the study of Layfield's et al. (2020) was that the lowest average score was when asked if the system was able to do everything the respondent wanted it to be able to do, 5.27. Layfield et al. (2020) also addressed those 23 persons dropped out from this research and telemedicine visits because of socioeconomic circumstances or low technology literature and they were unable to have video-based telemedicine visits.

Johnson et al. (2020) studied the usability of video-based telemedicine visits of breast cancer patients. The method used was Telehealth usability questionnaire (TUQ) with Likert scale 1-7 and there were 75 patients that completed the survey. The respondents answered the survey within 2 days after the video meeting. 26 of 75 respondents this was their first video-based telemedicine visit and 31 respondents have had 2 or more telemedicine visits before this one. The median age of respondents was 63 ranging between 25-83 years. Patients were generally satisfied with the video-based telemedicine visits, and they find it usable, the overall usability score being 5.5. Even though there were differences between respondents age and number of prior telemedicine visits Johnson et al. (2020) found that the usability scores were not associated with respondents age or prior telemedicine visits. Some patients had experienced some connection troubles and those patients were associated with lower usability scores than those who did not have connection issues.

In the research of Thelen-Perry's et al. (2018) 20 urology patients were interviewed about their experience about video-based visits. The telemedicine visits were return visits of the patient's initial in-clinic visits, and they were conducted by a single urologist. Patients were provided in advance step-by-step instructions on how to download a smartphone application and access to video meetings. Patients also had a possibility to get assistance with downloading the application by an officer. Thelen-Perry et al. (2018) conducted a semi-structured telephone interview with patients after the meeting and they found out that the patients were pleased with the video visit experience. Most patients reported that they only had minimal issues when joining the video meeting, but some had problems with downloading the video application. Patients did not have problems with the quality of video but there were some issues with audio. According to Thelen-Perry et al. (2018) the key factors for successful adoption of video-based telemedicine would be obtaining direct feedback from patients and identifying technical issues.

Despite several studies in the world agreeing that patients feel generally satisfied with telemedicine visits via video and that the usability of video-based telemedicine software is good, there is still a lack of Finnish research in that area. For example, there was not any Finnish-based research to be found when searching usability research of video-based telemedicine visits from patients' point of view from PubMed-or ProQuest Central-databases.

7 Research method

The method of this test was to act a made-up video-based telemedicine visit for control visit for high blood pressure where the subject would perform as a patient. So, the subject was suffering an imagined high blood pressure. Every participant conducted the same telemedicine visit with three different video-based telemedicine software and after using each software the questionnaires were answered. The order of software used for all participants was based on Latin square to ensure each software would occur evenly in the test order (See Table 3).

Table 3: Testing order of software for different groups

Group 1	Group 2	Group 3
Software 1	Software 2	Software 3
Software 3	Software 1	Software 2
Software 2	Software 3	Software 1

The test started with both researchers greeting and welcoming the subject. After that the researcher who would be the doctor in the test left for another room and the other researcher stayed with the subject (See Figure 9). Then the researcher conducted proper guidance to the subject about what the test will be, ending the subject granting a permission to participate in the research. After the granted permission the actual part of the test started.

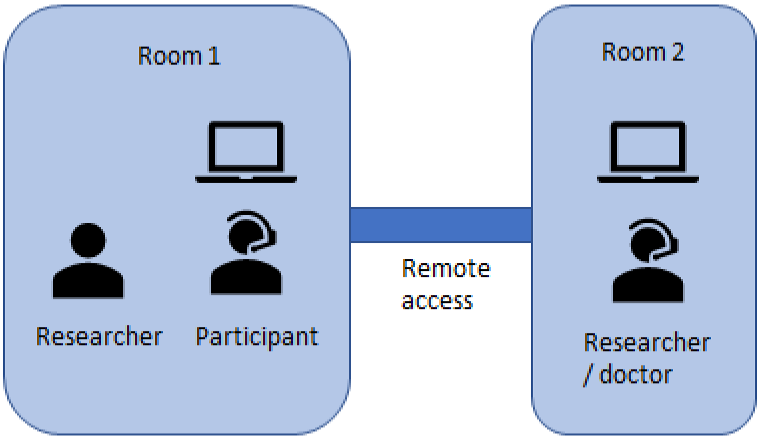

Figure 9: Research situation

As a start of the telemedicine visit the doctor, who was the other researcher, sent to the subject an email-invitation, which contained instructions how to join the telemedicine meeting. Both researchers had a made-up manuscript of how the visit with the doctor would advance and the subject was given a new task when the latter one had been successfully completed (See Appendices 4-6). The tasks were such as joining the meeting, having a made-up conversation with the doctor telling blood pressure results, muting and unmuting the microphone, setting a meeting to full-screen and leaving the meeting. Telling the blood pressure results happened conversing with software 1 and 3 but with software 2 it happened via chat-window because the software had that feature. The researcher, who was in the same room as the subject, observed the test situation and if the subject could not successfully complete the task the researcher helped the subject with the task. When all tasks were completed, the subject then conducted a usability questionnaire that included Telehealth Usability Questionnaire (TUQ) (See Appendix 1) and Nasa Task Load Index (NASA-TLX) (See Appendix 2). After filling the questionnaire then the subject was guided to test other 2 software, a new email-invitation was sent and test procedure with tasks started again (See Figure 10). After all three software were tested, a small interview section was held by researchers where the subject reviewed software and possible errors that had occurred in the test. In the end the subjects also ranked the software, which they would prefer the most in video-based telemedicine meeting and which they prefer the least (See Appendix 3).

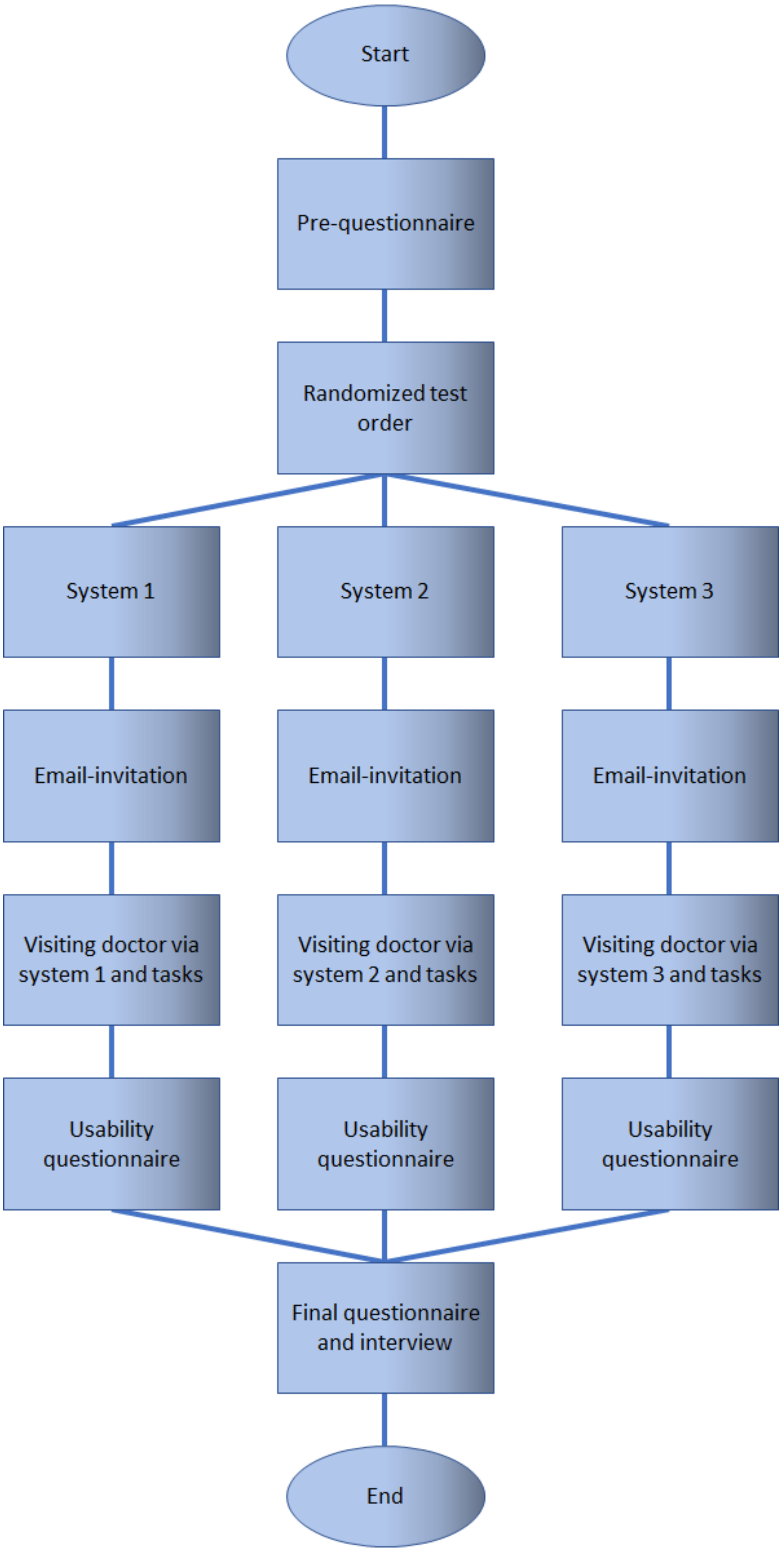

Figure 10: Research method

The research method of this study followed mainly the same procedure as in the usability research of Agnisarman's et al. (2017) on four different video-based telemedicine software. Some changes were still made. In the study of Agnisarman's et al. (2017) questionnaires used were Computer

System Usability Questionnaire (CSUQ) and NASA-TLX but in our research we changed CSUQ to TUQ because we wanted to use a more telemedicine-based usability questionnaire. The tasks in our research were mainly the same with only minor changes mainly because of the differences in the software of our research and Agnisarman's et al (2017) research.

Telehealth Usability Questionnaire was modified for this research because TUQ included questions about usefulness and reliability and to answer those questions reliably the subject needs to have experience and history of using telemedicine services. Therefore, TUQ in this study included only four components, which were ease of use and learnability, interface quality, interaction quality and satisfaction and future use. NASA Task Load Index Test was also slightly adjusted to meet the purposes of this study. Original NASA-TLX uses a 21-scale questionnaire, but we only use a scale from 1 to 5 so the questionnaire which contains TUQ and NASA-TLX will look in-line. The questionnaire was digital, and it was created in Google Forms.

For email-invitations for doctor's appointments a Gmail-account was created. The participants used that Gmail-account where the email-invitations were sent so the participants did not use their own email accounts. As told before the email-invitations contained instructions how to access the meeting via a tested system. The instruction basically included the right URL-address and notification to follow the instruction provided by the software. We wanted that there was available as little information as possible because according to Nielsen's usability heuristic #10 Help and documentation in optimal circumstances of usability there should not be any need of additional help or documentation (Nielsen, 2020)

Before the actual test with real subjects the test procedure was tested with 5 test subjects and the procedure was found out working as it should be and the research with real subjects could start. Though some adjustments with the test situation and procedure were made because of issues found with test subjects.

7.1 Software in research

The software introduced in this chapter were used either in the DigiSote-project or Ossi-project and that is the reason why those were tested in this research. The test was conducted using Windows 10-operating system and Google Chrome-browser because all software tested worked without any issues on that. For example, two software did not function at all with Internet Explorer- or Safari-browsers with 11-version or older because software used WebRTC for transmitting video signals and those browsers did not support WebRTC. The research was done in 2018 so there may have been changes in tested software after that, so the result of this study does not represent the status of these software at the date of the publication of the thesis.

7.1.1 Software 1

In Software 1 an organization has its own workplace where a patient can see all the available professionals that are logged in. This means that the patient needs to first access the workplace through the right URL address. Then the patient has to find the right professional from the contact list and then call the professional (See Figure 11).

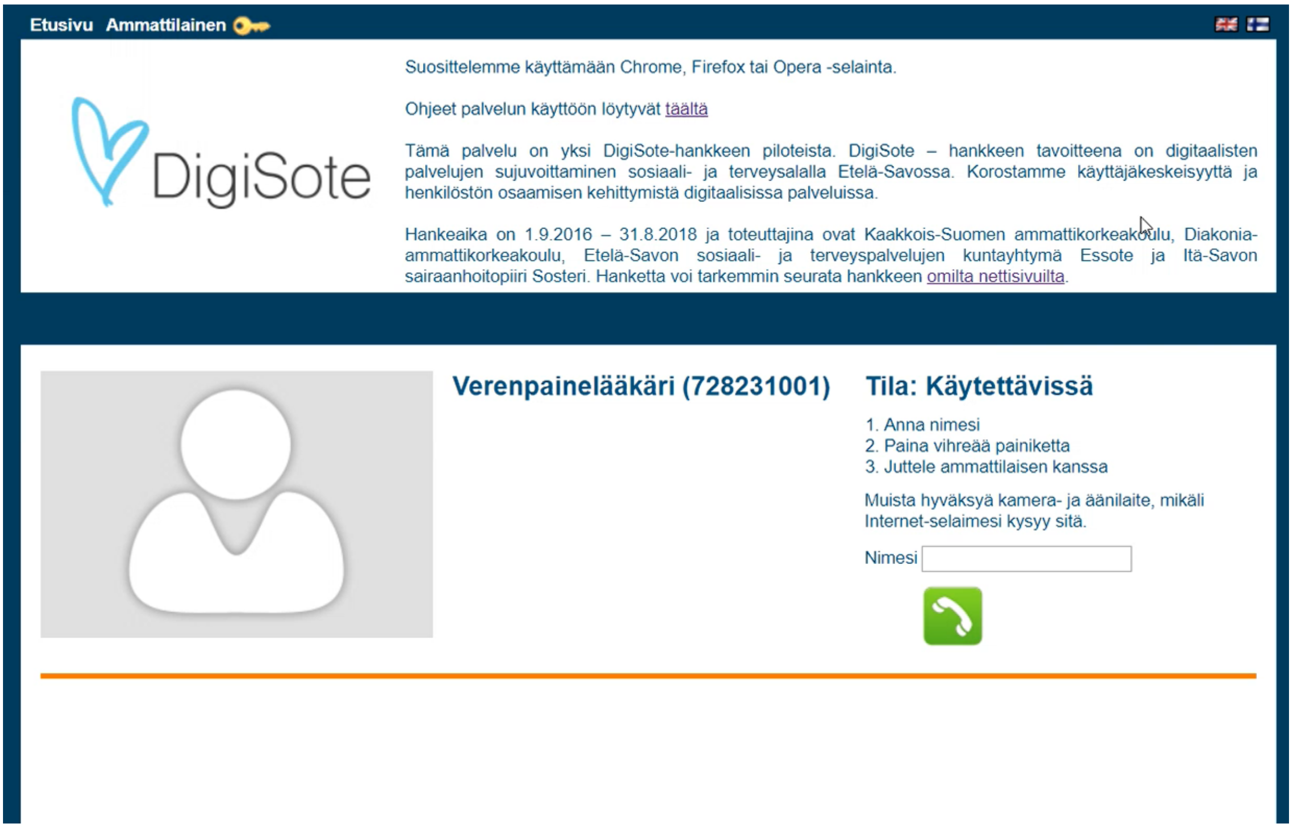

Figure 11: Accessing the video meeting with software 1

In a meeting there are buttons to mute microphone and audio and a red telephone for leaving the meeting. The doctor is in the center of the screen and the patient's own video is on the top right corner of the screen. Under the doctor's view there are two drop-down buttons for changing the source of the video and audio (See Figure 12).

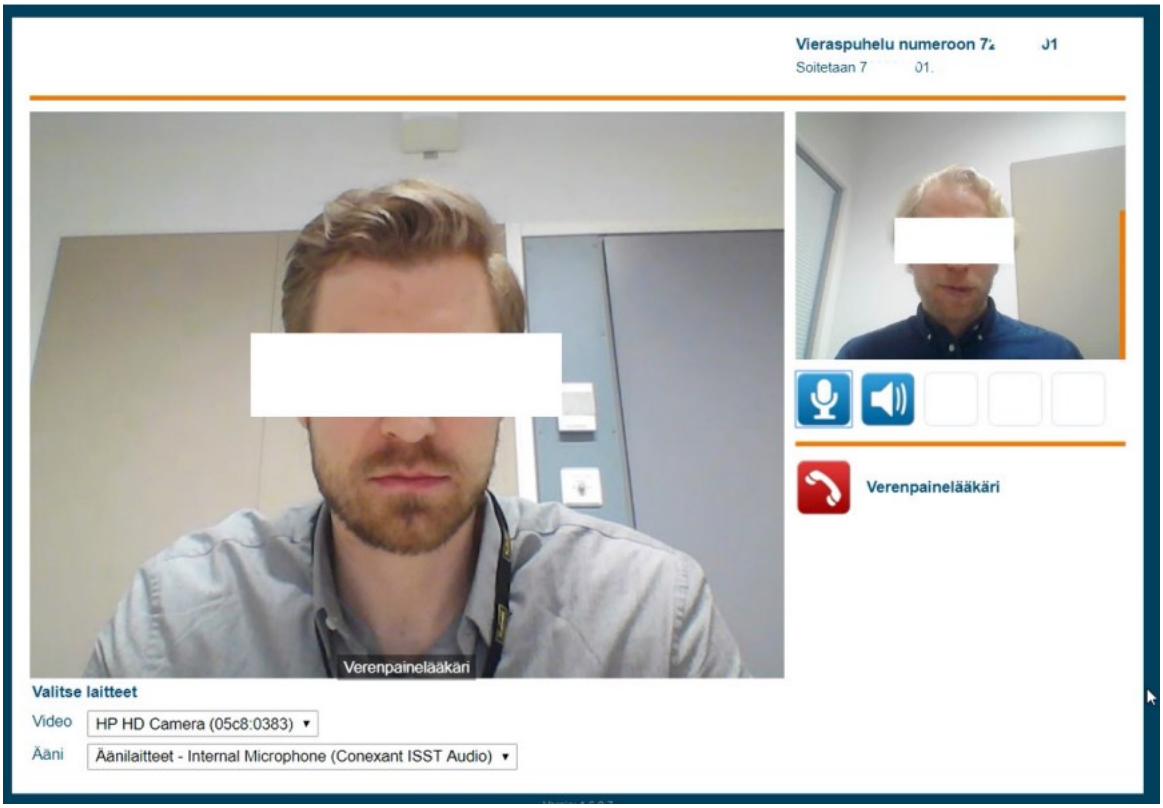

Figure 12: Meeting view of patient with software 1

7.1.2 Software 2

Software 2 is based on different meeting rooms where participants could access the meeting. The right URL-address of the correct meeting room should be provided in the forehand and access happens through that address by typing your own name and copy-pasting the URL address in the login-page (See Figure 13). In the Connect-button there is a drop-down menu for adjusting video and audio settings.

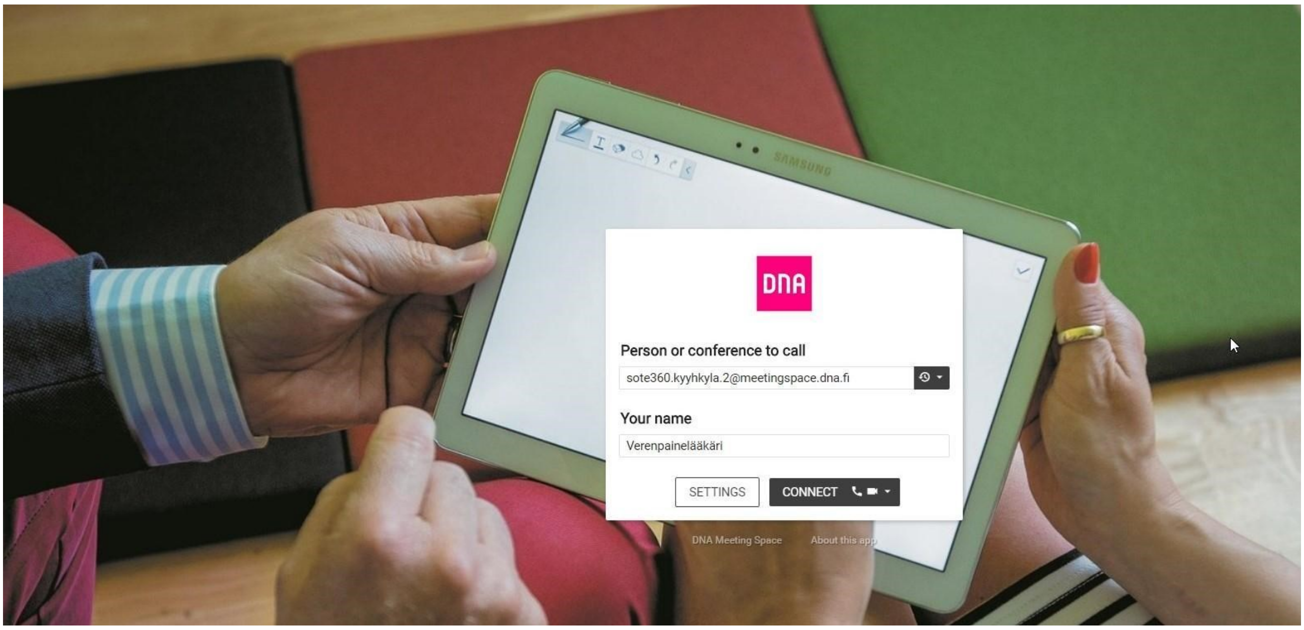

Figure 13: Login with software 2

When in a meeting with software 2 the patient can see the doctor in the middle of the screen and the patient is in the right top corner. Below the screen there are multiple buttons for example muting microphone or sharing an own screen. There is also a possibility to change a doctor's video in the full screen mode which can be done with a button that is in the down-right corner of the screen. In this software there is a chat-functionality and we used that in our research to type the correct blood pressure result which is on the left side of the screen (See Figure 14).

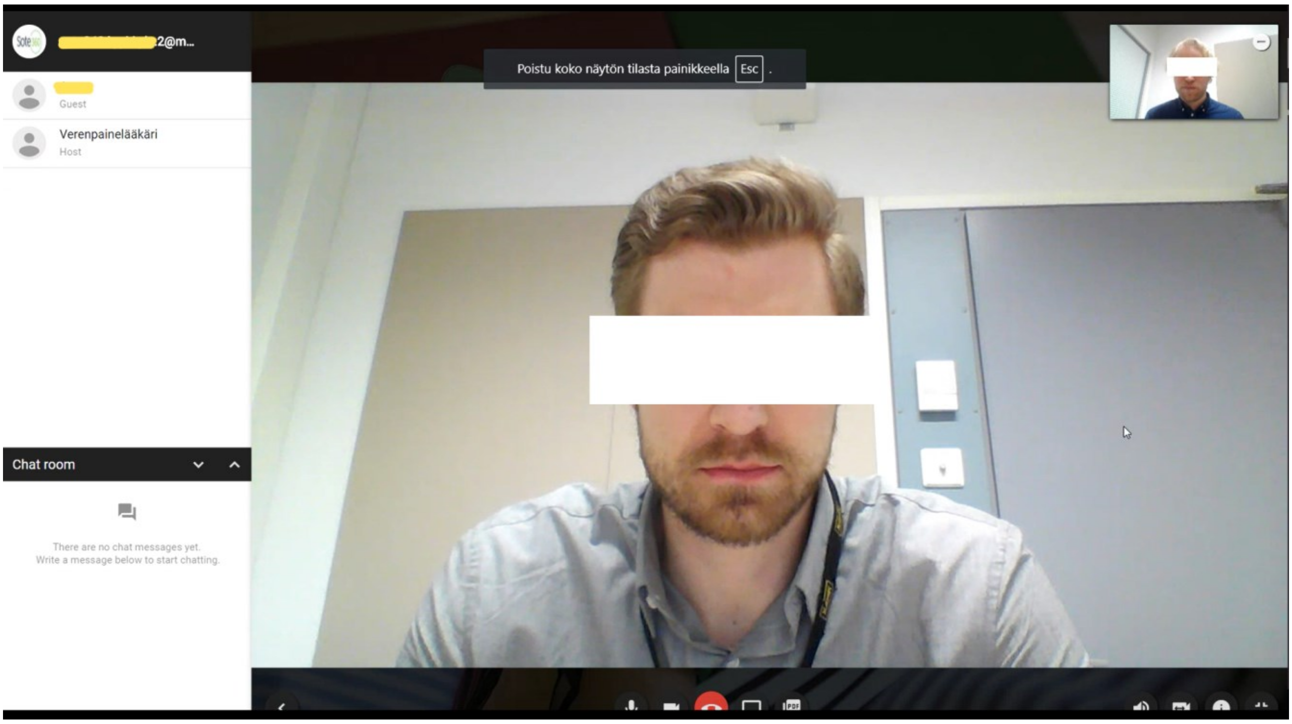

Figure 14: Meeting view of patient with software 2

7.1.3 Software 3

Accessing a meeting in Software 3 happens through an invitation message that is automatically sent by the software. Message contains a link that redirects the patient to the meeting. This software uses Vidyo Web-plugin, and the plugin needs to be downloaded when using this software for the first time. When using software 3 for the first time it redirects you automatically to download the Vidyo web-plugin (See Figure 15 & Figure 16).

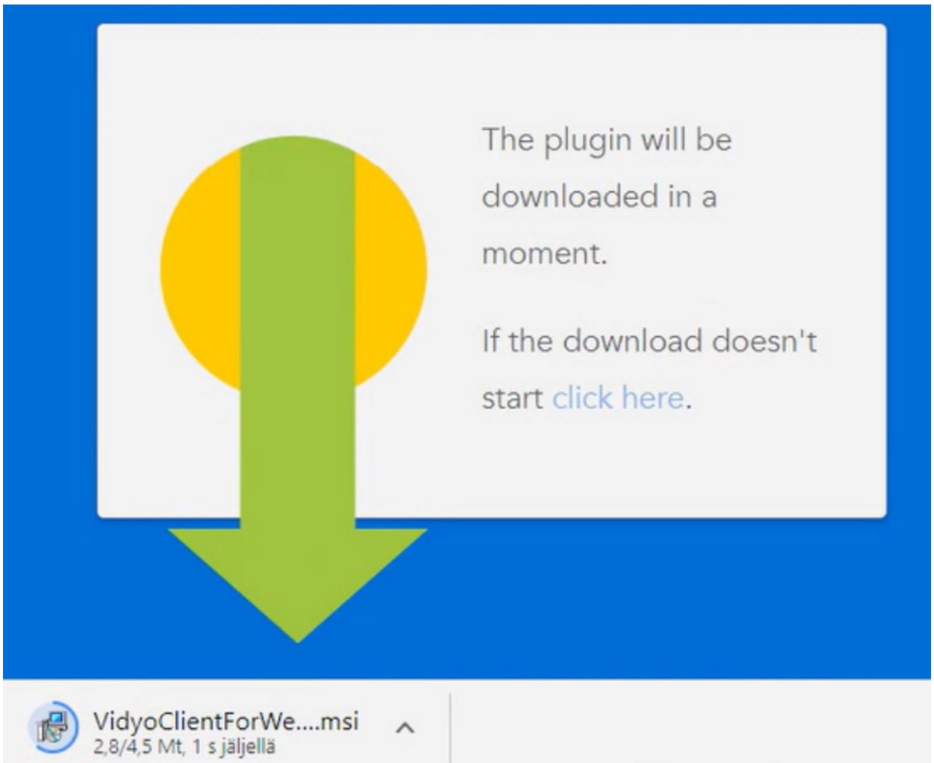

Figure 15: Downloading Vidyo web-plugin part 1

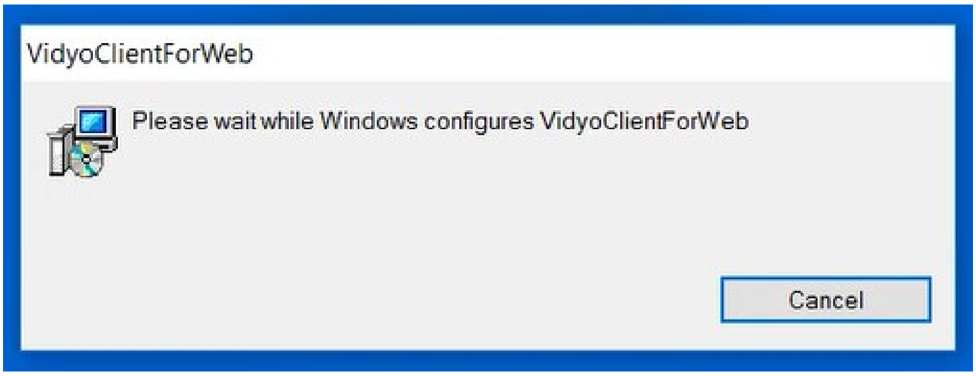

Figure 16: Downloading Vidyo web-plugin part 2

In a meeting the patient is in the center of the screen and the patient's own video is in the down-right corner. Below the patient's video there are buttons to change the doctor's video to full screen-mode and to leave the meeting. Far left from those buttons there are other buttons for muting microphone and audio and stopping your own video. On the left side of the screen the patient can see the participants of the meeting and is also able to share documents etc. (See Figure 17).

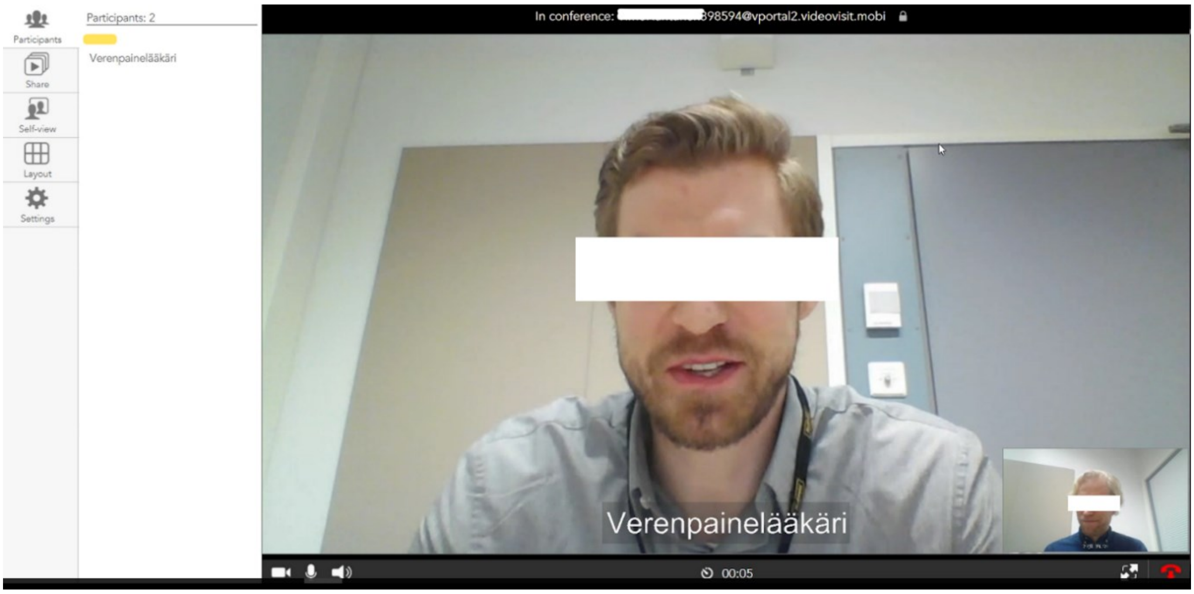

Figure 17: Meeting view of patient with software 3

7.2 Participants

The subjects in the study (n = 21) were teachers or students at South-Eastern Finland University of Applied Sciences (XAMK). There were 12 students and 9 teachers. The mean age of student's was 33 years and teacher's 51 years. The mean age of all participants was 41 years. The participants were all Finnish, so the questionnaires and manuscripts were all in Finnish. The participants recruited via email-advertisements. During the recruitment the participants conducted a pre-test survey where participants evaluated their experience and know-how on telemedicine software, Windows-operating system and Google Chrome-browser. To minimize the effect of learnability of tasks of the test and telemedicine software the participants were divided into three groups by Latin square, so the test order of software was randomized. Factors of the pre-test survey were considered in the division, so the groups matched each other by know-how on telemedicine software, Windows-operating system and Google Chrome-browser. As a prize of participation, the participants were given two movie tickets.

7.3 Statistical analysis

The averages and standard deviations were calculated in Microsoft Office Excel 2013-software ((Microsoft Corporation, Redmond, Washington, USA). Figures and tables of the results were created also in Excel to visualize the results. Though there were three different groups for participants the results were processed as a one group because of the Latin square that diminished the risk of learnability by ensuring that all software occurred evenly in the testing order.

7.4 Ethical analysis

This study was approved by the ethical committee of South-Eastern Finland University of Applied Sciences. The participants were recruited via email that were sent to the students and teachers at South-Eastern Finland University of Applied Sciences. The email contained information about the research method so that the participants would have a clear understanding about the research. Before the test the participants were given instructions and information about the research and the research method. After that the participant was asked to give a confirmation to the research. During the testing procedure the participant was able to stop the test whether the participants felt like it.

The information gathered from the participants were all in digital form. The information was basically the results from the questionnaires and the final interview which included ranking tested software from the most preferred to the least preferred and commenting on the used software. The questionnaires were conducted in Google Forms and the participants were given a unique ID which could not be linked to a certain person. The participants used that ID when responding to the questionnaires and the researcher used the same ID when writing down to Google Sheets the main findings of the final interview. When all the participants had participated in the study all the files from the Google Cloud were downloaded locally and all the files from the cloud were deleted. Also, the back-up files were created and those were in the network drive of South-Eastern Finland University of Applied Sciences.

8 Results

8.1 Telehealth Usability Questionnaire

Usability of three video-based telemedicine software were evaluated by modified Telehealth Usability Questionnaire with Likert-scale from 1 to 5. Total average scores can be seen in Figure 18. Software 1 received the highest average score 4.33 ± 0.80 while software 3 received the second highest score 4.13 ± 0.84. The lowest average score was for software 2 with an average of 2 3.39 ± 1.16.

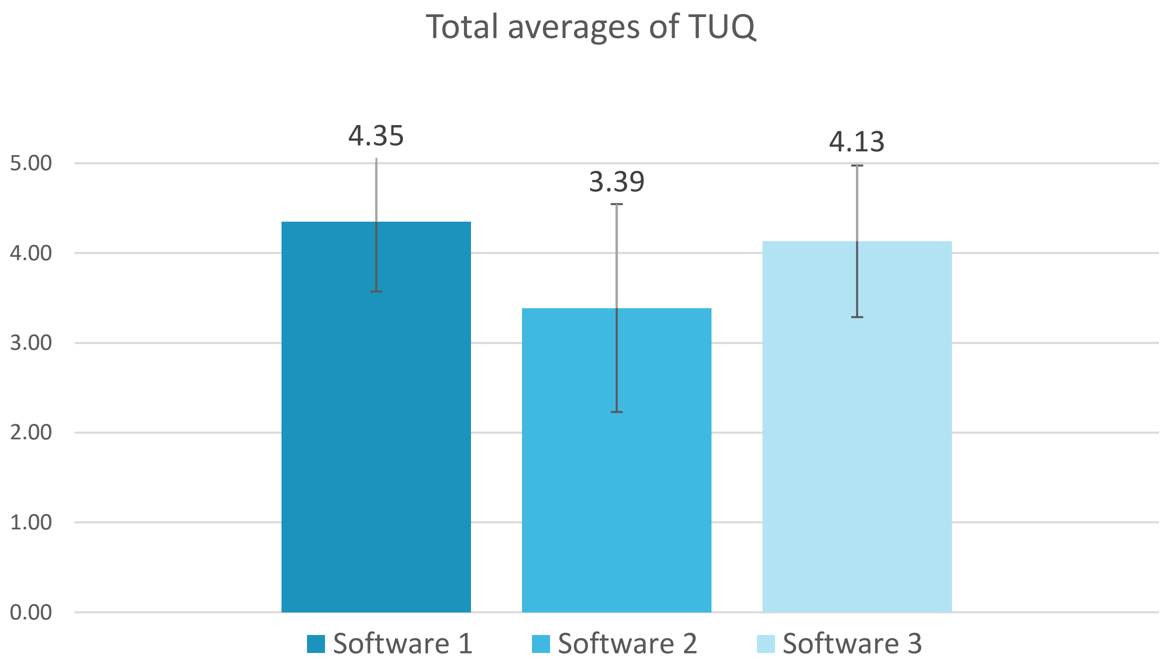

Figure 18: Total average results and deviations of Telehealth Usability Questionnaire

TUQ includes four different components, which are ease of use and learnability, interface quality, interaction quality and satisfaction and future use. The average scores of these components are seen in Figures 19, 23, 28 and 33. In this chapter there are also result-charts of each question following the component which they are included.

Software 1 received the highest averages in all four components. Ease of use and learnability was evaluated as average 4.49 ± 0.71. interface quality was 4.31 ± 0.78. interaction quality was 4.35 ±

0.88 and satisfaction and future use was evaluated 4.29 ± 0.91. The second averages received software 3 and its averages were in ease of use and learnability 4.16 ± 0.77, interface quality was 4.12 ± 0.86, interaction quality was 4.20 ± 0.94 and satisfaction and future use was 4.04 ± 0.79. The lowest average was evaluated to be in software 3 with ease of use and learnability' average and standard deviation being 3.04 ± 110, interface quality's 3.14 ± 1.09, interaction quality's 4.06 ± 0.95 and satisfaction and future use's 3.21 ± 1.20. The results of TUQ are further discussed and analyzed in the chapter 9.

8.1.1 Ease of use and learnability

There were three questions in the Ease of use and learnability-section and the questions varied from simplicity and learnability to producibility. The results of those questions can be viewed from the Figures 20 - 22. The total average scores of ease of use and learnability-component is seen in figure 19.

In the questions of this component software 1 received average scores 4.57 ± 0.81, 4.67 ± 0.58 and 4.24 ± 0.70. The lowest averaged received software 2 with the averages of 2.71 ± 1.06, 3.14 ± 1.20 and 3.29 ± 1.00 and software 3 got averages scores of 4.19 ± 0.81, 4.29 ± 0.72, 4.00 ± 0.77.

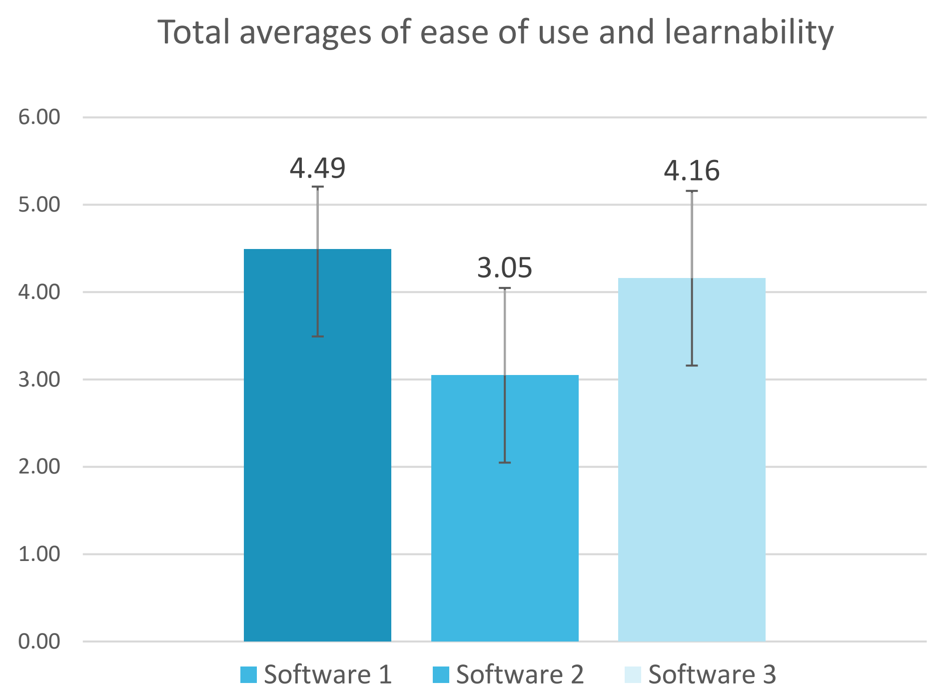

Figure 19: Average results and deviations of ease of use and learnability

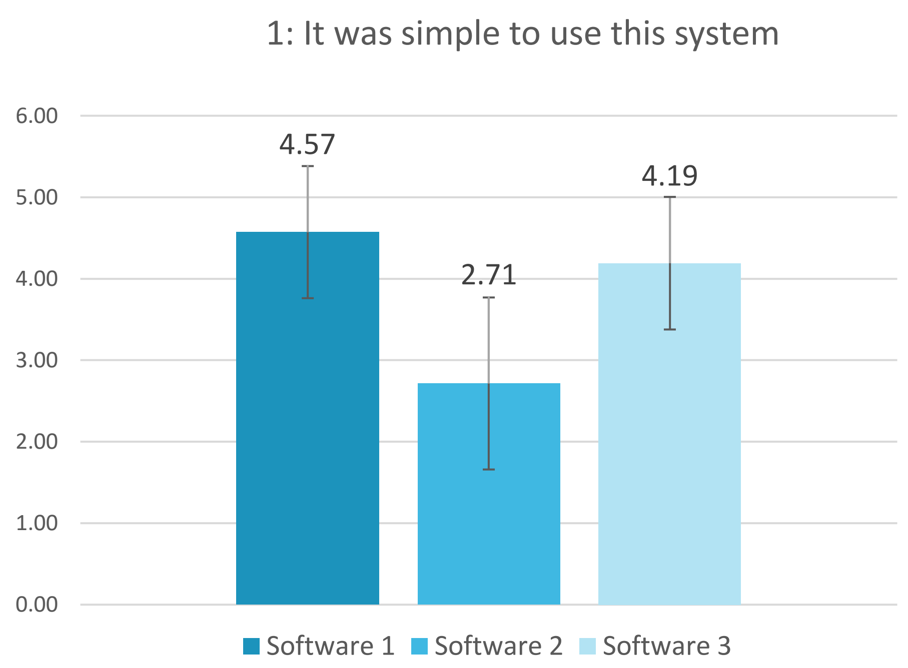

Figure 20: Average results and deviations of question 1 of TUQ

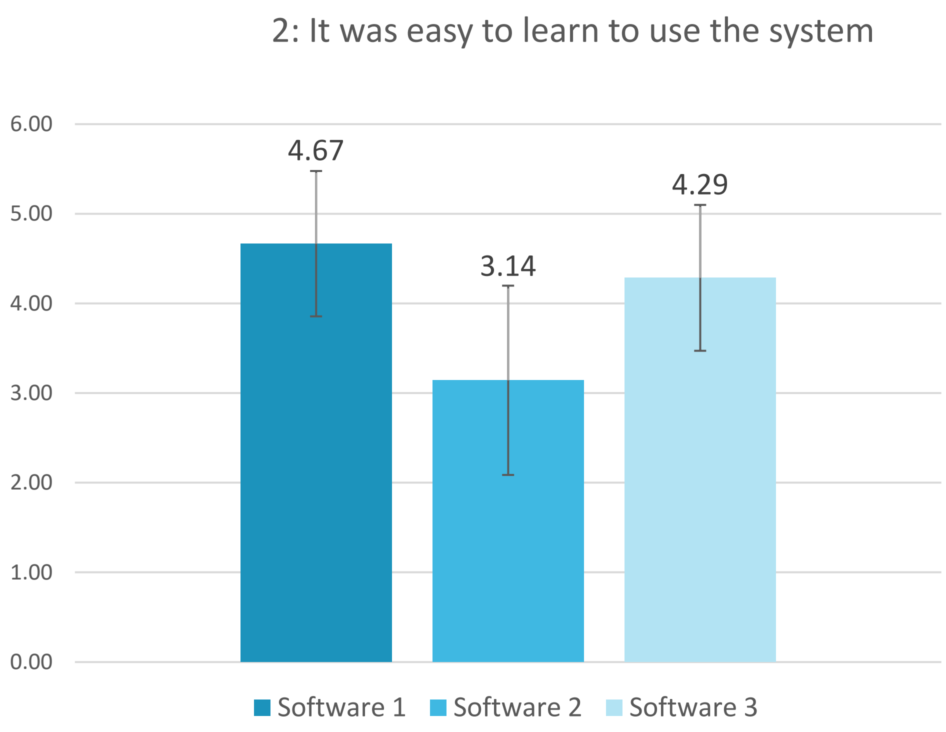

Figure 21: Average results and deviations of question 2 of TUQ

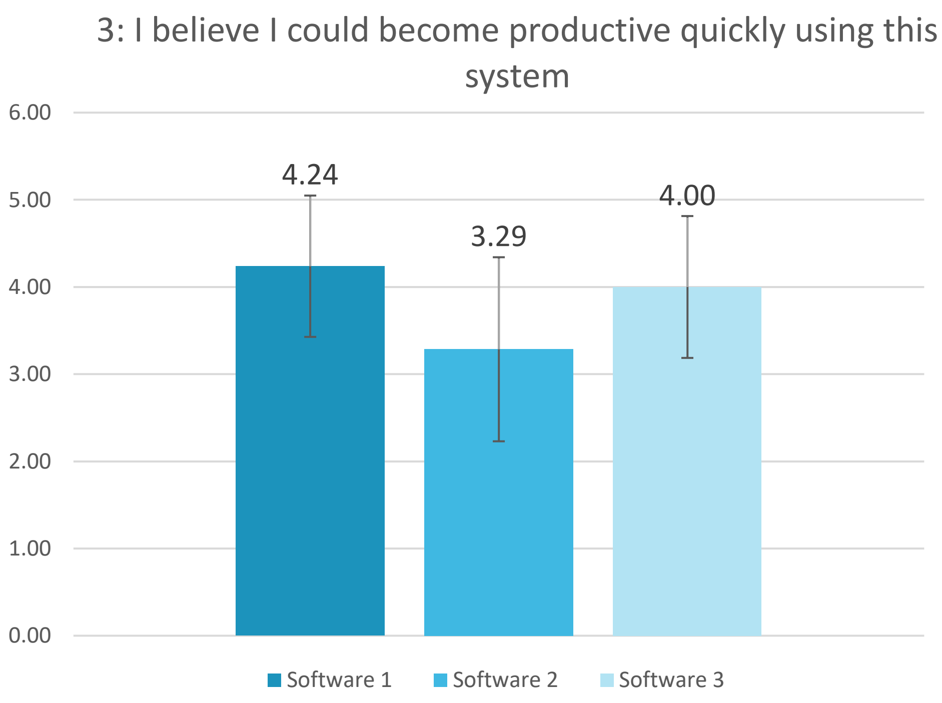

Figure 22: Average results and deviations of question 3 of TUQ

8.1.2 Interface quality

In the Interface quality-section there were four questions which asked respondents to rank the interaction with the software, did they like the software, was the software simple to understand and if the software was able to do everything they wanted. The results of those questions can be viewed from the Figures 24 - 27. The total average scores of interface quality-component is seen in figure 23.

In the questions of this component software 1 received average scores 4.29 ± 0.85, 4.43 ± 0.68, 4.38 ± 0.92 and 4.14 ± 0.65. Software 2 received the averages of 3.19 ± 1.21, 3.05 ± 1.02, 2.76 ± 1.14 and 3.57 ± 0.87 and software 3 got averages scores of 4.10 ± 0.77, 4.10 ± 0.94, 4.19 ± 0.98 and 4.10 ± 0.77.

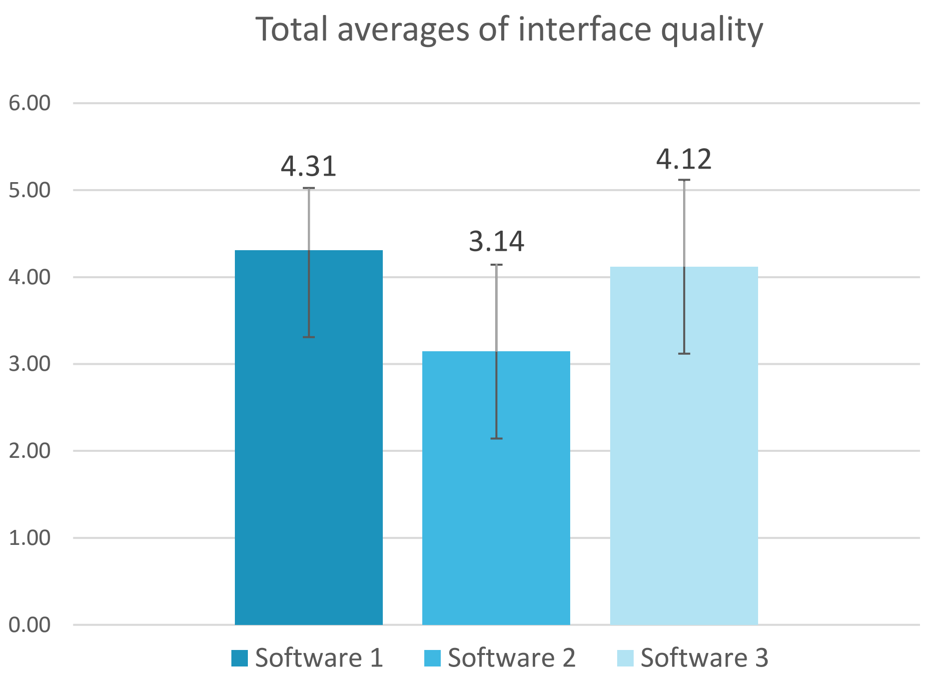

Figure 23: Average results and deviations of interface quality

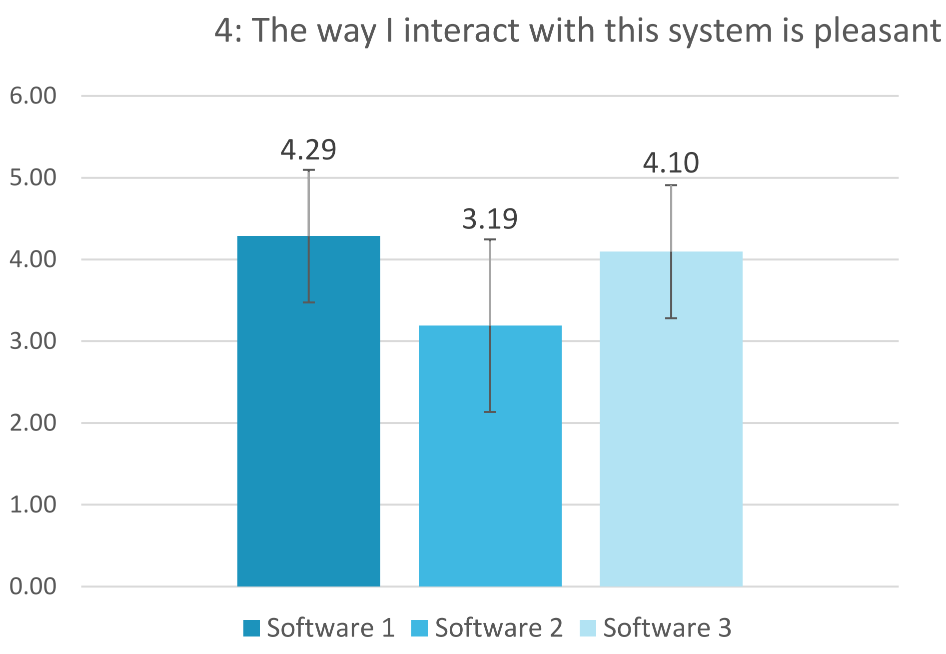

Figure 24: Average results and deviations of question 4 of TUQ

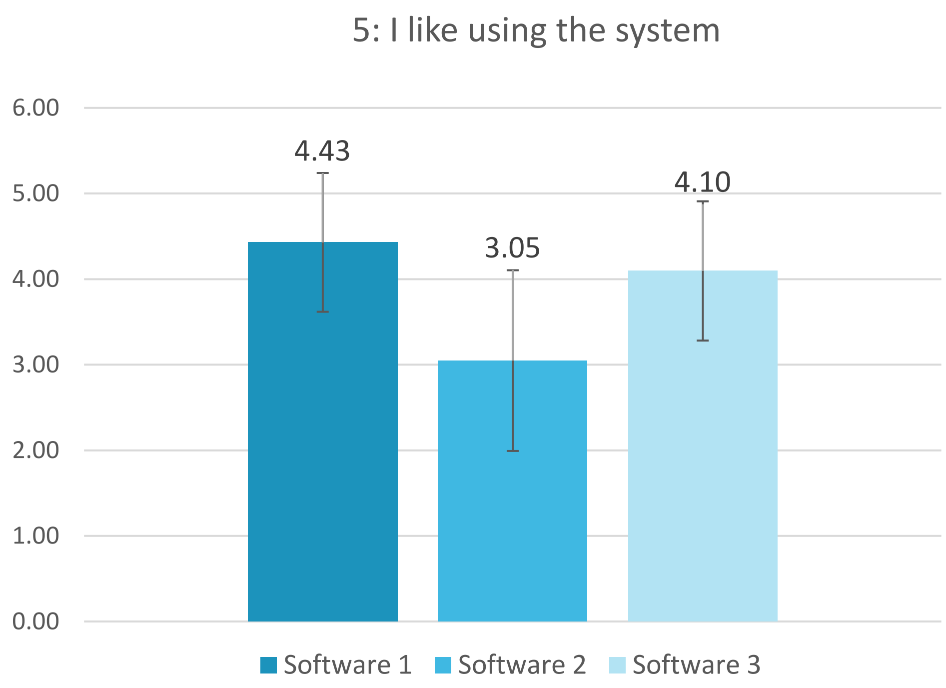

Figure 25: Average results and deviations of question 5 of TUQ

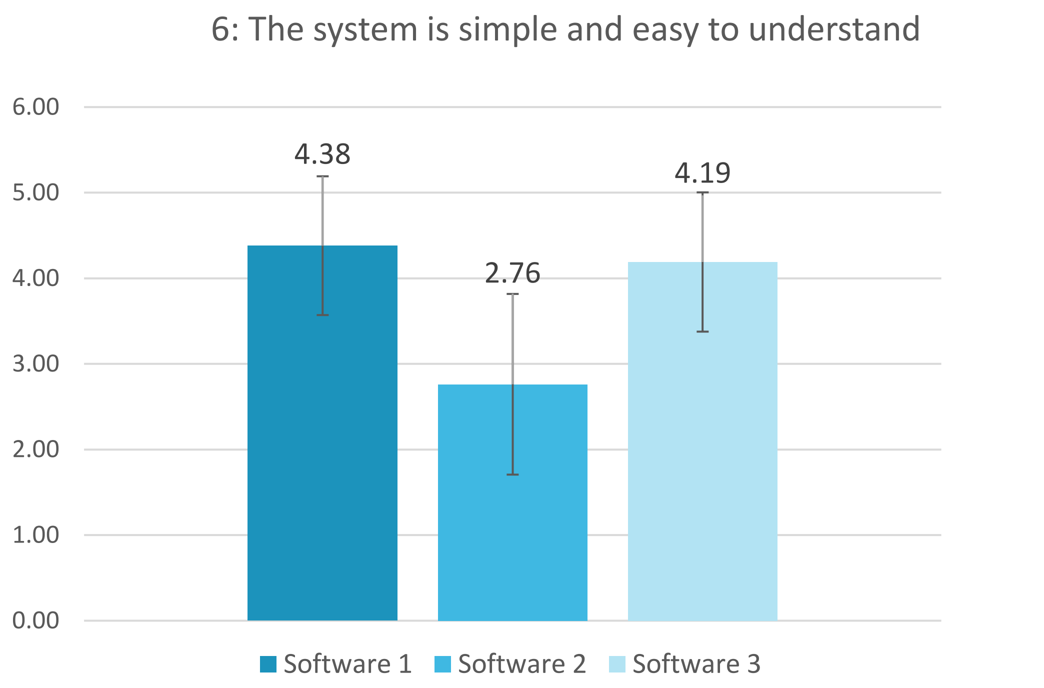

Figure 26: Average results and deviations of question 6 of TUQ

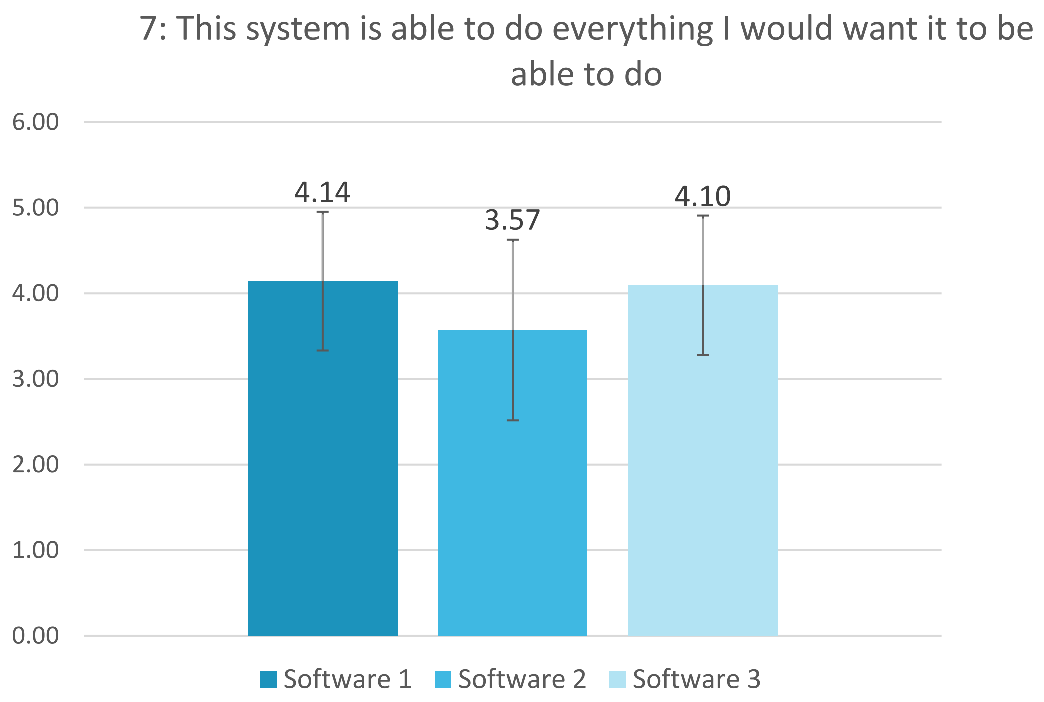

Figure 27: Average results and deviations of question 7 of TUQ

8.1.3 Interaction quality

Interaction quality-section contained four questions that inquired how easily the respondents could talk to or hear the clinician with the software and the respondents' feelings if they were able to express themselves effectively. The last question of this section was about if the respondent was able to see the clinician as well as met in person. The results of those questions can be viewed from the Figures 29 - 32. The total average scores of interaction quality-component is seen in figure 28.

In the questions of this component software 1 received average scores 4.29 ± 0.96, 4.71 ± 0.56, 4.43 ± 0.68 and 3.90 ± 1.09. Software 2 received the averages of 4.10 ± 0.83, 4.57 ± 0.75, 3.76 ± 1.09 and 3.81 ± 0.93 and software 3 got averages scores of 4.38 ± 0.80, 4.48 ± 0.93, 4.10 ± 0.89 and 3.86 ± 1.06.

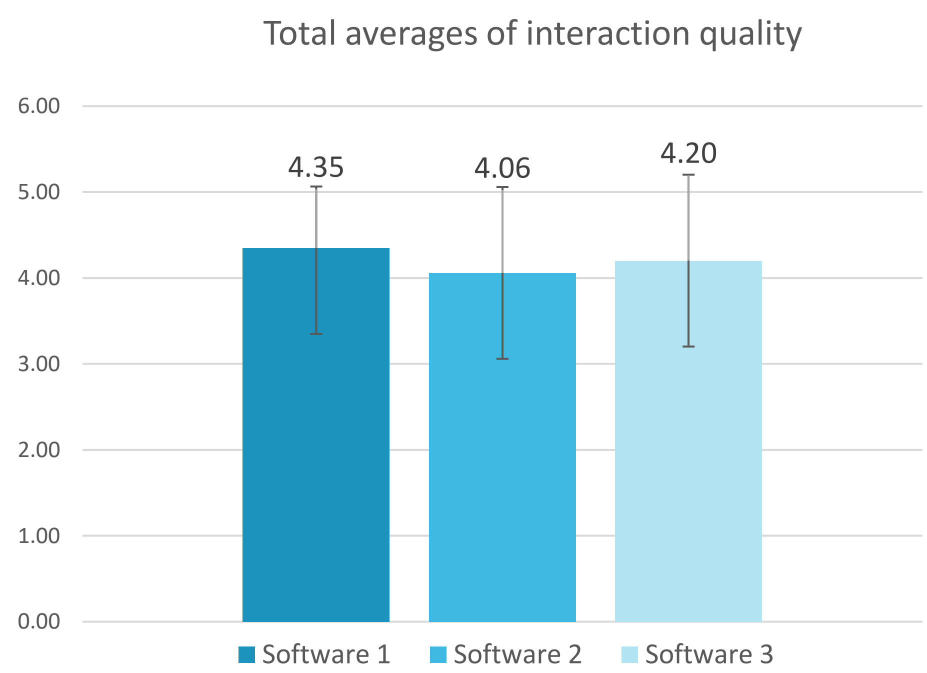

Figure 28: Average results and deviations of interaction quality

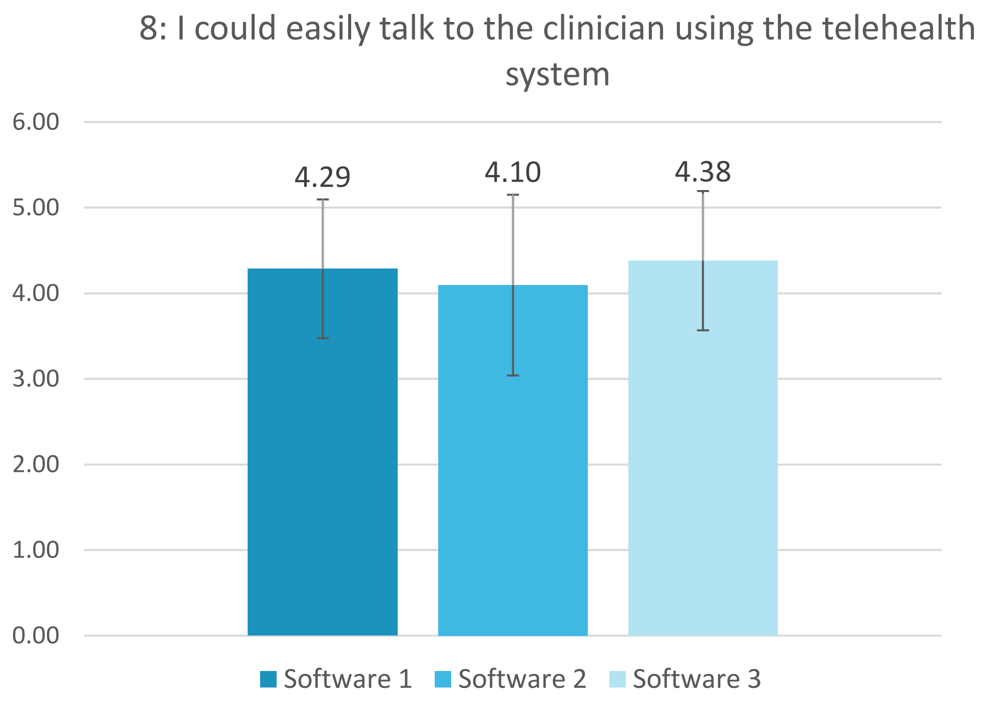

Figure 29: Average results and deviations of question 8 of TUQ

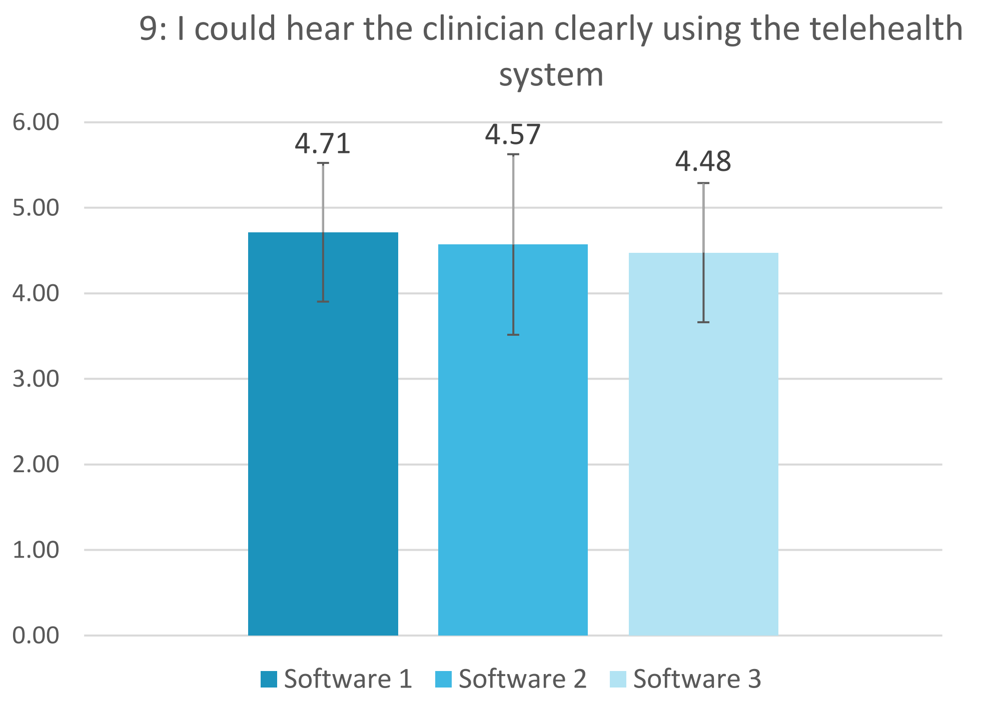

Figure 30: Average results and deviations of question 9 of TUQ

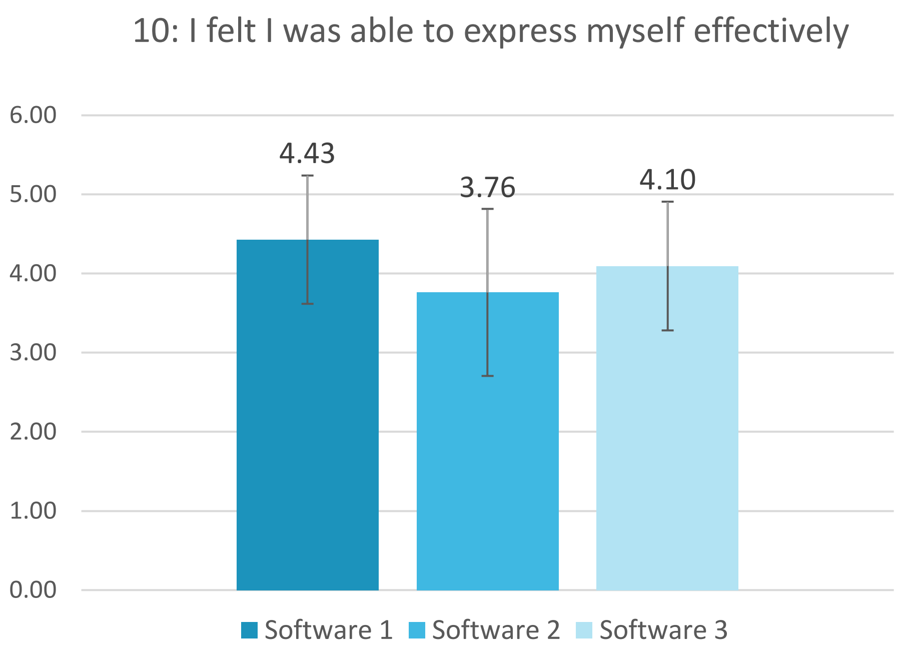

Figure 31: Average results and deviations of question 10 of TUQ

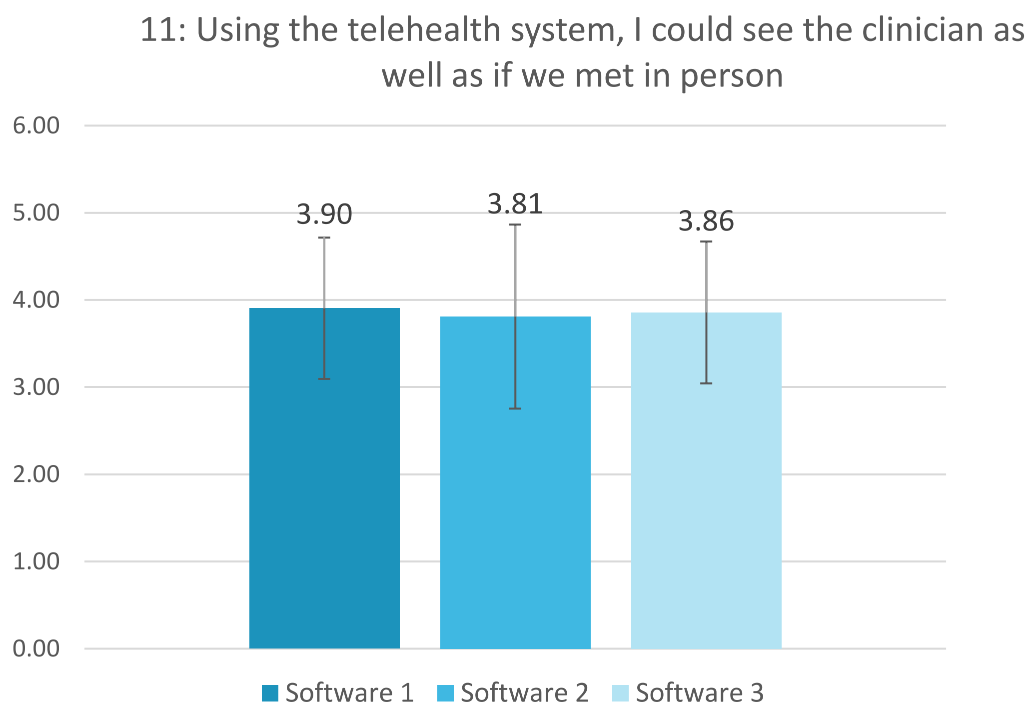

Figure 32: Average results and deviations of question 11 of TUQ

8.1.4 Satisfaction and future use

The last section of Telehealth Usability Questionnaire was Satisfaction and future use and it contained four questions where respondents gave their rank for every software if they felt comfortable communicating with the clinician with the software and if it is an acceptable way to receive healthcare services. This section also included the respondents to evaluate if they would use this software again in telehealth. The last question was if the respondent in overall was satisfied with the software. The results of those questions can be viewed from the Figures 34-37. The total average scores of satisfaction and future use-component is seen in figure 33.

In the questions of this component software 1 received average scores 4.00 ± 0.77, 4.29 ± 0.72, 4.48 ± 0.60 and 4.38 ± 0.67. Software 2 received the averages of 3.29 ± 1.15, 3.19 ± 1.12, 3.24 ± 1.41 and 3.14 ± 1.20 and software 3 got averages scores of 3.90 ± 0.83, 4.05 ± 0.74, 4.19 ± 0.81 and 4.05 ± 0.80.

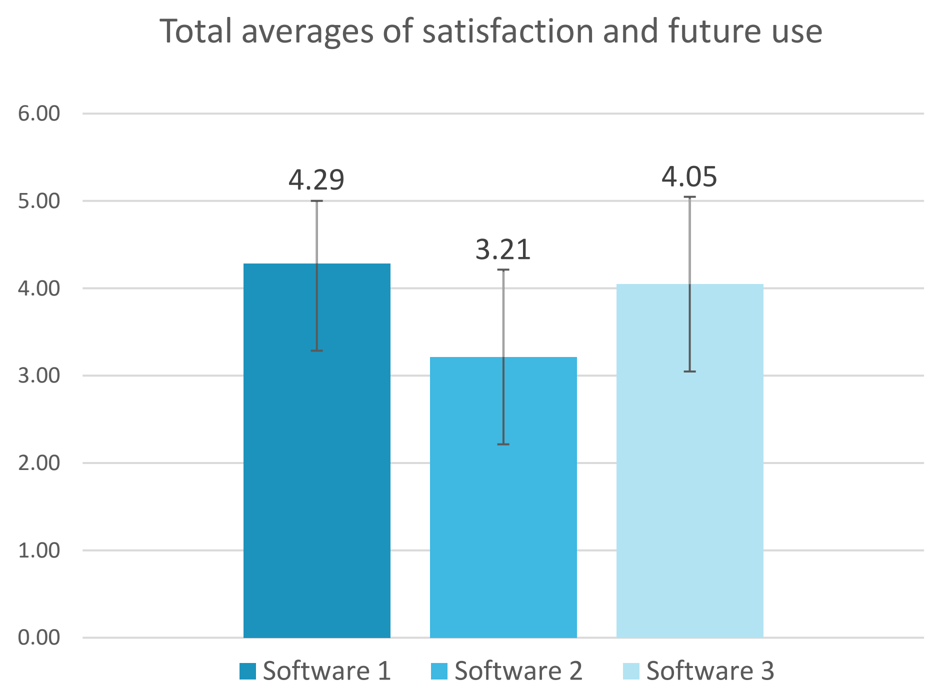

Figure 33: Average results and deviations of satisfaction and future use

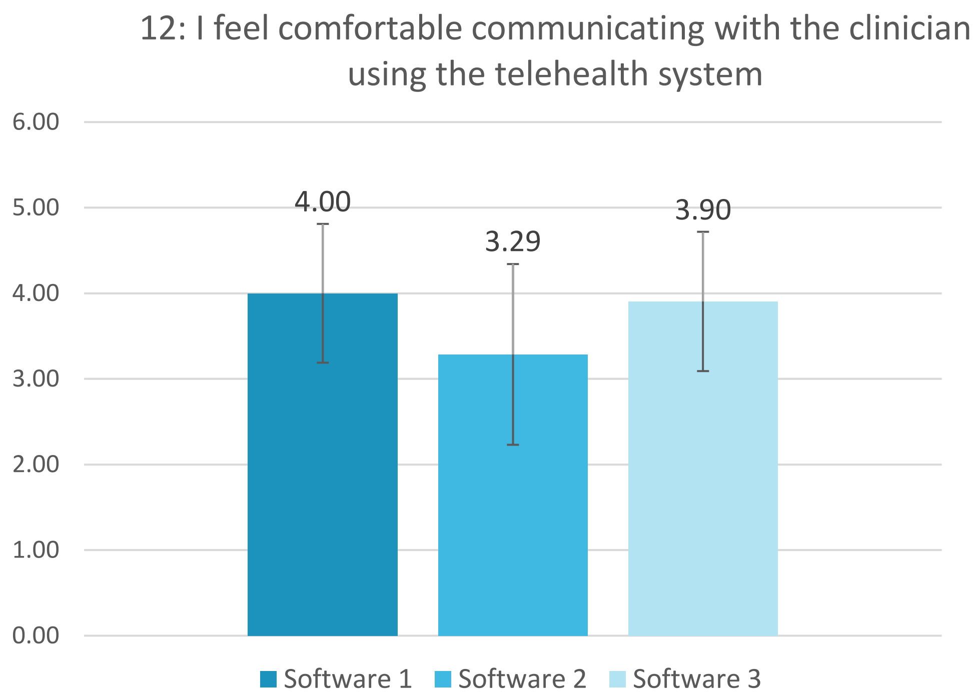

Figure 34: Average results and deviations of question 12 of TUQ

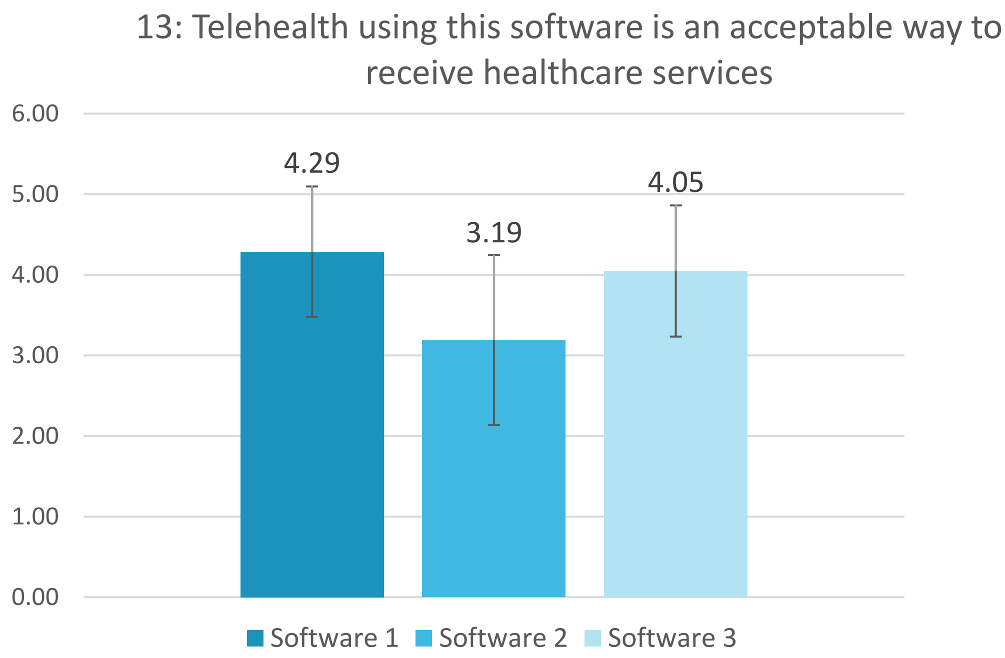

Figure 35: Average results and deviations of question 13 of TUQ

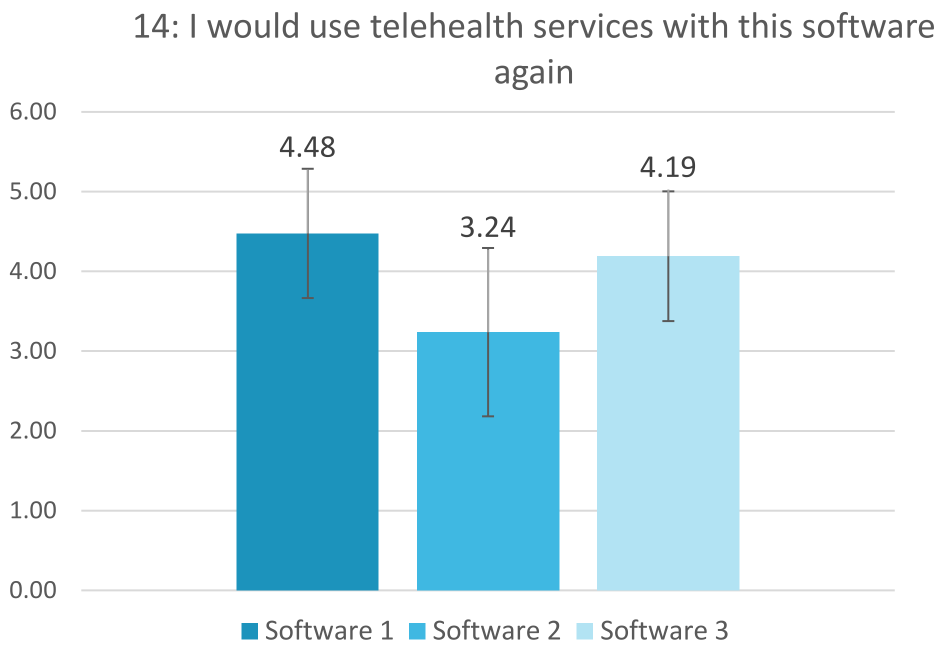

Figure 36: Average results and deviations of question 14 of TUQ

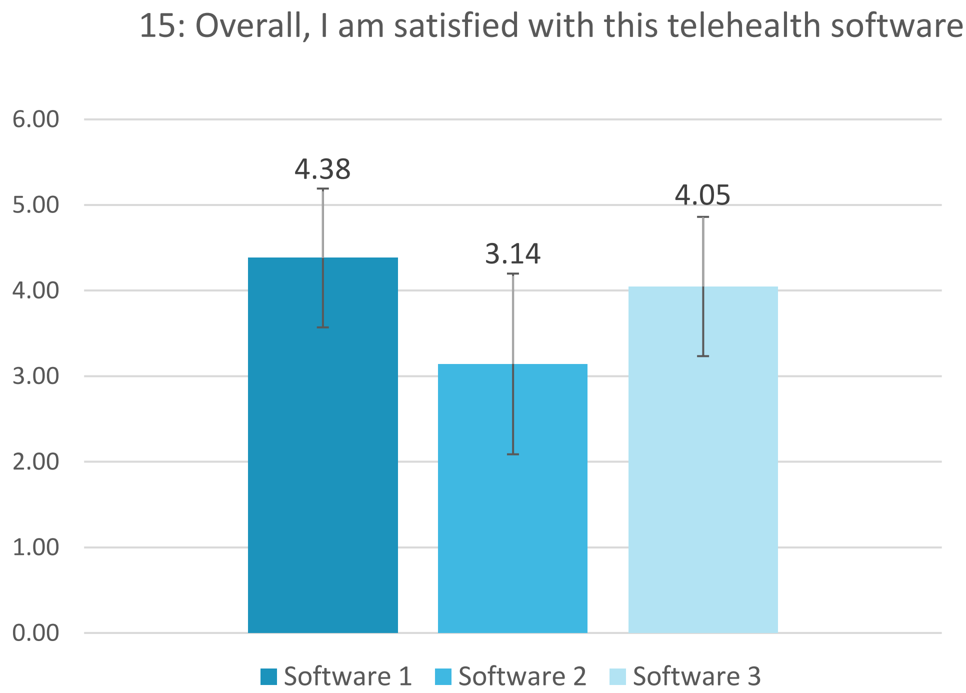

Figure 37: Average results and deviations of question 15 of TUQ

8.2 NASA-TLX

The participants' workload while performing tasks in video-based meetings with different software was measured with the Nasa Task Load Index Test (NASA-TLX). NASA-TLX evaluates the workload with six different components, mental demand, physical demand, temporal demands, frustration, effort and performance, with Likert-scale from 1 to 5. The lower average means lower workload in all components, except in the performance component where higher average means better performance. The results of NASA-TLX are shown in Figures 39 – 44. The total scores of NASA-TLX are seen in Figure 38.

The lowest workload while performing video-based telemedicine meetings was estimated to be in software 1 in five components. The averages with standard deviation were in mental demand 1.42 ± 0.60, in physical demand 1.19 ± 0.40, in temporal demand 1.29 ± 0.46, in effort 1.48 ± 0.75 and in frustration 1.14 ± 0.36. The second lowest averages received software 3 averages being in mental demand 1.85 ± 0.85, in physical demand 1.24 ± 0.43, in temporal demand 1.52 ± 0.68, in effort 2.04 ± 1.12 and in frustration 1.43 ± 0.74. The highest averages and highest workload in five components were with software 3 and averages were in mental demand 2.76 ± 0.94, in physical demand 1.43 ± 0.81, in temporal demand 1.90 ± 1.00, in effort 2.86 ± 1.01 and frustration 2.24 ± 1.09. The performance-component was drop out from the total average because of reliability issue that is further expressed in the chapter 9.4. In performance component software 1 and 3 received the highest average when both received the average of 4.19. The standard deviation of software 1 was 1.17 and software 3 was 0.98. Software 2's performance average was evaluated as 3.48 ± 1.21. The results of NASA-TLX are further discussed and analyzed in the chapter 9.

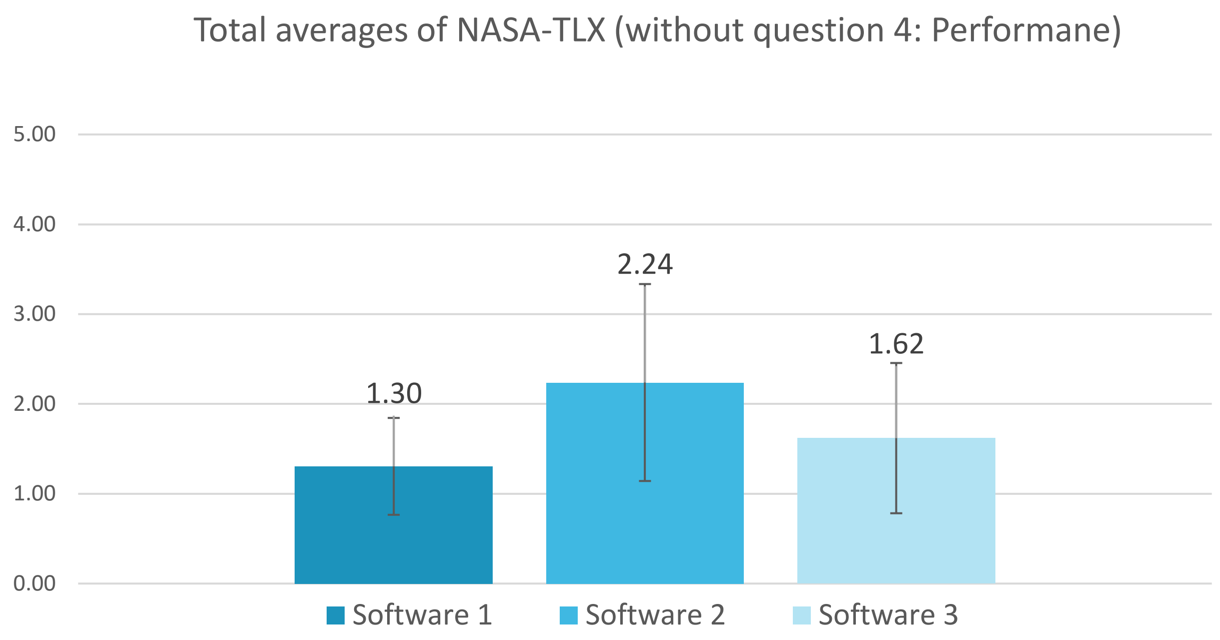

Figure 38: Total average results and deviations of NASA Task Load Index

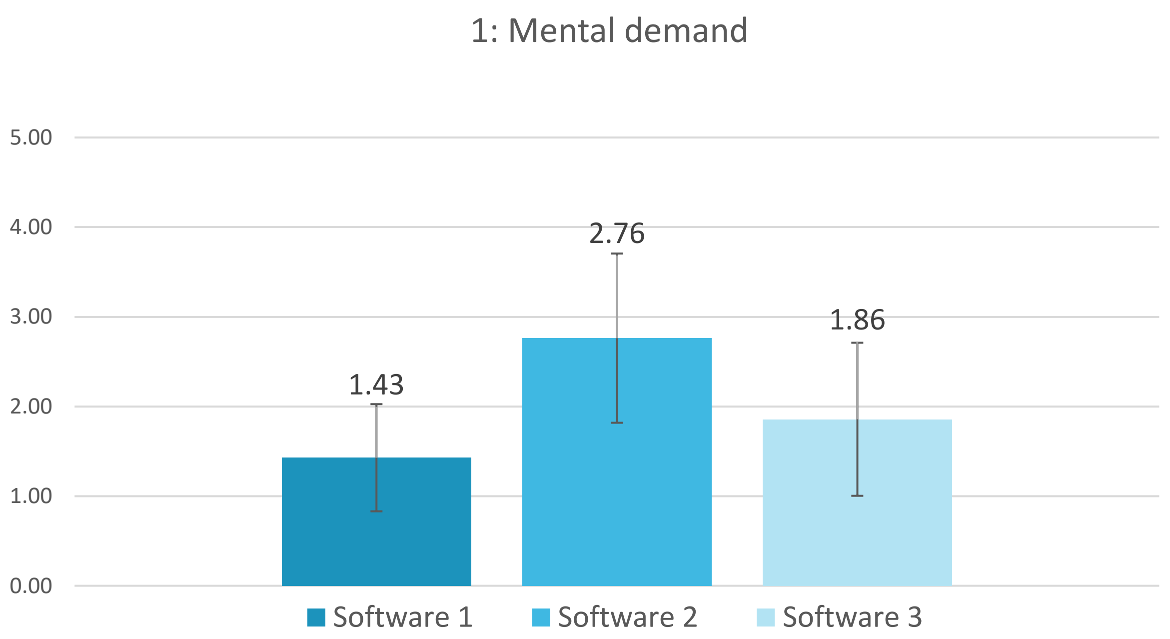

Figure 39: Average results and deviations of mental demand of NASA Task Load Index

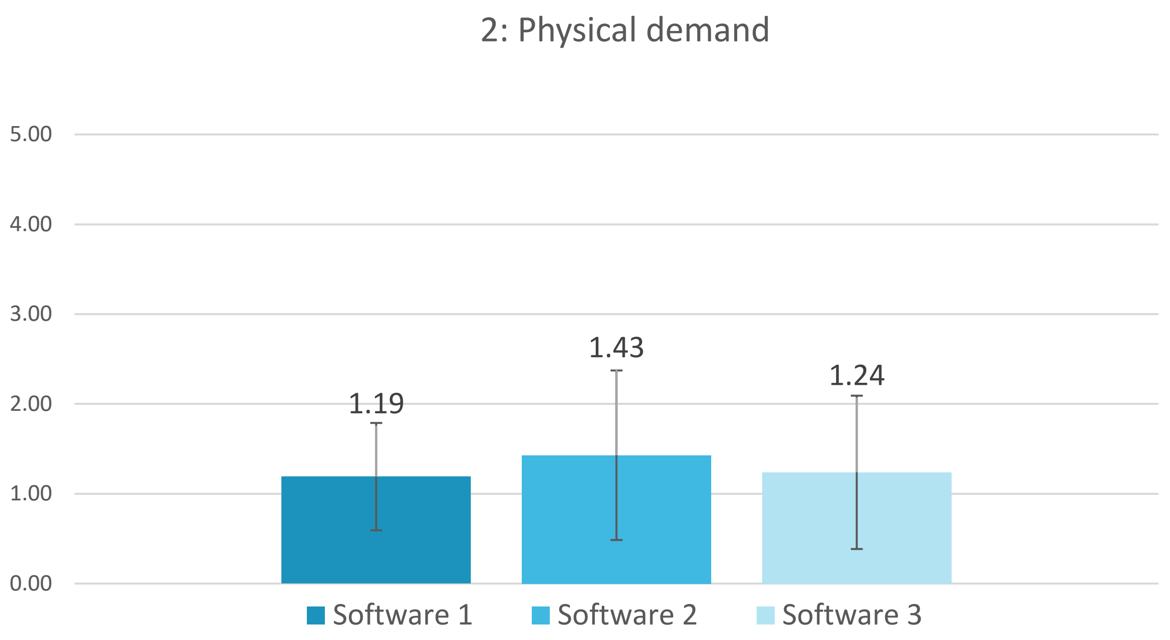

Figure 40: Average results and deviations of physical demand of NASA Task Load Index

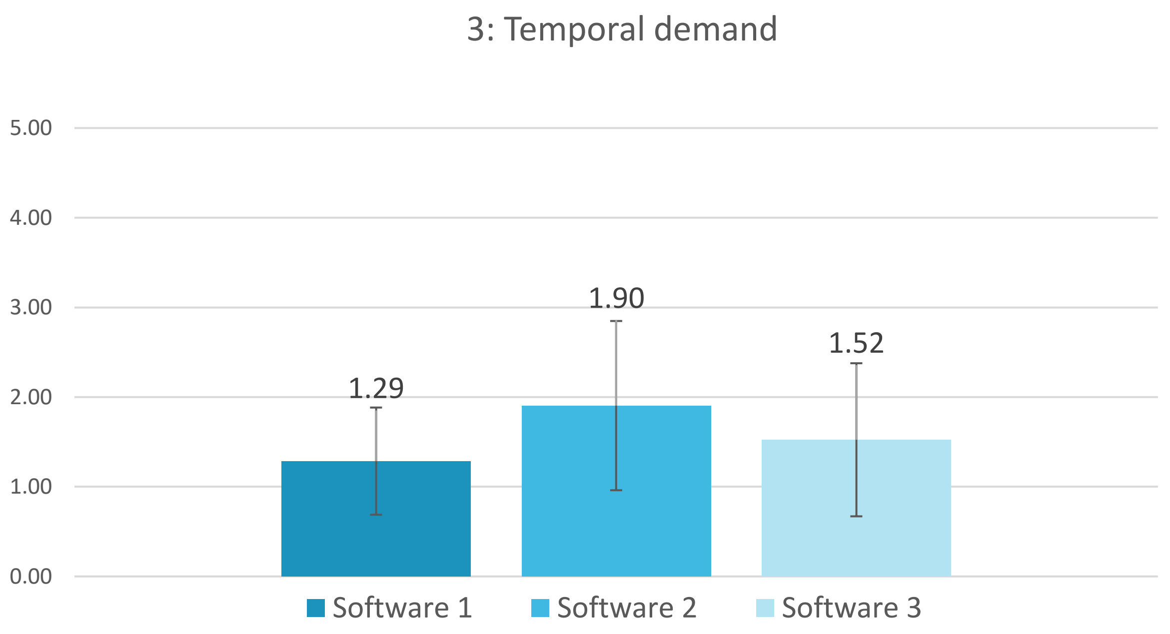

Figure 41: Average results and deviations of temporal demand of NASA Task Load Index

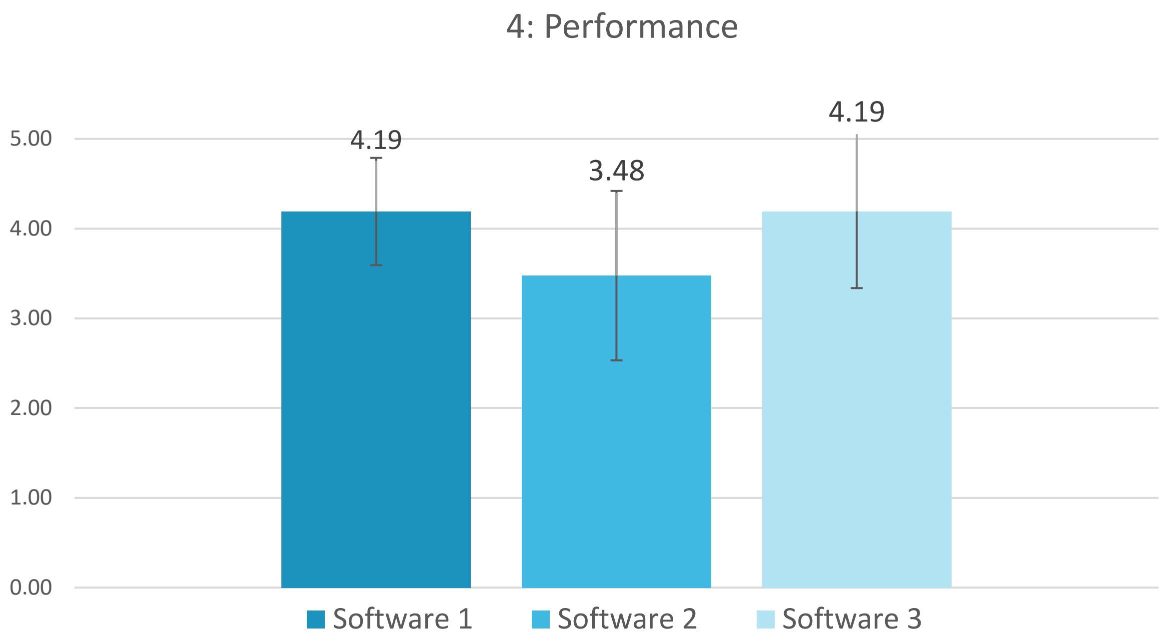

Figure 42: Average results and deviations of performance of NASA Task Load Index

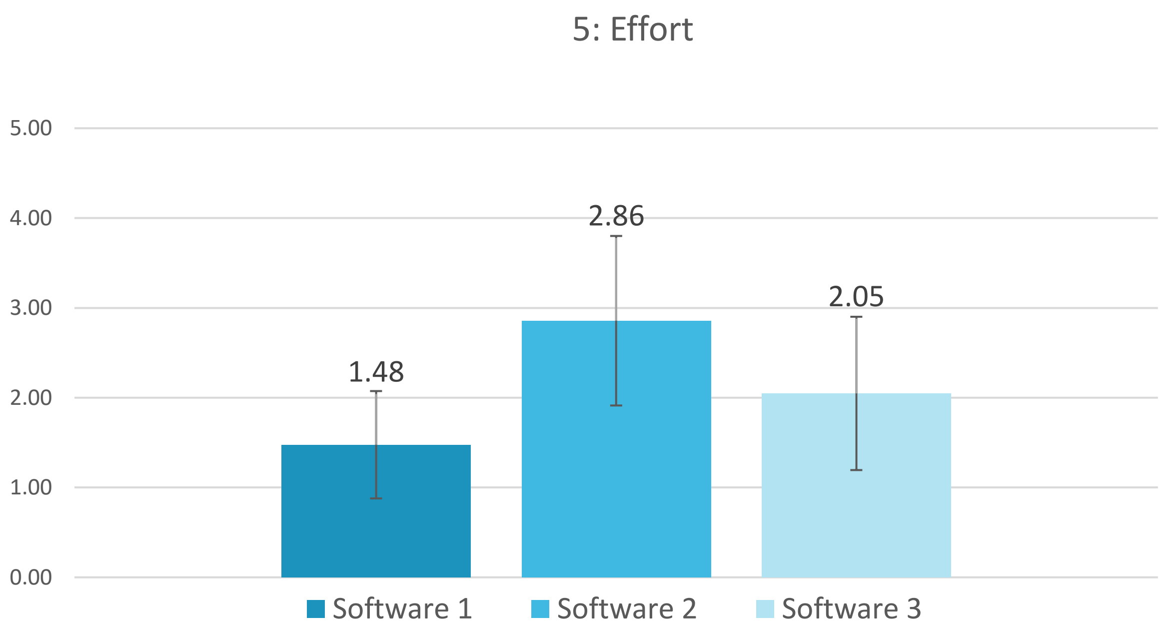

Figure 43: Average results and deviations of effort of NASA Task Load Index

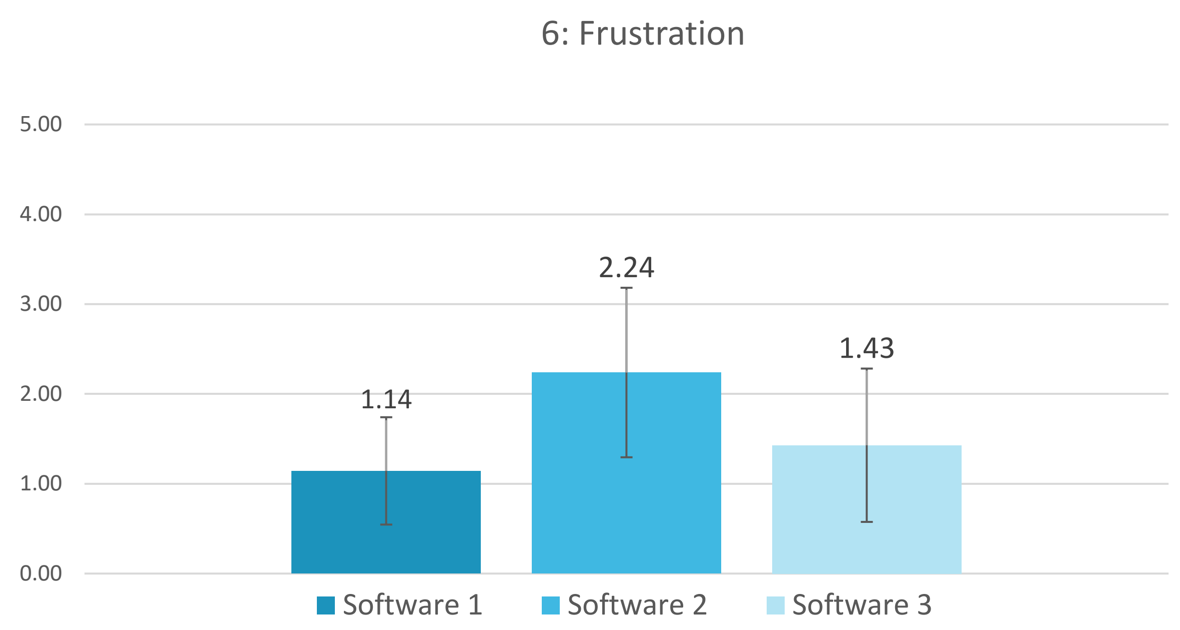

Figure 44: Average results and deviations of frustration of NASA Task Load Index

8.3 Participants' preferences

The participants ranked tested software, which they prefer to use in video-based telemedicine meetings. The rankings went from the most preferred software to the third preferred software. The result of rankings is shown in Table 4.

Table 4: Participants' references of tested software

	The most preferred software	The second most preferred software	The third most preferred software
Software 1	13	6	2
Software 2	1	6	14
Software 3	7	9	5

8.4 Participants' comments

The participants gave their comments about tested software in the end of the testing session. The comments were mostly made about usability errors in software and examples of those comments can be seen in Table 5.

Table 5: Examples of participants' qualitative comments about tested software. The comments are translated from Finnish to English.

Software 1	Software 2	Software 3
No extra waiting times or downloading extra features or add-on.	Bad usability. Too complex to get into the meeting.	The button for leaving the meeting was hard to discover.
Interface did not clearly communicate whether the microphone was on or off	Also chat-window was too small. Hard to read the text in it.	The icons were hard to understand what they meant.
Easy to use and functional	Trying to use the software almost caught all the attention from the meeting and communicating with the doctor	The first impression about the login was not pleasant.

9 Discussion

Aim of the thesis was to study usability and workload of video-based telemedicine software and get patients' point of view from those aspects in simulated video-visits with a doctor. The results of the study can be used by healthcare professionals to get further knowledge about the state of usability of video-based telemedicine software and they can improve their guidance to patients when they start to use video-based telemedicine software. The results can be benefit also by software providers to get valuable information about what should be taken care of when designing video-based telemedicine software.

In the usability test three video telemedicine software were tested. There were 21 participants who gave their opinion about software's usability and workload after they had used software. The opinions were gathered with Telehealth Usability Questionnaire (TUQ) and Nasa Task Load Index-test (NASA-TLX). Also, participants gave their preferences about tested software and qualitative comments were gathered from participants. In the test participants performed a simulated video-visit with a doctor with given tasks that considered using software and communicating with the doctor. There did not occur any major issues during tests and there were no dropouts. Still some participants experienced some connectivity problems which caused some audio problems. The experienced connectivity problems were still mild and there were not any big differences between the participants.

9.1 Usability

Usability of three video-based telemedicine software were evaluated by modified Telehealth Usability Questionnaire with Likert-scale from 1 to 5. Software 1 received the highest average score (4.33 ± 0.80) while the lowest average score (3.39 ± 1.16) was for software 2. Software 3 had an average score (4.13 ± 0.84) which was close with software 1. TUQ examined usability more deeply with four different components. Those components were ease of use and learnability, interface quality, interaction quality and satisfaction and future use. Three components, ease of use and learnability, interface quality and satisfaction and future use, had more or less the same average scores than the total averages of TUQ. In the Interaction quality-section there weren't any big differences and that is discussed more deeply in the chapter 9.1.3.

One major inspiration for this study was that there was a lack of video-based telemedicine usability studies that were based in Finland. Though there are international usability studies about the problem in question it is hard to use the results of those studies in Finland. One reason for that is that usability is a cultural issue. According to Amant (2017) external factors play a huge role in how users use different products. For example, in Finland there can be different external factors than in the United Kingdom or in the United States. When comparing the received TUQ scores to other video-based telemedicine usability studies that also used TUQ, the results were much in line with those studies. In the studies of Layfield's et al (2020) and Johnson's et al (2020) software received good ratings with TUQ. In Likert scale from 1 to 7 the ratings were 6.01 and 5.5. It is safe to say that perceived usability in average is much the same in this research than in the research of Layfield et al. (2020) and Johnson et al. (2020). It is good to keep in mind that there are some limitations that must be taken care of when comparing the results of this study to other studies because there were some differences in the research methods and setups. For example, in the research of Layfield's et al. (2020) the patients had the possibility to get help and instructions before having the actual video meeting with the medical personnel and in this study the participants were not able to get instructions or help. When comparing the results of this study to Agnisarman's et al. (2017) which had the similar test method and setup, we can see that their results are also much similar. In the study of Agnisarman et al (2017) they had a different usability survey, but the survey was using the same Likert-scale that this study. In the study of Agnisarman et al. (2017) four different video platforms received usability scores between 3.28 – 4.15 and in this study three different software received usability scores between 3.39 – 4.33.

There were not any major connection problems in this research mainly because the testing environment was in the campus of Xamk, and the participant and the doctor were in the rooms next to each other. But when looking to studies that had participants that had a real-life experience with the video-based telemedicine software we can see that connections problems are an issue. For example, in the studies of Thelen-Perry's et al. (2018), Patel's et al. (2021) and Johnson's et al. (2020) there were documented connections issues and those issues affected usability negatively. So, the usability scores in this study may be a bit higher than when using these software in real-life because according to other studies the connection problems do occur in real-life video-based telemedicine meetings, and it makes usability worse.

According to Karapanos et al. (2009) during the user's first experience with a product the most appreciated sectors of usability are learnability and aesthetic stimulation. We can see that in the Learnability-section of Telehealth Usability Questionnaire software 2 received the lowest average score of tested software. Also, software 2 received the worst score in Interface quality- and Satisfaction and future use-sections. It could be that software 2 received the lowest average score because respondents thought it was the hardest software to learn to use and it affected other sections of usability because learnability is one of the most appreciated sectors of usability during the first time using a product. The same effect could also be in aesthetic stimulation because software 2 received the lowest score also in that section.

9.1.1 Ease of use and learnability

The first question of this section inquired about whether the software was simple to use. The simplest software to use was software 1 with an average of 4.38 ± 1.12. According to respondents, software 1 was also easiest to learn (4.67 ± 0.58) and they believed they could become productive quickliest with software 1 (4.24 ± 0.70). The lowest averages received software 2 with the averages of 2.71 ± 1.06, 3.14 ± 1.20 and 3.29 ± 1.00 and software 3 got averages scores of 4.19 ± 0.81, 4.29 ± 0.72, 4.00 ± 0.77.

The interesting factor in this section is that tested software included different login steps. In the research of Agnisarman et al. (2017) the extra steps in the initiation phase caused lower usability and higher mental demand. In software 1 the patient was able to get the right meeting with the right URL address and choose the right professional from the contact list. To access the right meeting with software 2 the patient only needs the right URL address. With software 3 after accessing the right URL the patient needs to download Vidyo-plugin with Google Chrome-browser's own web store features. Though the software 2 had the least steps in the login-phase it still received the worst scores from tested software in this section. This differed from the results of Agnirman's et al (2017) where it seemed that the most steps affected negatively to usability and mental demand. Software 2 also received bad qualitative comments about too much complexity to get into the meeting which could indicate that software 2 had a more difficult initiation phase compared to software 1 and 3.

9.1.2 Interface Quality

Software 1 received the highest averages in this section and total average was 4.31 ± 0.78 while software 3 received 4.12 ± 0.86. Software 2 got the lowest average scores of this section and the total average of this section was 3.14 ± 1.09. The highest average of different question in this section was found in question 5, which inquired if the participant liked the software. Software received the average 4.43 in that question while software 2 got 3.05 ± 1.02 and software 3 got 4.10 ± 0.94. The lowest average scores in different question in this section was in question 6 which was about if the software was simple and easy to understand. In that question Software 2 gained the average of 2.76 ± 1.14. The averages scores of this section were much like the averages scores in the first section of TUQ, Ease of use and learnability.

Aesthetic stimulation according to Karapanos et al. (2009) is one of the most appreciated sectors of usability when using a product for the first time. When taking a closer look to the questions of TUQ we can see that the Interface quality-section holds questions that can be drawn to aesthetic stimulation. So, the participants may have been fooled by the aesthetic stimulation when they have answered to these questions and the differences of this section could be explained by aesthetic stimulation. But it is also possible that the reason for the differences in this section could be explained by the same reason as the reason explained in the Ease of use and learnability-section, which was too complicated login-methods.

Jakob Nielsen's (2020) eighth usability heuristic considered aesthetic and minimalist design. This heuristic said that the interface should only contain the needed information and even rarely information should be removed. According to Nielsen (2020) every extra unit in the interface diminishes the relative visibility of other units. Software 2 included chat-window and other extra-features that was not in another tested software so it could be possible that because there were more functionalities in software 2 it confused the participants. Still, it is debatable if for example a well-designed chat-window would distract the interface.

9.1.3 Interaction Quality

All software received almost the same average in this section and the averages were in the software order 4.35 ± 0.88, 4.06 ± 0.95 and 4.20 ± 0.94. The biggest differences in this section between the averages was in question 10, which asked if the respondent felt they were able to express himself/herself effectively. In this question the highest average gained software 1 with the average of 4.43 ± 0.68 while others gained 3.76 ± 1.09 and 4.10 ± 0.89.

There were not any such differences in this section when comparing the results of this section to other sections of TUQ. And the question that had the lowest difference in average scores of tested software of the whole TUQ questions was in this section. In question 11 that inquired if the participant was able to see with the software doctor as well if they met in person the tested software received average scores 3.90 ± 0.77, 3.81 ± 0.93 and 3.86 ± 1.06. The reason for the small differences in these sections could be that the communication with the doctor and tasks to perform in the visit were made-up. Though according to qualitative comments given by the participants there were some usability issues and some participants said that using software 2 distracted communicating with the doctor. Still the usability did not seem so bad that it would have strongly affected the communication with the doctor because there were not big differences with the software average scores of this section.

9.1.4 Satisfaction and future use

The participants were generally satisfied with the tested software though there were some differences between average scores of this section. Software 1 received the highest average scores 4.29 ± 0.91 while software 2 got the lowest average scores 3.21 ± 1.20. Software 3 gained the total average of 4.05 ± 0.79 in this section. The highest difference was in question 14, which inquired if the participants would use telehealth service with this software again, and in question 15, where the participants estimated if they were satisfied with that telemedicine software. Software 1 received 1.24 better average scores in those questions than software 3. Interesting part of this section was question number 15 where software 2 received the lowest average score of Telehealth Usability Questionnaire of this study which was 3.14. The result suggests that the participants had a mediocre feeling with software 2 in this research.

9.2 Workload

The workload of participants' while doing the tasks in the test during the video meeting with each tested software was measured with the Nasa Task Load Index Test (NASA-TLX). It has six components, which are mental demand, physical demand, temporal demands, frustration, effort and performance and with each component the participants evaluated how much was the workload.

In five components software 1 received the lowest workload. and the second lowest workload received software 3. The highest averages and highest workload in five components were with software 3. In performance component software 1 and 3 received the most positive score when both received the average of 4.19 and participants evaluated that they performed better with those software than with software 2. Software 2's performance average was evaluated as 3.48.

As told in chapter 8.1.1 in Agnisarman's et al (2017) the initiation phase led to lower usability and higher mental demand. As in Ease of use and learnability-section of Telehealth Usability Questionnaire also in Mental demand-section of NASA-TLX the software that had least steps in the login phase had the highest mental demand.

The interesting point in this section was that Software 3's Vidyo-plugin download did not affect the usability according to TUQ-scores, but the Vidyo-plugin download possibly increased the effort that users had to do when using the software 3 according to Effort-scores of NASA-TLX. The difference of average scores of Effort-questions between Software 1 and 3 was the highest when comparing all the questions in the Nasa Task Load Index-questionnaire.

There were not any differences in the physical demand. This can be explained by low physical demand that normally is associated with computer use. Also, the software used in this research were designed in a manner, so the users do not need any specific physical skills, for example fine motor skills.

The question 4 of NASA-TLX, which inquired about the performance demand, had reliability issues because the scale should have been the other way. In the NSA-TLX questionnaire of this research the software who had the most positive performance according to respondents had the highest averages when in other questions the most positive endpoint was the lowest averages. The average deviation of this question also indicates that reliability of this question is debatable because software 1 and 2 received the highest deviations in this question when comparing the other questions of NASA-TLX.

9.3 Participants' preferences

The participants also gave their preferences about tested software and which they prefer to use in video-based telemedicine meetings. The preferences scale went from the most preferred software to the third preferred software.

Software 1 received 13 rankings to the most preferred software while 7 participants preferred the most software 3 and 1 participant valued the most software 2. Software 2 also received the biggest number of rankings to the third most preferred software and the total amount was 14.

Software 2 was estimated to be less preferred than software 1 and 3. This follows the results of Telehealth Usability Questionnaire and NASA Task Load Index-questionnaire where software 2 was evaluated to be harder to use and mentally more demanding. When comparing the results of the participants' preferences and the results of TUQ we can see that there could be a connection with usability and user experience and with usability and system acceptability as Nielsen (2010, pp. 25) and Sonderegger et al (2019) have stated.

9.4 Reliability

The results of this study should be taken with a little consideration. For example, as told in chapter 9.2 the question 4 of NASA-TLX had the scale problems which may have affected reliability of that question. The software included different features and the test procedure was slightly different for each software. For example, tested software had a different login-methods and with software 1 it was impossible to set the video meeting in full screen-mode and because of that software 1 did not have that task in the procedure. Also, a larger number of respondents would have made the results more reliable. According to Nielsen (2010, pp. 224) at least 30 respondents is the adequate number to get reliable results of a usability questionnaire. On the other hand, the adequate number of participants is debatable because in the study of Agnisarman et al. (2017) there were only 19 participants.

Despite some reliability issues the results of this study can be seen reliable. The questionnaires used in this research were considered reliable. As Hajesmaeel-Gohari & Bahaadinbejy (2021) researched the Telehealth Usability Questionnaire is the most common tool used in telehealth studies. And, the second most common tool is Telemedicine Satisfaction Questionnaire, which is a basis of the usability factors of telehealth in Telehealth Usability Questionnaire (Parmanto et al. 2016). Also, Nasa Task Load Index Test is popular and useful tool to measure workload while performing a task in different environments (Hart, 2006; Colligan et al. 2015; Agnisarman et al. 2017).

The method of this study was much similar than in the study of Agnisarman et al. (2017) so the method can also be seen as a valid way to measure the usability of video-based telemedicine software. There were also some other features in this study, which increased the reliability of this study, for example the Latin square, that secured a reliable test order between the different software.

9.5 Limitations

The research contains some limitations which should be considered in the results of the study. The participants of this study were teachers and students at South-Eastern Finland University of Applied Sciences (XAMK) so there cannot be any assumptions made at the whole population-level from the results of this study. It is fair to say that these results are directional how higher educated feel the usability of video-based telemedicine software.

The results also indicate the state of usability of tested software during the first time. When using these software again, the usability can be different because for example there is no need to download the Video-plugin with software 3 when using that software for the second time. According to Karapanos et al. (2009) and Sonderegger et al. (2012) also the appreciated aspects of usability changes during the usage of product. For example, during the first-time usage the most appreciated aspects are aesthetics and learnability but during long-term usage the most appreciated aspect is usefulness.

It is also good to keep in mind that the research was done in 2018 so there may have been changes in tested software after that, so the result of this study does not represent the status of these software at the date of the publication of this thesis.

9.6 Further research and development

This study concentrated on only the usability when using the first time the video-based telemedicine software. To gain more in-depth knowledge about usability of the video-based telemedicine software the same kind of study should be implemented when the patients have used software multiple times. As told before the appreciated aspects of usability varies during the usage of software, so it would be important to gain more insight about usability when the patients' have more experience with video-based telemedicine software (Karapanos et al., 2009; Sonderegger et al., 2012).

Also, the study should be implemented with real patients with real doctors in real situations. It is good to keep in mind that this was a made-up doctor's appointment with a made-up manuscript so in a real-life situation the situation could change the experienced usability. Usability is a cultural issue and external factors have a big influence on usability and in real-life situations the external factors may differ from the made-up test environment (Amant, 2017). But when doing this kind of research with real patients the ethical aspects should be taken care into notice more deeply.

The results of this study helped the social and health care authorities to give a better and detailed guidance to patients who receive care via video-based telemedicine software. Also, the results of this study helped software vendors to get clearer view of the usability of their software. Though some vendors whose software were tested in this study did major technical updates after the study before having a chance to look the results.

9.7 Conclusion

In this study there were some differences in usability of video-based telemedicine software. The biggest difference was in ease of use and learnability, which may have influenced the different login-methods with different software. Also, when interviewing the participants, it came clear that there were some usability errors in software for example some button icons were hard to discover, which also may influence ease of use and learnability. The smallest difference of usability was in the quality of interaction and the reason for this may be that the communication with the doctor was made-up and the same for every respondent. Also, there were differences in the workload experienced while performing a video meeting with a doctor with different software. The reason for the differences could be drawn to the same reasons as the differences in usability. According to results of Telehealth Usability Questionnaire and Nasa Task Load Index of this study software 1 had a better usability which according to Nielsen (2010, pp. 25) and Sonderegger et al. (2019) means that software 1 potentially has better system acceptability and user experience.

This research contains several limitations, and this research clearly points out how complicated issue usability is. The actual state of usability is hard to discover with just one time use, especially when it is the first time using the software. As discussed earlier, according to Karapanos et al. (2009) and Sonderegger et al. (2012) the aesthetic is much more appreciated when using a product for the first time, which influences usability. So, the differences received in this study could be also explained by differences in aesthetic and not just by usability. To gain more knowledge of the usability of the video-based telemedicine software the usability questionnaires should be done again when respondents have more experience with software and preferably with real patients.

References

Agnisarman, S. O., Chalil Madathil, K., Smith, K., Ashok, A., Welch, B., & McElligott, J. T. (2017). Lessons learned from the usability assessment of home-based telemedicine systems. *Applied ergonomics*, *58*, 424–434. https://doi.org/10.1016/j.apergo.2016.08.003

Amant, K.S. (2017). The cultural context of care in international communication design: a heuristic for addressing usability in international health and medical communication. *Communication Design Quarterly Review*, *5*, 62-70. https://doi.org/10.1145/3131201.3131207

Amberg, M., Hirschmeier, M., & Wehrmann, J. (2004). The Compass Acceptance Model for the analysis and evaluation of mobile services. *International Journal of Mobile Communication*, *2*, 248-259. http://doi.org/10.1504/IJMC.2004.005163

Chisnell, D. (2016). Democracy is a design problem. *Journal of Usability Studies*, *11*, 124-130.

Colligan, L., Potts, H. W., Finn, C. T., & Sinkin, R. A. (2015). Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International journal of medical informatics*, *84*(7), 469–476. https://doi.org/10.1016/j.ijmedinf.2015.03.003

Davis, F.D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int. J. Man Mach. Stud*., *38*, 475-487. https://doi.org/10.1006/imms.1993.1022

Dhahri, A. A., Iqbal, M. R., & Pardoe, H. (2020). Agile Application of Video Telemedicine During the COVID-19 Pandemic. *Cureus*, *12*(11), e11320. https://doi.org/10.7759/cureus.11320

Enlund, T. (2018). Potilastietojärjestelmän vioista haetaan hyvitystä miljoonavahingoille – Lifecare ei toimi Päijät-Hämeessä vieläkään [The refunds are seeked because the errors of EHR's – Lifecare still does not work at Päijänne Tavastia]. https://yle.fi/uutiset/3-10165266

Eura2014. (n.d.) *Project Plan of DigiSote-project*. https://www.eura2014.fi/rrtiepa/projekti.php?projektikoodi=S20798

Hajesmaeel-Gohari, S., & Bahaadinbeigy, K. (2021). The most used questionnaires for evaluating telemedicine services. *BMC Medical Informatics and Decision Making, 21*, 36. https://doi.org/10.1186/s12911-021-01407-y

Hart, S.G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*, 904 - 908. https://doi.org/10.1177/154193120605000909

Hyppönen, H., Aalto A-M., Reponen, J., Kangas, M., Kuusisto-Niemi, S. & Heponiemi, T. 2018. *Kansalainen – pystyn itse? Kokemuksia sosiaali- ja terveydenhuollon sähköisistä palveluista kansalaisille [The citizens experiences of electronic services of social and healthcare].* The Finnish Institute for Health and Welfare.

Hyppönen, H. & Ilmarinen, K. 2016. Sosiaali- ja terveydenhuollon digitalisaatio [The digitalization of social and healthcare]. The Finnish Institute for Health and Welfare.

Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., Shields, A., Rosenbaum, S., & Blumenthal, D. (2009). Use of electronic health records in U.S. hospitals. *The New England journal of medicine*, *360*(16), 1628–1638. https://doi.org/10.1056/NEJMsa0900592

Johnson, B. A., Lindgren, B. R., Blaes, A. H., Parsons, H. M., LaRocca, C. J., Farah, R., & Hui, J. (2021). The new normal? Patient satisfaction and Usability of Telemedicine in breast cancer care. *Annals of surgical oncology*, *28*(10), 5668–5676. https://doi.org/10.1245/s10434-021-10448-6

Kaipio, J., Lääveri, T., Hyppönen, H., Vainiomäki, S., Reponen, J., Kushniruk, A., Borycki, E., & Vä-nskä, J. (2017). Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland. *International journal of medical informatics*, *97*, 266–281. https://doi.org/10.1016/j.ijmedinf.2016.10.010

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J. (2009). User experience over time: an initial framework. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/1518701.1518814

Klaic, M., & Galea, M. P. (2020). Using the Technology Acceptance Model to identify factors tha predict likelihood to adopt tele-neurorehabilitation. *Frontiers in neurology, 11*, 580832. https://doi.org/10.3389/fneur.2020.580832

Knox, L., Gemine, R., Rees, S., Bowen, S., Groom, P., Taylor, D., Bond, I., Rosser, W., & Lewis, K. (2020). Using the Technology Acceptance Model to conceptualise experiences of the usability and acceptability of a self-management app (COPD.Pal®) for Chronic Obstructive Pulmonary Disease. *Health and technology*, 1–7. Advance online publication. https://doi.org/10.1007/s12553-020-00494-7

Krug, S. (2014). Don't make me think, revisited: a common sense approach to web usability. New Riders Publishing. https://doi.org/10.5860/choice.51-6218

Lah, U., Lewis, J.R., & Šumak, B. (2020). Perceived Usability and the Modified Technology Acceptance Model. *International Journal of Human–Computer Interaction*, *36*, 1216 - 1230. https://doi.org/10.1080/10447318.2020.1727262

Layfield, E., Triantafillou, V., Prasad, A., Deng, J., Shanti, R. M., Newman, J. G., & Rajasekaran, K. (2020). Telemedicine for head and neck ambulatory visits during COVID-19: Evaluating usability and patient satisfaction. *Head & neck*, *42*(7), 1681–1689. https://doi.org/10.1002/hed.26285

Lewis, J. L. (1995) IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human–Computer Interaction*, *7*(1), 57-78. https://doi.org/10.1080/10447319509526110

Majrashi, K. & Hamilton, M. (2014). *User Experience of University Websites.* LAP Lambert Academic Publishing.

Malindi, P. (2011). QoS in Telemedicine. In G. Graschew & S. Rakowsky (Eds.), Telemedicine Techniques and Applications. *IntechOpen*. https://doi.org/10.5772/20240

McGowan, J. J. (2008). The pervasiveness of telemedicine: adoption with or without a research base. *Journal of general internal medicine*, *23*(4), 505–507. https://doi.org/10.1007/s11606-008-0534-z

Nguyen, M., Fujioka, J., Wentlandt, K., Onabajo, N., Wong, I., Bhatia, R. S., Bhattacharyya, O., & Stamenova, V. (2020). Using the technology acceptance model to explore health provider and administrator perceptions of the usefulness and ease of using technology in palliative care. *BMC palliative care*, *19*(1), 138. https://doi.org/10.1186/s12904-020-00644-8

Nielsen, J. (2010). *Usability Engineering.* Academic Press

Nielsen, J. (2020). *10 Usability Heuristics for User Interface Design.* https://www.nngroup.com/articles/ten-usability-heuristics/

Nord, G., Rising, K.L., Band, R.A., Carr, B.G., & Hollander, J.E. (2019). On-demand synchronous audio video telemedicine visits are cost effective. *American Journal of Emergency Medicine*, *37*, 890–894. https://doi.org/10.1016/j.ajem.2018.08.017

Norman, D. 2010. *Why design education must change*. http://www.core77.com/posts/17993/why-design-education-must-change-17993

Oulasvirta, A. (2011). *Ihmisen ja tietokoneen vuorovaikutus [Human-Computer interaction]*. Gaudeamus Helsinki University Press

Parmanto, B., Lewis, A. N., Jr, Graham, K. M., & Bertolet, M. H. (2016). Development of the Telehealth Usability Questionnaire (TUQ). *International journal of telerehabilitation*, *8*(1), 3–10. https://doi.org/10.5195/ijt.2016.6196

Patel, M., Miller, R., Haddad, H., An, L., Devito, J., Neff, A., Rajkumar, A., & Ellimoottil, C. (2021). Assessing patient usability of video visits. *mHealth*, *7*, 22. https://doi.org/10.21037/mhealth-20-30

Qiang, J. K., & Marras, C. (2015). Telemedicine in Parkinson's disease: A patient perspective at a tertiary care centre. *Parkinsonism & related disorders*, *21*(5), 525–528. https://doi.org/10.1016/j.parkreldis.2015.02.018

Reponen, J. (2015). Terveydenhuollon sähköiset palvelut murroksessa [The electronic services of healthcare in a turning point]. *Duodecim*, *131*, 1275–1276.

Rodriguez, J. A., Betancourt, J. R., Sequist, T. D., & Ganguli, I. (2021). Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. *The American journal of managed care*, *27*(1), 21–26. https://doi.org/10.37765/ajmc.2021.88573

Sonderegger, A., Zbinden, G., Uebelbacher, A., & Sauer, J. (2012). The influence of product aesthetics and usability over the course of time: a longitudinal field experiment. *Ergonomics*, *55*(7), 713–730. https://doi.org/10.1080/00140139.2012.672658

Sonderegger, A., Uebelbacher, A., Sauer, J. (2019). The UX Construct – Does the usage context influence the outcome of user experience evaluations? In D. Lamas, F. Loizides, L. Nacke, H. Petrie, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2019*, pp. 140-157. https://doi.org/10.1007/978-3-030-29387-1

Staggers, N., Jennings, B. M., & Lasome, C. E. (2010). A usability assessment of AHLTA in ambulatory clinics at a military medical center. *Military medicine*, *175*(7), 518–524. https://doi.org/10.7205/milmed-d-09-00285

Thelen-Perry, S., Ved, R., & Ellimoottil, C. (2018). Evaluating the patient experience with urological video visits at an academic medical center. *mHealth*, *4*, 54. https://doi.org/10.21037/mhealth.2018.11.02

Yip, M. P., Chang, A. M., Chan, J., & MacKenzie, A. E. (2003). Development of the telemedicine satisfaction questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. *Journal of telemedicine and telecare*, *9*(1), 46–50. https://doi.org/10.1258/135763303321159693

XAMK. (n.d.). *Summary of OSSI-project*. https://www.xamk.fi/tutkimus-ja-kehitys/ossi/

Appendices

Appendix 1. Telehealth Usability Questionnaire

Etäteknologian käytettävyyskysely

Tässä tutkimuksessa selvitetään kokemuksia etäyhteyden välityksellä toteutetusta etävastaanotosta. Tehtäväsi on arvioida kokemuksiasi etäteknologian käytöstä palveluiden vastaanottamisessa arviointiskaaloilla, joita on tässä tutkimuksessa yhteensä n. 30 jaettuna kahdelle eri kyselysivulle. Lisäksi sinulta kysytään laadullista palautetta.

Arviointiskaalojen avulla voit arvioida kokemustasi etäteknologian käytettävyydestä ja kuormittavuudesta. Arviointeja tehdessäsi mieti jokainen väittämä erikseen, ennen kuin annat arvioinnin. Jos juuri sopivaa vastausvaihtoehtoa ei tunnu löytyvän, valitse vaihtoehto, joka on lähinnä omaa mielipidettäsi.

Tutkimuksen toteuttaa Kaakkois-Suomen ammattikorkeakoulun Mikkelin yksikkö. Vastaukset käsitellään nimettöminä ja luottamuksellisesti. Tutkimuksen tuloksia hyödynnetään etäteknologioiden käytön kehittämisessä sosiaali- ja terveysalalla. Tulokset raportoidaan siten, että vastaajien henkilöllisyyttä ei ole mahdollista tunnistaa tuloksista.

Ennen kuin aloitat, ole hyvä ja vastaa lyhyeen alla olevaan taustatietokyselyyn:

*Pakollinen

Esitiedot

1. **Oma ID-numero**

2. **Mitä järjestelmää arvioit?**
Merkitse vain yksi soikio.

- Software 1
- Software 2
- Software 3

Kokemukseni etäyhteysjärjestelmästä

Seuraavaksi arvioi etäyhteyden käyttöä etävastaanotossa ja sen vaikutuksia seuraavien väittämien avulla. Arvioinneissa "etäyhteysjärjestelmä" tarkoittaa juuri testaamasi videoneuvottelujärjestelmää. "Asiantuntija" puolestaan viittaa etävastaanottoa tarjonneeseen ammattilaiseen.

3. **Etäyhteysjärjestelmän käyttö oli yksinkertaista** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä						Olen täysin samaa mieltä

4. **Oli helppoa opetella etäyhteysjärjestelmän käyttö** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä						Olen täysin samaa mieltä

5. **Uskon saavani nopeasti parempaa palvelua tätä etäyhteysjärjestelmää käyttämällä** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

6. **Etäyhteysjärjestelmän käyttö tietokoneella oli miellyttävää** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

7. **Pidin etäyhteysjärjestelmän käyttämisestä** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

8. **Etäyhteysjärjestelmä oli yksinkertainen ja helppo ymmärtää** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

9. **Etäyhteysjärjestelmä teki kaiken sen minkä haluankin sen tekevän asiantuntijan tapaamisessa** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

10. **Mielestäni asiantuntijan kanssa oli helppo jutella etäyhteysjärjestelmää käyttäen** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

11. **Pystyin kuulemaan asiantuntijan puheen selvästi etäyhteysjärjestelmää käyttäessäni** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

12. **Tunsin pystyväni ilmaisemaan itseäni tehokkaasti etäyhteysjärjestelmää käyttäessäni** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	◯	◯	◯	◯	◯	Olen täysin samaa mieltä

13. **Etäyhteysjärjestelmää käyttäessä pystyin näkemään asiantuntijan aivan kuin olisimme tavanneet kasvotusten** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	◯	◯	◯	◯	◯	Olen täysin samaa mieltä

14. **Mielestäni tapaamiset tätä etäyhteysjärjestelmää käyttäen ovat samanlaisia kuin tapaamiset kasvotusten** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	◯	◯	◯	◯	◯	Olen täysin samaa mieltä

15. **Aina, kun tein virheen etäyhteysjärjestelmää käyttäessäni, pystyn korjaamaan sen helposti ja nopeasti (jos et tehnyt virheitä, älä vastaa tähän kysymykseen)**
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	◯	◯	◯	◯	◯	Olen täysin samaa mieltä

16. **Etäyhteysjärjestelmän antamat virheilmoitukset kertoivat selvästi, kuinka ongelmat voi korjata (jos et tehnyt virheitä, älä vastaa tähän kysymykseen)**
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	◯	◯	◯	◯	◯	Olen täysin samaa mieltä

17. **Tunnen oloni mukavaksi käyttäessäni etäyhteysjärjestelmää tapaamiseen asiantuntijan kanssa** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	◯	◯	◯	◯	◯	Olen täysin samaa mieltä

18. **Tämä etäyhteysjärjestelmä on soveltuva tapa vastaanottaa sosiaali- ja terveyspalveluita** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	◯	◯	◯	◯	◯	Olen täysin samaa mieltä

19. **Käyttäisin tätä etäyhteysjärjestelmän kautta järjestettäviä palveluja uudestaankin** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

20. **Olen kaiken kaikkiaan tyytyväinen tässä etäyhteydessä käyttämääni järjestelmään** *
Merkitse vain yksi soikio.

	1	2	3	4	5	
Olen täysin eri mieltä	⬭	⬭	⬭	⬭	⬭	Olen täysin samaa mieltä

Appendix 2. NASA-Task Load Index

Kuormittavuus

Seuraavaksi arvioi etäyhteyden käytön kuormittavuutta etävastaanottoa. Arvioinneissa "etäyhteysjärjestelmä" tarkoittaa juuri testaamaasi videoneuvottelujärjestelmää.

21. **Kuinka henkisesti vaativaa etäyhteysjärjestelmän käyttö oli?** *

Merkitse vain yksi soikio.

	1	2	3	4	5	
Erittäin vähän	◯	◯	◯	◯	◯	Erittäin paljon

22. **Kuinka fyysisesti vaativaa etäyhteysjärjestelmän käyttö oli?** *

Merkitse vain yksi soikio.

	1	2	3	4	5	
Erittäin vähän	◯	◯	◯	◯	◯	Erittäin paljon

23. **Kuinka kiireistä etäyhteysjärjestelmän käyttö oli?** *

Merkitse vain yksi soikio.

	1	2	3	4	5	
Erittäin vähän kiireiseltä	◯	◯	◯	◯	◯	Erittäin kiireiseltä

24. **Kuinka onnistuneesti pystyit tekemään sen, mitä sinun piti tehdä etäyhteysjärjestelmän kautta?** *

Merkitse vain yksi soikio.

	1	2	3	4	5	
Epäonnistuin	◯	◯	◯	◯	◯	Onnistuin

25. **Kuinka kovasti sinun täytyi yrittää pystyäksesi käyttämään tätä etäyhteysjärjestelmää?** *

Merkitse vain yksi soikio.

	1	2	3	4	5	
Erittäin vähän	◯	◯	◯	◯	◯	Erittäin paljon

26. **Kuinka epävarma, lannistunut, ärtynyt, stressaantunut tai vihainen olit?** *

Merkitse vain yksi soikio.

	1	2	3	4	5	
Erittäin vähän	◯	◯	◯	◯	◯	Erittäin paljon

Appendix 3. Preferences questionnaire

Loppukysely

Pisteytä testaamasi järjestelmät siten, että 1. on mieluisin vaihtoehto lääkärin etävastaanottoon osallistumiseen. Täytä jokainen kenttä siten, että jokaisessa valinnassa on eri pisteytys. Eli järjestä testaamasi järjestelmät omasta mielestäsi paremmuusjärjestykseen.

*Pakollinen

1. **Oma ID-numero**

2. **Aseta järjestelmät järjestykseen sen mukaan, mitä käyttäisit mieluiten lääkärin etävastaanottoon osallistumiseen** *

   Merkitse vain yksi soikio riviä kohden.

	1.	2.	3.
Software 1	◯	◯	◯
Software 2	◯	◯	◯
Software 3	◯	◯	◯

Appendix 4. Manuscript with Software 1

Vastaanoton käyttötapaus – Software 1

Katso ohje sähköpostistasi ja kirjaudu järjestelmään

Ammattilainen: Hei, miten voit?
Potilas: Hyvin
Ammattilainen: Tervetuloa etäyhteysvastaanotolle. Ennen kuin aloitamme, haluaisin että hiljennät mikrofonin, niin voin näyttää sinulle muutaman toiminnon mitä tällä etäyhteysjärjestelmällä voi tehdä. Kun olet hiljentänyt mikrofonin, sano jotain niin voin varmistaa, etten kuule sinua.

Säädä mikrofoni äänettömäksi

Potilas: Ok.
Ammattilainen: Voit ottaa mikrofonin takaisin käyttöön. Sanotko jotain, että kuulen että mikrofonisi on jälleen käytössä.

Laita mikrofoni takaisin päälle

Ammattilainen: Miten viikko on kulunut?
Potilas: Ihan hyvin, olen tarkkaillut verenpaineita ja elänyt terveellisesti.
Ammattilainen: Hieno homma, minkälaisia arvoja olet saanut, kun olet mittailut verenpainettasi?
Potilas: Tänä aamuna yläpaine oli 138 ja alapaine 72.
Ammattilainen: Mahtavaa! Mittasitko eilen myös verenpaineesi?
Potilas: Kyllä, mittasin verenpaineeni myös eilen aamulla.
Ammattilainen: Hienoa! Voisitko kertoa tulokset eilisestä verenpainemittauksestasi?
Potilas: Tottakai!

Kerro nyt ohjaajalta saamasi verenpainearvot lääkärille

Ammattilainen: Hienoa, että sinulla menee noin hyvin! Näyttää siltä, että olisi hyvä, jos näkisimme uudestaan kuukauden päästä.
Potilas: Se käy oikein hyvin!
Ammattilainen: Hieno juttu, olemme yhteydessä. Voit nyt katkaista puhelun. Heippa!

Katkaise puhelu ja sulje ikkuna
Testi loppui. Siirry vastaamaan käytettävyyskyselyyn.

Appendix 5. Manuscript with Software 2

Vastaanoton käyttötapaus – Software 2

Katso ohje sähköpostistasi ja kirjaudu järjestelmään

Ammattilainen: Hei, miten voit?
Potilas: Hyvin
Ammattilainen: Tervetuloa etäyhteysvastaanotolle. Aloitetaan siten, että aseta koko ruudun tila käyttöön niin voit nähdä minut hieman paremmin.

Avaa videoneuvottelu koko ruudun kokoiseksi

Potilas: Selvä.
Ammattilainen: Ennen kuin aloitamme, haluaisin että hiljennät mikrofonin, niin voin näyttää sinulle muutaman toiminnon mitä tällä etäyhteysjärjestelmällä voi tehdä. Kun olet hiljentänyt mikrofonin, sano jotain niin voin varmistaa, etten kuule sinua.

Säädä mikrofoni äänettömäksi

Potilas: Ok.
Ammattilainen: Voit ottaa mikrofonin takaisin käyttöön. Sanotko jotain, että kuulen että mikrofonisi on jälleen käytössä.

Laita mikrofoni takaisin päälle

Ammattilainen: Miten viikko on kulunut?
Potilas: Ihan hyvin, olen tarkkaillut verenpaineita ja elänyt terveellisesti.
Ammattilainen: Hieno homma, minkälaisia arvoja olet saanut, kun olet mittailut verenpainettasi?
Potilas: Tänä aamuna yläpaine oli 138 ja alapaine 72.
Ammattilainen: Mahtavaa! Mittasitko eilen myös verenpaineesi?
Potilas: Kyllä, mittasin verenpaineeni myös eilen aamulla.
Ammattilainen: Hienoa! Voisitko kirjoittaa tulokset eilisestä verenpainemittauksestasi chattilaatikkoon?
Potilas: Tottakai!

Kirjoita nyt ohjaajalta saamasi verenpainearvot chattilaatikkoon

Ammattilainen: Hienoa, että sinulla menee noin hyvin! Näyttää siltä, että olisi hyvä, jos näkisimme uudestaan kuukauden päästä.
Potilas: Se käy oikein hyvin!
Ammattilainen: Hieno juttu, olemme yhteydessä. Voit nyt katkaista puhelun. Heippa!

Katkaise puhelu ja sulje ikkuna
Testi loppui. Siirry vastaamaan käytettävyyskyselyyn.

Appendix 6. Manuscript with Software 3

Vastaanoton käyttötapaus – Software 3

Katso ohje sähköpostistasi ja kirjaudu järjestelmään

Ammattilainen: Hei, miten voit?
Potilas: Hyvin
Ammattilainen: Tervetuloa etäyhteysvastaanotolle. Aloitetaan siten, että aseta koko ruudun tila käyttöön niin voit nähdä minut hieman paremmin.

Avaa videoneuvottelu koko ruudun kokoiseksi

Potilas: Selvä.
Ammattilainen: Ennen kuin aloitamme, haluaisin että hiljennät mikrofonin, niin voin näyttää sinulle muutaman toiminnon mitä tällä etäyhteysjärjestelmällä voi tehdä. Kun olet hiljentänyt mikrofonin, sano jotain niin voin varmistaa, etten kuule sinua.

Säädä mikrofoni äänettömäksi

Potilas: Ok.
Ammattilainen: Voit ottaa mikrofonin takaisin käyttöön. Sanotko jotain, että kuulen että mikrofonisi on jälleen käytössä.

Laita mikrofoni takaisin päälle

Ammattilainen: Miten viikko on kulunut?
Potilas: Ihan hyvin, olen tarkkaillut verenpaineita ja elänyt terveellisesti.
Ammattilainen: Hieno homma, minkälaisia arvoja olet saanut, kun olet mittailut verenpainettasi?
Potilas: Tänä aamuna yläpaine oli 138 ja alapaine 72.
Ammattilainen: Mahtavaa! Mittasitko eilen myös verenpaineesi?
Potilas: Kyllä, mittasin verenpaineeni myös eilen aamulla.
Ammattilainen: Hienoa! Voisitko kertoa tulokset eilisestä verenpainemittauksestasi?
Potilas: Tottakai!

Kerro nyt ohjaajalta saamasi verenpainearvot lääkärille

Ammattilainen: Hienoa, että sinulla menee noin hyvin! Näyttää siltä, että olisi hyvä, jos näkisimme uudestaan kuukauden päästä.
Potilas: Se käy oikein hyvin!
Ammattilainen: Hieno juttu, olemme yhteydessä. Voit nyt katkaista puhelun. Heippa!

Katkaise puhelu ja sulje ikkuna
Testi loppui. Siirry vastaamaan käytettävyyskyselyyn.