Aalto-yliopisto
Aalto-universitetet
Aalto University

Master’s Programme in Computer, Communication and Information Sciences

Improving the usability of online symptom checkers to avoid the digital exclusion of vulnerable user groups

Felipe González Carceller

Master’s Thesis
2021

Aalto-yliopisto
Aalto-universitetet
Aalto University

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Felipe González Carceller

Title of thesis Improving the usability of online symptom checkers to avoid the digital exclusion of vulnerable user groups

Programme Computer, Communication and Information Sciences

Major Software and Service Engineering

Thesis supervisor PhD Sari Kujala

Thesis advisor DA Kaisa Savolainen

Date 08.08.2021 **Number of pages** 67 + 12 **Language** English

Abstract

This research aims to study how usable existing online symptom checkers are from the point of view of vulnerable user groups and what to consider when developing these services in the future. For this, usability tests were conducted on two of the most prevalent symptom checkers in Finland, Omaolo symptom checker and Klinik Access. A total of eight participants were recruited for this study, four people with mild intellectual disabilities and four older adults. These user groups were selected since they can be considered to be at risk of digital exclusion, which then may lead to social exclusion. This may happen as health services become progressively more digitized restricting some people's access to them. By developing usable services for people of all ages, abilities, and skill levels, everyone can benefit, as states the principle of universal design.

A combination of the think-aloud method, observations, questionnaires, and semi-structured interviews was used to better understand the presented problem. The methods used yielded various usability issues that hindered the participants' ability to fill in the symptom checkers in an efficient manner. The median System Usability Scale score of the Omaolo symptom checker was 72/100 and for Klinik Access it was 68/100.

Based on the gathered data and analyzed results, this study presents 13 guidelines that address issues found in the usability tests ranging from the insufficient visibility of relevant information to the clearness of the different options in the services. These guidelines are an effort to make online symptom checkers usable by people with the widest range of capabilities and therefore minimize service quality gaps between the different user groups, no matter their age or capabilities. Learning from the usability issues revealed in this study and applying this acquired information in the future will improve the chances to take every user into consideration and ultimately result in universally more usable online symptom checkers.

Keywords usability, universal design, symptom checkers, eHealth, vulnerable users

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Felipe González Carceller

Työn nimi Sähköisten oirearvioiden käytettävyyden parantaminen heikommassa asemassa olevien käyttäjien digitaalisen syrjäytymisen välttämiseksi

Koulutusohjelma Computer, Communications and Information Sciences

Pääaine Software and Service Engineering

Työn valvoja FT Sari Kujala

Työn ohjaaja TaT Kaisa Savolainen

Päivämäärä 08.08.2021 **Sivumäärä** 67 + 12 **Kieli** Englanti

Tiivistelmä

Tämän tutkimuksen tarkoituksena on tutkia, kuinka käytettäviä olemassa olevat sähköiset oirearviot ovat heikommassa asemassa olevien käyttäjäryhmien näkökulmasta ja mitä on pidettävä mielessä näitä palveluita kehitettäessä tulevaisuudessa. Tätä varten suoritettiin oirearvioiden käytettävyystestit Omaolo oirearviolla ja Klinik Accessilla. Tähän tutkimukseen rekrytoitiin yhteensä kahdeksan osallistujaa, neljä lievästi kehitysvammaista ja neljä ikäihmistä. Nämä käyttäjäryhmät valittiin, koska niiden voidaan katsoa olevan digitaalisen syrjäytymisen vaarassa, mikä voi sitten johtaa sosiaaliseen syrjäytymiseen. Näin voi käydä, kun terveydenhuoltopalvelut digitalisoituvat mikä saattaa rajoittaa joidenkin ihmisten pääsyä palveluihin. Kun palveluita voi käyttää iästä, kyvyistä tai taidoista riippumatta, jokainen käyttäjä hyötyy.

Jotta esitettyä ongelmaa voitaisiin ymmärtää paremmin, käytettävyystestin aikana tietoa kerättiin ääneen ajattelun, havaintojen, kyselylomakkeiden sekä haastatteluiden avulla. Käytetyt menetelmät toivat esille erilaisia käytettävyysongelmia, jotka estivät osallistujia täyttämästä oirearvioita tehokkaasti. System Usability Scale -kyselystä Omaolo oirearvio sai keskimäärin 72/100 pistettä ja Klinik Access 68/100 pistettä.

Kerätyn tiedon ja analysoitujen tulosten perusteella tässä tutkimuksessa esitetään 13 ohjetta, jotka käsittelevät käytettävyystesteissä löydettyjä käytettävyysongelmia kuten esimerkiksi relevantin tiedon puutteellista näkyvyyttä sekä eri vastausvaihtoehtojen selkeyttä. Näiden ohjeiden avulla pyritään tekemään sähköisistä oirearvioista mahdollisimman käytettäviä heikommassa asemassa oleville käyttäjille ja huomioimaan mahdollisimman moni käyttäjä, tarjoten kaikille käyttäjille saman käyttökokemuksen ja mahdollisuuden käyttää palveluita. Tässä tutkimuksessa löydetyistä käytettävyysongelmista voidaan ottaa oppia ja tätä opittua tietoa voidaan soveltaa tulevaisuudessa, antaen ihmisille mahdollisuuden käyttää verkkopalveluita itsenäisesti lisäten heikommassa asemassa olevien käyttäjien itsenäisyyttä ja oirearvioiden helppokäyttöisyyttä.

Avainsanat käytettävyys, universal design, oirearvio, eHealth, heikommassa asemassa olevat käyttäjät

Preface

I want to give a special thanks to my supervisor Sari Kujala as well as my advisor Kaisa Savolainen for mentoring me and allowing me to constantly learn throughout the writing of this thesis. They both helped me immensely during my time as a research assistant and I cannot thank them enough. I would also like to thank everyone at DigiIN for giving me invaluable feedback and guiding me in the right direction.

I would also like to thank people at DigiFinland, Klinik Healthcare Solutions, the Selkeästi meille -initiative, Väylä ry, and all the participants who made this study possible. This thesis could not have happened without you.

A big thank you goes also to my family as they supported me through my struggles and were always there for me while I was working on this thesis.

Espoo, 8 August 2021
Felipe González Carceller

Contents

1 Introduction

As the world becomes more and more digital, some services can be found exclusively online. A shift towards digital services is also happening in the healthcare field (Mehta et al., 2020). However, not all patients have the same opportunities or knowledge to adapt to the rapid digitalization of the modern world and this can lead to some unwanted or unaccounted consequences. Some groups of people might be unintentionally excluded and receive limited service or no service at all for something they have no control over.

Online symptom checkers are used by people seeking health related guidance as these services typically provide an urgency assessment and suggest a course of action based on the reported symptoms (Chambers et al., 2019). The user can then use this information to make better informed decisions regarding their health and decide whether they should seek medical attention as soon as possible or whether they can treat the symptoms on their own. Symptom checkers utilize algorithms to provide the user with an initial screening or triage to determine the most appropriate next step the user should take (Semigran et al., 2015). Depending on the symptom checker, entered information can also be relayed to a doctor to give them advanced understanding of the patient's condition before even meeting them. Furthermore, symptom checkers can empower users as they facilitate patient involvement in their health care (Meyer et al. 2020). This study focuses on the two most prevalent online symptom checkers in Finland, Omaolo symptom checker and Klinik Access.

Public services are typically aimed for a wider audience so it can sometimes be difficult to provide the same service for everybody. However, there are some steps that can be taken to get closer to this goal. Namely, the term universal design, also referred to as design for all, conceptualizes a design philosophy which aims to develop products and services that are usable to the greatest extent possible by all kinds of people regardless of their age as well as their abilities or disabilities (Story et al., 1998). This kind of approach benefits user groups that are vulnerable or in danger of digital exclusion which, then again, may lead to social exclusion (Martin et al., 2016). Universal design combats this issue by aiming to provide the users with the means to use the service with equal opportunities. Especially in the case of intellectually disabled individuals as well as older adults, creating a situation where they can use a service on their own without the feeling of being a burden to others can make them feel empowered and creates a positive user experience (Chan et al., 2013; Wass & Safari, 2020). Nevertheless, vulnerable users' needs might be neglected as studies suggest that vulnerable user groups are often not involved in the design process of eHealth applications (Ware et al., 2017; Oudshoorn et al., 2020).

Previous work has discussed the usability aspect of symptom checkers but little to no research has focused on vulnerable user groups such as older adults and intellectually disabled individuals (Kim, 2017; Knitza et al., 2021; Liu et al., 2021). The aim of this research is to propose a set of guidelines to address the design choices of public digital health services, and more specifically online symptom checkers, as an attempt to minimize the gap of service received by different groups of people. Thus, the research problem that will be answered is

“**How can the usability of digital health services be improved to avoid the digital exclusion of vulnerable user groups?**”. This research problem is further supported by two additional research questions:

1. How usable are current online symptom checkers for older adults and the mildly intellectually disabled?
2. How to design symptom checkers for all users?

These questions help to better understand the problem of designing public services for a large and heterogeneous user base. To answer these research questions, a literature review will be conducted and followed up by an empirical study. The literature review will present a summary of what has been accomplished in this field so far while the empirical part of the study will take the form of a usability test in order to identify what works in the current design of the services and where changes need to possibly take place.

The thesis is structured as follows. Section 1 introduces the topic of the thesis, the research problem along with the supporting research questions, and the scope of the study. Section 2 further introduces usability, universal design, and accessible services and their relevance in this study. This sections also present examples of previous studies in this field and what can be learned from them. Section 3 goes over the methodology of this study. This includes the research approach, an introduction to the tested services and the participants, the data collection methods and the reasonings why they were chosen, a step-by-step description of how each usability test was conducted, and how the collected data was analyzed. Then, Section 4 presents the results of the empirical study. An overview of the participants is given followed up by the results of the usability tests. Section 4 also proposes a set of guidelines that are based on the usability issues identified during the tests. Finally, Section 5 discusses the study as a whole by answering the research problem and research questions, going over the limitations of the study, and then proposing future directions for studies to follow.

2 Background

2.1 Usability

The Human-centered design for interactive systems' standard ISO 9241-11:2018 (International Organization for Standardization, 2018) defines usability as "the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". According to Bevan et al. (2015), usability has previously been used to measure the user's performance or how well they used a system, product, or service to solve tasks and reach goals, that is, its effectiveness and efficiency. The emphasis was more on the applicability of the system in a specific situation rather than on the user's experience when using the system, which is also an important aspect to consider states Bevan et al (2015). However, over the years usability has come to replace the term "user friendly" (Bevan et al., 1991; Nielsen, 1993). In a similar manner, Bevan et al. (2015) argue that there is more to usability evaluation than measuring just effectiveness, efficiency, and satisfaction, referring to Nielsen's (1993) definition of usability which consists of a total of five usability attributes adding learnability, memorability, and errors to the previously mentioned efficiency and satisfaction. Nielsen defines usability as a quality attribute that assesses how easy a user interface is to use. In other words, the easier a webpage is to use, the better its usability.

Following Nielsen's (1993) definition, learnability describes how easy it is for users to accomplish basic tasks the first time they interact with the system. For a system to be usable, it should be easy for a new user to learn and to understand (Shackel, 2009). In the case of a system being challenging to learn and to comprehend, users might be reluctant to accept it and to use it (Rizzo et al., 2005). By using words, phrases, and concepts familiar to the user, the user can associate interactions to past experiences and make connections which improve learnability (Nielsen, 1994a).

Efficiency measures how quickly and successfully users can perform tasks once they have learned the design and are familiar with the system (Nielsen, 1993). This can be evaluated by recording the time it takes a user to complete a high-level task once they have had time to get used to the system (Kim, 2017). Although the user may be able to perform the tasks efficiently, said tasks might not be effectively completed. Effectiveness measures the accuracy and completeness with which the user achieves a specified goal, where accuracy describes how well the actual outcome matches the intended outcome and completeness describes how many of the intended outcomes were actually achieved (International Organization for Standardization, 2018). Effectiveness can be evaluated by determining the task completion rate (Arain et al., 2016).

Memorability consists in the user's ability to interact with a system again and reestablish proficiency after a period of not using it (Nielsen, 1993). A human's information processing has a limited capacity and any additional memory used in tasks other than the one at hand may reduce resources for comprehension (Dalal et al., 2000). Making objects, actions, and options visible to the user and allowing the user to recognize rather than recall certain functions and interactions minimizes their memory load (Nielsen, 1994a).

Errors can and will happen when engaging with user interfaces. Good usability allows and also helps users to recognize, diagnose, and recover from errors (Nielsen, 2012). People like to feel they are in control and not bound by a system's limits. This leads to the need to provide users with an option to go back on their actions in the case of an unwanted reaction (Nielsen, 1994a). In other words, supporting undo is important. This said, small errors or slips can be prevented by constraining users' actions and by limiting their choices (Laubheimer, 2015a). The number of slips can also be reduced by providing users with predefined choices and presenting them with a ready list of possible interactions as opposed to have them form their own input (Laubheimer, 2015a). Eliminating or checking for error-prone conditions and presenting users with a confirmation option before they commit to an action is part of a well-designed system. Similarly, small unwanted mistakes can be prevented applying thoughtful and easy to understand design such as providing users with a preview before applying major changes or before performing destructive actions (Laubheimer, 2015b).

Satisfaction considers how pleasant a device or a system is to use and what emotions it evokes in the user (Nielsen, 1993). According to Jordan (1998), pleasure and the feeling of satisfaction when using a product play a big role in usability when interacting with a service. The happier the user is while using the system, the easier it is to sustain the user's attention and interest in using the system which then facilitates any ongoing learning process.

2.1.1 Universal design

Since the goal is to make services usable by everyone, the principle of universal design is adopted. Universal design strives to design products, environments, and services to be usable by all people and to the greatest extent possible, without the need for adaptation or specialized design (Mace, 1988). In other words, the designs insist on being universally applicable to all people of all ages, abilities, and skill levels. A service should be designed to meet the needs of all people who wish to use it (National Disability Authority, 2021). This does not mean that a special requirement is met to benefit only a minority of the population but instead if the service is accessible, usable, convenient, and a pleasure to use, everyone will benefit (Centre for inclusive Design, 2020). Supporting this, a study conducted by Harrington et al. (2017) suggests that designing non-stigmatizing products and services for users who have been shown to have the most difficulties in learning and adopting new technologies, will conceivably result in products and services that are usable and therefore adopted by a wider audience. For example, in the physical world there are lowered curbs at crosswalks to provide easy passage for people in wheelchairs. Moreover, these design choices also benefit other people using the crosswalk such as people riding bikes or parents pushing their child's stroller. A similar approach can be taken in the digital world. By lowering the curb for people with limited capabilities, we assure everyone's inclusion (Hesse, 1995). The aim is to create a structured environment to be equally accessible by all users, regardless of experience level or physical and mental ability (Hesse & Shneiderman, 2007).

Universal design consists of seven principles established by The Center of Universal Design (Connell et al., 1997). These principles, although initially crafted to be applied in architecture, aim to educate all designers as well as consumers about the characteristics of more usable products, environments, and services (Story et al., 1998). The principles, and the guidelines within them, served as inspiration and were kept in mind during the planning of the empirical study and the evaluation of the test results. Table 1 consists of each of universal design's seven principles, their definition, and a list of key elements that ought to be present in a design that complies with the principle.

Table 1. Principles of universal design (Connell et al., 1997).

Principle	Guidelines
1. Equitable Use. The design is useful and marketable to people with diverse abilities.	**1a.** Provide the same means of use for all users: identical whenever possible; equivalent when not. **1b.** Avoid segregating or stigmatizing any users. **1c.** Provisions for privacy, security, and safety should be equally available to all users. **1d.** Make the design appealing to all users.
2. Flexibility in Use. The design accommodates a wide range of individual preferences and abilities.	**2a.** Provide choice in methods of use. **2b.** Accommodate right- or left-handed access and use. **2c.** Facilitate the user's accuracy and precision. **2d.** Provide adaptability to the user's pace.
3. Simple and Intuitive. Use of the design is easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level.	**3a.** Eliminate unnecessary complexity. **3b.** Be consistent with user expectations and intuition. **3c.** Accommodate a wide range of literacy and language skills. **3d.** Arrange information consistent with its importance. **3e.** Provide effective prompting and feedback during and after task completion.
4. Perceptible Information. The design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities.	**4a.** Use different modes (pictorial, verbal, tactile) for redundant presentation of essential information. **4b.** Provide adequate contrast between essential information and its surroundings. **4c.** Maximize "legibility" of essential information. **4d.** Differentiate elements in ways that can be described (i.e., make it easy to give instructions or directions). **4e.** Provide compatibility with a variety of

	techniques or devices used by people with sensory limitations.
5. Tolerance for Error. The design minimizes hazards and the adverse consequences of accidental or unintended actions.	**5a.** Arrange elements to minimize hazards and errors: most used elements, most accessible; hazardous elements eliminated, isolated, or shielded. **5b.** Provide warnings of hazards and errors. **5c.** Provide fail safe features. **5d.** Discourage unconscious action in tasks that require vigilance.
6. Low Physical Effort. The design can be used efficiently and comfortably and with a minimum of fatigue.	**6a.** Allow user to maintain a neutral body position. **6b.** Use reasonable operating forces. **6c.** Minimize repetitive actions. **6d.** Minimize sustained physical effort.
7. Size and Space for Approach and Use. Appropriate size and space is provided for approach, reach, manipulation, and use regardless of user's body size, posture, or mobility.	**7a.** Provide a clear line of sight to important elements for any seated or standing user. **7b.** Make reach to all components comfortable for any seated or standing user. **7c.** Accommodate variations in hand and grip size. **7d.** Provide adequate space for the use of assistive devices or personal assistance.

2.1.2 Universal design in eHealth

Universal design is particularly important in the eHealth sector as it can make services more inclusive and usable for vulnerable user groups (Kadir & Jamaludin, 2013; Ruzic & Sanfod, 2017; Irish, 2020). That being said, usability is perhaps the biggest barrier stopping older adults from fully embracing the use of digital channels when it comes to healthcare (Zapata et al., 2015). Bad usability of a system can also lead to users not adopting the use of the service or abandoning it completely (Greenhalgh et al., 2017; Harrington et al., 2017). One solution to this issue is to involve users in the design process. User involvement results in more accurate user requirements and in improved user satisfaction (Kujala, 2003). Similarly, studies by Raviselvam et al. (2016) and Gkouskos & Burgos (2017) show that universal design has the potential to empower users and that supporting user participation can improve the overall usability of products and services. Ruzic & Sanfod (2017) concluded that even if a set of guidelines are intended to promote universal design, they still need to be validated by applying them and evaluating them with users with a wide range of abilities. However, since the two user groups involved in this study do not represent the average user of an eHealth application, their role and involvement in the design process may be minimal or nonexistent. Oudshoorn et al. (2020) suggest that the majority of studies do not take into consideration the preferences or needs of people with mild intellectual disabilities when it comes to eHealth applications. This inconsideration applies to older adults as well (Ware et al., 2017).

2.2 Accessible services for vulnerable users

While universal design focuses on designing products, environments, and services that are usable by the widest range of users operating in varying situations, accessibility strives to make products, environments, and services usable by people with disabilities as well as older adults (Petrie & Bevan, 2009). One of the groups affected by the digitization of health services is the intellectually disabled (Lussier-Desrochers et al., 2017). Intellectually disabled individuals have been observed to encounter significantly more difficulties with finding information on the internet as well as understanding information from online sources compared to people from the general population (Alfredsson Ågren et al., 2020). Moreover, it has been estimated that around 1% of the general population has some degree of an intellectual disability (Patel et al., 2020). In Finland, this ratio is roughly the same, accumulating to nearly 50 000 Finns being classified as intellectually disabled (Westerinen, 2018; Kehitysvammaliitto, 2021; Tukiliitto, 2021). Studies show that the majority of people with an intellectual disability have a mild intellectual disability, meaning that they may have limitations when it comes to conceptual development and daily living skills but they can still do basic self-care and function in ordinary life with minimal support (National Academics of Sciences, Engineering, and Medicine, 2015; Patel et al., 2018). Developing more accessible services would allow intellectually disabled individuals to find information concerning their health more easily (Alfredsson Ågren et al., 2020).

Both of the previously mentioned user groups, the mildly intellectually disabled and older adults, can be considered as vulnerable user groups as Vines et al. (2013) state that people can be vulnerable in a number of ways, for example, physically, cognitively, and socially. Existing literature shows that these two user groups face, among other things, cognitive challenges when interacting with technology (Marston et al., 2019; Setchell et al., 2021). Consequently, studies suggest that not being able to use digital technology, or in other words being excluded to an extent due to the lack of digital capabilities, can in some cases lead to social exclusion as well (Martin et al., 2016; Seifert et al., 2018; Seifert et al., 2020).

Furthermore, people with intellectual disabilities tend to have problems in the field of language meaning that they have a reduced understanding of speech as well as a reduced ability to communicate (Parnell, 2014). Such limitations impact the person’s access to information and therefore lead to a lack of health-related knowledge (Larson et al., 2005). Moreover, this makes it difficult for people with intellectual disabilities to make assessments and decisions regarding their health needs (Feldman et al., 2015; Shogren, 2017). This is why some intellectually disabled people rely on caregivers when it comes to health-related decision-making (Geukes et al., 2019). Nevertheless, eHealth has the potential to offer opportunities to support people with mild intellectual disabilities in different contexts of their daily lives (Oudshoorn et al., 2020). Giving people with intellectual disabilities the opportunity to make their own decisions about daily matters, makes them feel more empowered and in control of their own life (Wass & Safari, 2020).

Studies show that people with intellectual disabilities can benefit from eHealth applications in the same way as the rest of the population (Raspa et al., 2018; Watfern et al., 2019). However, this requires some additional aspects to be considered in order to facilitate the users' adaptation to digital healthcare systems. The Matching Person & Technology model (Scherer & Craddock, 2002) is used to assist people in selecting the most appropriate support solution for a certain individual. According to this model, there are three main areas that need to be assessed for eHealth to be effective: 1) the environmental factors influencing use, 2) the consumer's needs and preferences, and 3) the functions and features of the eHealth application. Following this, Oudshoorn et al. (2020) lists three main aspects related to using eHealth in order to support people with mild intellectual disabilities in performing daily activities.

First, Oudshoorn et al. (2020) conclude that the people around the user have a significant impact on how the user sees and adapts the system. Their research shows that people around a mildly intellectually disabled individual are rarely involved in the process of selecting and adapting the use of an eHealth application. People with mild intellectual disabilities can have difficulties with generalizing and applying their learned skills to a new context and they need support throughout their life (Thompson et al., 2009). Therefore, it is essential to receive as much support as possible from key stakeholders, whether they are family members or hired support staff.

Second, Oudshoorn and her colleagues' (2020) results show that when working with intellectually disabled individuals, studies often neglect the user's needs and preferences. This lack of personalization impedes users from exploring their preferences and finding suitable solutions. Using the needs and preferences of mildly intellectually disabled users as a starting point can help find the most appropriate eHealth applications that cater to the user's needs bringing more value to the user. Each user is widely different and hence it is essential to offer customization. Studies regarding assistive technologies and eHealth emphasize the importance of personalization and its impact on the user's ability to fulfil their personal goals (Collins & Collet-Klingenberg, 2018; Boot et al., 2018; Frielink et al., 2020).

Third, Oudshoorn et al. (2020) imply that many eHealth applications can be implemented following the conventions of universal design. Developing eHealth applications with the guidelines of universal design in mind increases the likelihood of users adapting the use of the systems in their daily life and it allows intellectually disabled people to benefit from eHealth systems in the same way as the rest of the population, given that they are involved in the design process from the very beginning (Raspa et al., 2018; Watfern et al., 2019). This universal ease of use can make services easier to use even for people who do not use digital means often or are not that familiar with technology (Gassmann & Reepmeyer, 2008). Furthermore, a broader application of accessibility may benefit everyone, not only people with disabilities.

3 Research methods

The research method used for this study was a usability test as testing with real users helps uncover problems in the design and learn about the users' behavior and preferences. In the usability test, people from two different user groups, older adults and people with mild cognitive disabilities, acted as test users and performed an evaluation with the help of standardized clinical vignettes on the two most prevalent symptom checkers in Finland, Omaolo symptom checker (Omaolo, 2021) and Klinik Access (Klinik Healthcare Solutions, 2021). The following sections describe the research approach and introduce the services being tested and the participants involved in this study. Then, the usability evaluation methods are presented after which the data analysis is described.

3.1 Research approach

This study was conducted as part of the DigiIN project, a project that strives towards a socially inclusive digital society (DigiIN, 2021). It does this by creating practical solutions which ensure that the social welfare and healthcare sector's digital services are available and accessible by everyone and by making information more accessible to decision-makers and any other interested party. The DigiIN project is a cooperation between the Finnish Institute for Health and Welfare (THL), Aalto University, University of Helsinki, University of Jyväskylä, Laurea University of Applied Sciences, and Age Institute, aiming to prevent the marginalization of vulnerable users as digital services become more prominent in society thus requiring a more active role from the users themselves.

The study was conducted as a case study where the usability of two online symptom checkers was evaluated in order to make the services more accessible to everyone. A case study helps to understand a real-world scenario as well as take into account the specific conditions of the case (Yin, 2002). Case study research utilizes selective sampling instead of random sampling in order to achieve the most accurate and representative results regarding the research problem and the focused user group (Easterbrook et al., 2008).

Nielsen (1993) argues that usability testing with real users is irreplaceable, since it provides direct information about how people use the services and gives insight to their exact problems with the interface being tested. The usability test is a qualitative test which helps identify problems in the design of the service, uncovers opportunities to improve, and gives insight to the users' behavior and preferences (Moran, 2019). In other words, the goal of usability testing is to improve the usability of the service being tested by uncovering its problems. According to Dumas & Redish (1999), by observing the performance as well as the comments of the participant while they complete the usability test, valuable information is gathered about the current usability of the service. Dumas & Redish argue that asking the opinion of the participant is also an important part of usability testing as this helps understand how the user sees their experience with the service.

Previous usability testing on eHealth applications has included data collection methods such as questionnaires, task completion, interviews, and the think-aloud method to name a few listed by Maramba et al. (2019). In this study, data was collected using a combination of the think-aloud method, observations, a questionnaire, and a semi-structured interview. In addition, standardized clinical vignettes were used as they provide some common scenarios as to why people visit health centers. They also allow the classification of symptoms based on their severity and make the results comparable with previous studies.

3.2 Online symptom checkers

Online symptom checkers are digital health services that allow users to seek health related guidance (Chambers et al., 2019). They have existed for a number of years already and became even more prominent during 2020 due to the COVID-19 pandemic as a way to facilitate and help healthcare customers to self-diagnose and self-triage without the need of physically going to the health center (Aboueid et al., 2021a). These artificial intelligence-enhanced tools provide users with an initial diagnosis and triage based on the symptoms they are experiencing or displaying (Tsai et al., 2021). Symptom checkers may also be used to get suggestions of what are the next steps the user should take and they may offer some self-care options in cases when the displayed symptoms are not too grave (Semigran et al., 2015). Furthermore, some use them to self-diagnose before going to the health center while others double check the symptoms afterwards to have a second opinion and to get further information about the doctor's diagnosis (Meyer et al., 2020). Typically, symptom checkers present a series of questions where the user is asked about their symptoms. These answers are then processed by the service's artificial intelligence, making each diagnosis personal to that specific user and their case. As an additional benefit, symptom checkers have the potential to decrease the number of unnecessary medical visits (Chambers et al., 2019; Kujala et al., 2020). The two most prevalent symptom checkers in Finland are Omaolo symptom checker (Omaolo, 2021) and Klinik Access (Klinik Healthcare Solutions, 2021) and they are the two services tested in this study.

Omaolo is a digital service and service channel for healthcare, which supports self-service in personal care and directs patients to appropriate assistance when necessary (DigiFinland, 2020). The Omaolo symptom checker is composed of 16 specialized symptom checkers, each specializing in a different part of the body, ranging from headaches to back pain. Each of the symptom checkers asks the user a set of questions and then, based on the user's answers, suggests next steps the user should follow to address their issue. The user may also enter their home municipality in which case the service will display the corresponding city's recommendations and refer the user to the nearest health center if the symptoms require it.

Much like Omaolo, Klinik Access is widely used in Finland. It allows users to take care of their health concerns online by answering to a set of questions presented by the symptom checker. Similar to the Omaolo symptom checker, Klinik Access first asks the user to locate the ailment on the body. Once the user has chosen the location, more specific questions are presented to better understand the possible illness. In case the user cannot pinpoint the symptom's exact location, they may also choose a corresponding option stating no specific location. The entered information acts as a preliminary questionnaire which the user's doctor can access, giving them time to familiarize themselves with the case before meeting their patient. If the artificial intelligence-driven service deems the entered symptoms or their combination serious, the user is urged to contact a health center for immediate medical attention, assuring that the patients are directed safely and efficiently to the right point of care (Klinik Healthcare Solutions, 2021). This helps relief the stress placed on the general practitioners taking care of the patients.

This study was done in collaboration with the people behind Omaolo and Klinik Access. They provided demo environments where the tests could be conducted ensuring that the fictitious data entered into the systems would interfere with the systems' function as little as possible. After the usability tests' results were analyzed, each of the service providers received their own detailed report and presentation where they were able to present questions regarding the participants' interactions and experiences that they were more curious about.

3.3 Participants

The goal was to recruit at least eight participants for the study. Given the scope of the study as well as the COVID-19 pandemic, around eight test participants was seen as a reasonable number to provide an overview of the research problem and highlight the main usability issues within the systems. Sinkkonen et al. (2006) argue that a usability test should include three to six participants after which most of the major usability flaws are discovered. Nielsen's (1994b) study also shows that with just five participants, around 80% of main concerns and problems can be found and with eight participants, like in this study, around 90% of the of the most prevalent issues would be brought to attention when incorporating the think-aloud method.

As the focus of the study was on people at risk of digital exclusion, the participants recruited for the study consisted of a balanced mix of mildly intellectually disabled individuals and older adults, four from each group respectively. Two of the four intellectually disabled test participants were recruited from the Selkeästi meille-initiative, a joint project between The Finnish Association on Intellectual and Developmental Disabilities (FAIDD) and Kehitysvammatuki 57 ry. The other two intellectually disabled participants were recruited from Väylä ry, a non-profit organization that aims to employ intellectually disabled individuals and provide proper compensation to them for their work. Meetings were held with the two establishments to discuss about any special arrangements the test participants might require in order to make the test environment more suitable for them and more representative of a real-world scenario. This included assistive technology and other details that might have had to be taken into consideration. The test's older adults were recruited by utilizing personal contacts. They also had to satisfy the requirement of being at least 75 years old with knowledge of how to use a computer, making them potential users of the services being tested.

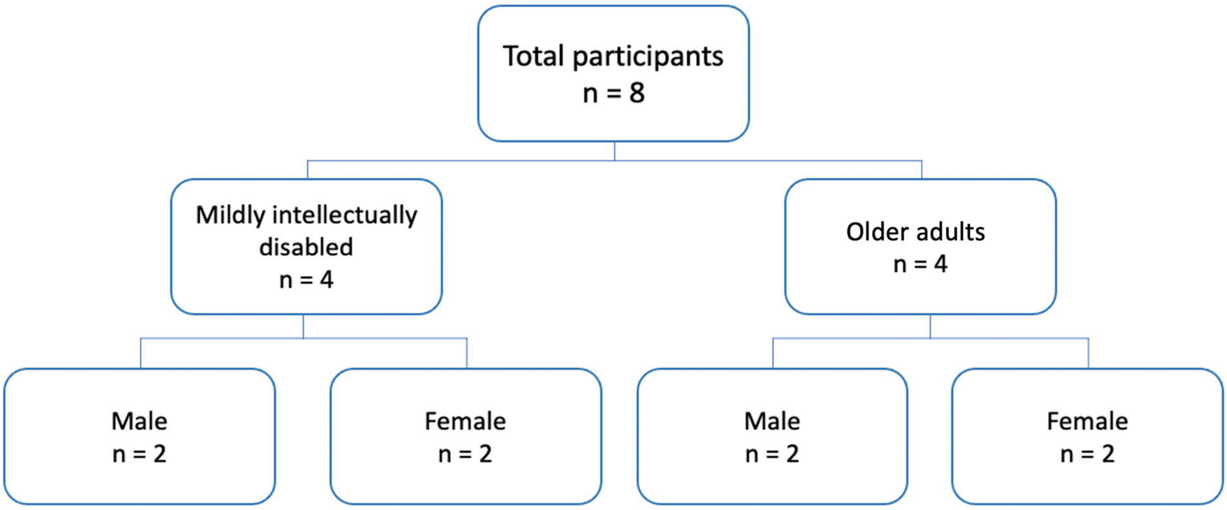


Figure 1. Distribution of participants.

3.4 Usability evaluation

A variety of data collection methods and their combination were used to better understand the research problem. Each method has its advantages and collecting data from different perspectives using multiple methods helps offset the limitations of each method leading to a more comprehensive understanding of the research problem (Easterbrook et al., 2008). The methods used for data collecting were the think-aloud method, observations, a questionnaire, and a semi-structured interview.

3.4.1 Thinking aloud

Participants were asked to say out loud what they were thinking during the tests regarding the usability of the services and the overall user experience allowing for the recording of the participants' thoughts and opinions. They were encouraged to voice even the minute problems they encountered while conducting the usability test. The think-aloud method gives insight to how the participant sees and interprets things as well as the reason why they see things in a certain way (Van Den Haak et al., 2003). The method also highlights issues in the moment so the observations made by the participant are recorded even if the participant forgets them later on. Since the usability tests were conducted remotely, the think-aloud method is especially useful as non-verbal cues might be hard to detect over the videocall (Vasalou et al., 2004).

3.4.2 Observation

Observation was used as a data collecting method during the usability test to perceive how the participants interact with the services and to catch behavior patterns that the participants might be unaware of. The test facilitator who observed the participants' actions and reactions live during the tests and then another time when watching the recording of the test a day or two after the test had taken place took notes of any encountered issues. The facilitator also kept track of the number of successfully completed tasks, number of errors committed, amount of assistance required to successfully complete the tasks, and the time it took to complete them. During the observation, the facilitator did not interfere with the test unless the participant asked for assistance or was unable to proceed on their own in which case the facilitator helped as was needed in order to allow the participant to continue with their test.

3.4.3 Questionnaires

Three questionnaires were used in this study to collect data. These questionnaires helped to get a more rounded picture of the participants and facilitate the comparison of the results as all of the participants are asked the exact same questions in the same order. First, the participants were asked to fill in a questionnaire regarding their background information before the usability test (see Appendix A). The background information consisted of the participant's gender, age, the number of times they had visited a doctor during the last two years, how many medical conditions diagnosed by a doctor they have, how many times they have previously used digital healthcare services, and how often they use digital devices such as smart phones or computers. These questions substantiated that the participants fit the study's target demographic concerning their age and ability to use an electronic device such as a computer on their own.

In addition to the background questionnaire, the participants were also asked to fill in a health literacy survey in order to establish the participants' health literacy level. In this case, health literacy is described as the ability of a person to gain access, understand, and then apply resources to make decisions regarding their own health (Sørensen et al., 2013). Morse et al. (2020) argue that health literacy is something that should be taken into consideration when evaluating symptom checkers and that it is a metric that can be used to measure whether using a symptom checker augments the understanding and management of the user's illnesses. An assessment about the health literacy of the participant also results in a more accurate description of the participant as a user of the service (Aboueid et al., 2021b). The first 16 questions of the complete European Health Literacy Questionnaire, HLS-EU-Q47, address health care while the rest of the questions focus more on disease prevention and health promotion (Jovanić et al., 2018). Therefore, a 16-question variation of the survey was used instead of the complete questionnaire with 47 items as it was seen a more suitable choice given the scope of this study (see Appendix B). This 16-question health literacy survey, HLS-EU-Q16, is a 4-point Likert scale where the answering options range from "very difficult” to “very easy” and has also the option of “I don’t know” which, for the purpose of collecting data, is interpreted as a missing answer (Eronen et al., 2019). The 16-question variant has been shown to be a feasible option in a number of different countries as well as when studies have included vulnerable user groups (Storms et al., 2017; Eronen et al., 2019; Lorini et al., 2019; Bas-Sarmiento et al., 2020). The health literacy survey used was based on the Finnish translation by Eronen et al. (2019).

After each of the usability tests, the participants were given a System Usability Scale (SUS) questionnaire to measure the perceived usability of the system (see Appendix C). The SUS questionnaire contains 10 questions on a 5-point Likert scale where the answering options range from “Strongly agree” to “Strongly disagree” (Brooke, 1996). The scored answers result in a numeric value that can be used to roughly compare the usability of different systems or alternatively, it can be used to detect any improvement in the usability of a newer version of the same system. According to Brooke (1996), this quick and quite effortless way of assessing the services’ usability alternates between positive and negative questions in order to avoid response bias as the participants need to think about the questions and cannot simply mark the same answer for each question presented. Other versions of the SUS were also taken into consideration in case the participants needed a cognitively easier and less straining version of the questionnaire. Holden’s (2020) simplified SUS, aimed at the cognitively impaired and older adults, as well as The Usability Metric for User Experience Lite (UMUX-Lite) proposed by Lewis et al. (2013), which contains just two positive questions using a 7-point Likert scale, were amongst the ones considered. However, these were discarded after some consulting with the mildly intellectually disabled participants’ employers and the older adults’ family members when it was assured that the people participating in the usability test would have no issues with the original and slightly lengthier SUS.

3.4.4 Semi-structured interview

Since the usability test was conducted on a small scale, the focus of data collection was mostly on gathering qualitative data rather than quantitative. Interviews fit this purpose as they are useful for understanding a certain group of people’s needs, problems, ideas, and expectations (Stickdorn et al., 2018). Similarly, interviews can help gather more ideas and bring up points that might not have been addressed in the structured part of the study. There are many types of interviews but the one seen as the most suitable for this study was the semi-structured approach as it offers a mixture of more specific questions, that were prepared prior to the interview, as well as open-ended questions, that address issues the participants might have encountered during the test. Semi-structured interviews are designed to obtain foreseen information along with possible unexpected types of information (Seaman, 1999). They also give the participant the opportunity to express themselves more freely and to further clarify why they liked or did not like certain design choices and interactions encountered while performing the tests. In order to ease the participant into the interview, they were simply asked what they felt about the test and then they were asked to comment on the services they had just tested, if they were easy to use, or difficult perhaps, and if so, if they could elaborate on their remarks. Easing into the more structured part of the interview gave the participants time to gather their thoughts and get more comfortable to discuss about their experience. The questions for the structured part of this study's interview are partly influenced by the requests of the service providers and their interests and can be found in Table 2.

Table 2. Interview's questions.

Question	
1	Would you use the service again in the future?
2	Was the summary and the instructions of what to do next clear enough?
3	Would you actually follow the instructions given?
4	Given the option, would you use the service using your phone?

3.4.5 Symptom vignettes

To assist the usability test and to avoid the need for participants to enter their personal medical information into the services, each participant was given two standardized clinical vignettes depicting ready symptoms. A total of six conditions with varying severity levels were selected from a list composed by Semigran et al. (2015). Conditions with different severity levels were chosen since people reportedly use symptom checkers in both urgent and non-urgent situations (North et al., 2012). Following the suggestions made by Semigran et al. (2015), the vignettes chosen were from three categories of triage urgency: conditions which require emergent care, conditions which require non-emergent care, and conditions where a medical visit is seen unnecessary and which can be treated with self-care. The selected conditions were translated from English to Finnish and simplified in order to make them more understandable and relatable by the participants. The different symptom vignettes used, their urgency, a more detailed description of which vignette was given to which participant, and the order in which participants tested the two symptom checkers can be found in Appendix D. During the tests, the participants were only given the symptom description, the correct diagnosis and the symptoms' urgency were omitted in order to influence the outcome as little as possible.

3.5 Study setup

The test sessions were held in the Microsoft Teams video conferencing platform which allows screen sharing, screen recording, and voice recording. This made interacting with the participants easy and created an opportunity to observe the participant and their actions while they conducted the usability test. The test sessions were primarily one-on-one meetings between the test participants and the test facilitator but all participants except for one had a support person physically present who then helped with setting up the test environment and provided technical assistance when needed. All the test sessions were conducted in Finnish as it was the primary language of all participants.

The participants were given the option of a face-to-face meeting where the test facilitator would be physically present when conducting the study in case the participant could not participate online. However, due to the COVID-19 pandemic, online means were preferred in order to minimize the risk of exposure of anyone involved in the study. The decision depended on each of the participants individually but eventually all the participants decided to follow the state-wide social distancing recommendations and opted to participate virtually.

In the beginning of a test session, the test facilitator would start by introducing himself and properly explaining to the participant what the purpose of the study is and why does the participant’s contribution matter. When the participant was first contacted, they were given a short introduction to the study and told about the expected length of the session but now the facilitator has the opportunity to explain things in more detail and answer any questions the participant might have at this point. The questionnaires concerning the participant’s background information and health literacy, as well as a consent form to cover the legal and ethical aspects of the study, were also sent before the remote test session in order for the participants to have time to fill them in at their own pace and as an effort to not make the test session too long for the participant.

The facilitator would then explain how the collected data will be used and stored and for how long. After this, the participant was informed that 1,5 hours had been reserved for the test so they could advance at their own pace and that they could stop the test at any time if they wished to do so. The facilitator would then provide the participant with the prepared vignettes and explain their role in the test. Each participant was given two vignettes with different severities, one that requires urgent care and one with milder symptoms. A more detailed assignment of vignettes can be seen in Appendix D. The participants were also told they could come up with their own answers in case they were asked questions that had no straight answer in the provided vignettes. In addition to this, the participants were instructed to imagine a world without the COVID-19 pandemic so any questions related to COVID-19 were to be skipped and ignored.

After this, the facilitator would set the scenario by asking the participant to picture themselves in a situation where they are not feeling well and experiencing the symptoms described in the vignette. The participant’s goal was to find out what was wrong with their health and what they should do next. The participant would try to conduct the test to the best of their abilities and first try independently but could ask for help if clarification was needed or if they got stuck and did not know how to proceed. Once the participant familiarized themselves with the provided material the test could start.

When the participant was ready to begin the test, the facilitator explained that the session would be recorded and, once the participant understood and approved this, the facilitator turned on the voice and video recording. At this point, the participant was asked to share their screen so that the facilitator could follow the test being conducted and so that the screen recording would capture all the necessary data. By following the provided instructions, the participant then navigated to one of the services' webpage on their computer and start entering data according to one of the given vignettes. The facilitator kept track of which service the participant tested first and switch the services' order for the next participant in order to avoid learning bias. While performing the test, the participant was also asked to think aloud and express verbally what they were experiencing when using the service and the reason behind their actions.

Once the participant had completed all the steps of the usability test but before the interview and further discussion, they were given the SUS questionnaire to fill in, as instructed by Brooke (1993). The facilitator was also ready to ask the aforementioned questionnaire's questions verbally and fill in the responses on the participant's behalf in case filling the questionnaire turned out to be a difficult task but all the participants were capable of filling it in on their own. After the questionnaire, an initial interview was conducted to understand the participant's thoughts while the experience was still fresh in memory. When the participant was ready to continue, they were instructed to return to the landing page of the service and asked to use the service again but this time using the second vignette as a list of symptoms. Once they were done, all of the previously mentioned steps were repeated on the second service.

After both services had been tested, a semi-structured interview was conducted. The participant was also asked to compare the two services and express their pros and cons respectively. When the discussion had reached its end, the participant was thanked and told they could be in contact with the facilitator if additional feedback came to mind after the tests. Then, if the participant had no more questions, the video call and screen recording was ended.

3.6 Pilot study

Before involving actual participants, a small-scale pilot study following all the steps just described was carried out to find any inconsistencies in the questions and assure that the instructions were understandable and easy to follow. Studies have shown that this results in more accurate test results and gives validity to the questions asked (Van Teijlingen & Hundley, 2001; Thabane et al., 2010). The questionnaires and interview questions were peer reviewed by other researchers as well as people involved with intellectually disabled individuals and a step-by-step execution of the study was conducted with relatives representing the older adults -user group. Based on the received feedback, some inconsistencies were identified and improvements, such as adapting plain language, were made in order to make the instructions as well as the questions more readable and understandable.

3.7 Data analysis

Upon completion of all the usability tests, the recordings from the test sessions were transcribed word for word to facilitate the coding of the gathered material. Coding helps categorize, sort, and give meaning to the gathered data (Williams & Moser, 2019). While going through the recordings, notes were also taken regarding the participants' actions and behavior. In addition, the transcriptions were further supplemented by the facilitator's notes taken during the observation part of the test sessions. These notes consist of the usability issues mentioned by the participants while thinking aloud as well as the ones observed by the facilitator and they were written down by the facilitator with corresponding timestamps to make locating them easier if there is a need to go back to them. With the timestamp, the test recording can be opened at the right moment to inspect the circumstances leading to the issues in more detail.

Qualitative analysis was done manually via the open coding method. First, each comment and observation about the services' usability was highlighted from the transcriptions and notes. Then, as new incidents were recorded, they were compared to existing incidents in order to find similarities as well as differences as suggested by Corbin & Strauss (2008). If multiple incidents were found to be conceptually similar, they were grouped together under a high-level and descriptive concept.

4 Results

The following sections present the results from the usability test. Based on the results, a set of guidelines are proposed along with examples of how they address issues that occurred during the usability tests. Throughout the following sections, the Omaolo symptom checker service will be referred to as Service A and the Klinik Access symptom checker will be referred to as Service B.

4.1 Participant characteristics

A total of eight people participated in the study, four representatives from each user group. The mildly intellectually disabled participants were between 20 and 35 years of age and the older adults who participated in the study were 75-77 years old. In both groups there was an equal number of male and female participants.

Based on the background information collected during the study, the participants had visited the doctor three to five times within the last two years on average. All the participants also answered that they use digital devices such as computers, tablets, and smart phones on a daily basis, satisfying the requirement of having enough knowledge to operate a computer. This suggests that the participants are familiar with how webpages work and have already established expectations of how to navigate on a webpage. A more detailed description of the participants can be found in Table 3. Around half of the participants had some previous experience on using digital health services while the rest reported that they had no experience. This information was omitted from the table below in order to increase the anonymity of the participants so they could not be easily identified. Some of the other answers were generalized for the same reason.

Table 3. Background information of the participants.

Participant	Gender (Male/Female)	Age (Years)	Number of visits to the doctor in the past 2 years	Number of diagnosed medical conditions	How often uses digital devices?
P1	M	25-29	0-5	0-1	Multiple times a day
P2	M	35-39	0-5	0-1	Multiple times a day
P3	F	20-24	0-5	2-3	Multiple times a day
P4	F	30-34	0-5	0-1	Multiple times a day

P5	M	75-79	0-5	2-3	Multiple times a day
P6	F	75-79	0-5	2-3	Multiple times a day
P7	M	75-79	0-5	0-1	Multiple times a day
P8	F	75-79	10+	4-5	Multiple times a day

Each participant completed a health literacy survey to give a more rounded depiction of the people who participated in the study. The health literacy survey used was the 16-question European Health Literacy Survey Questionnaire, HLS-EU-Q16. Only the scores of participants who answered at least 80% of the questions, which in this case meant 13 out of the 16 questions, were calculated, as recommended by Sørensen et al. (2015). All the participants satisfied this requirement so no participant's results were excluded from the study at this point. The questionnaire's answers were scored by going through the individual answers and giving one point if the answer was "fairly easy" or "very easy", and giving them zero points if the answer was "fairly difficult" or "very difficult". This resulted in a scale from 0 points being the lowest to 16 points being the highest. The scores were then compared to the three levels of health literacy: inadequate (0-8), problematic (9-12), and adequate (13-16).

From the health literacy levels depicted in Figure 2 and the questions in the survey, it is clear that the majority of the people who participated in the study have a low health literacy. They struggle with obtaining information regarding their health in the context of everyday life, processing said information, and making well-informed decisions based on their understanding of it. Making digital health services accessible and easy to use for these users could have a positive effect on their health literacy levels and provide additional assistance that was not previously available to them.

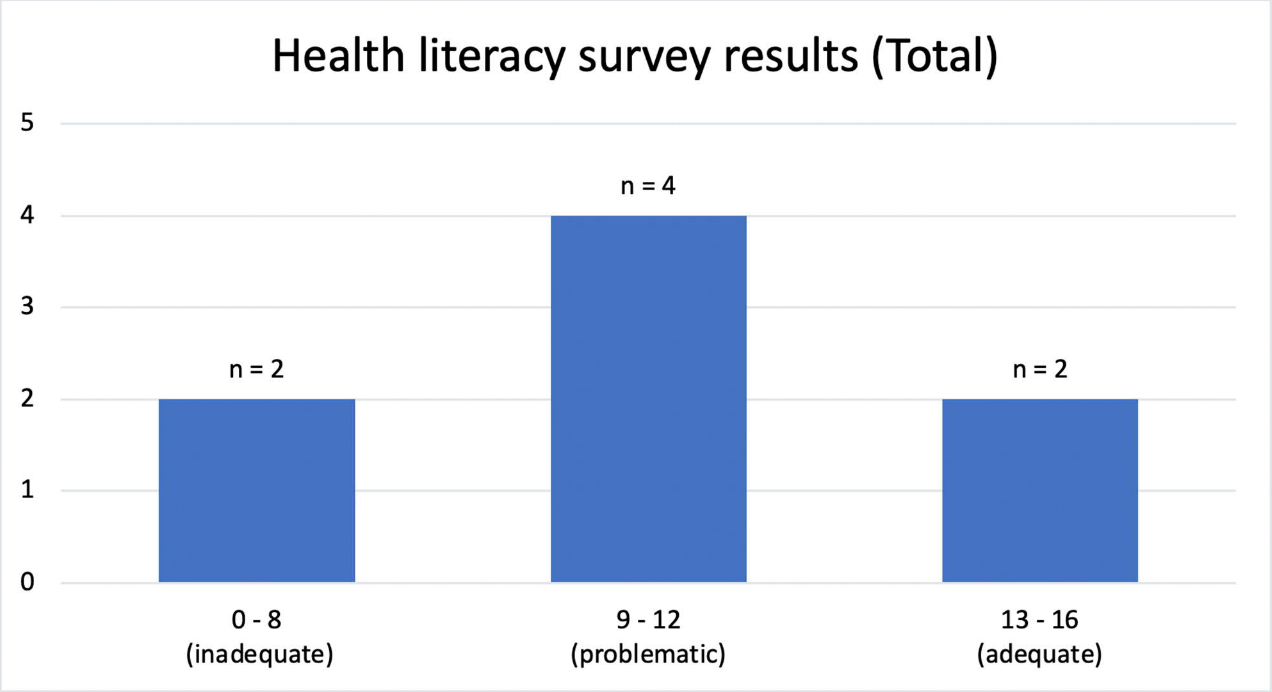


Figure 2. Health literacy scores of all participants and the distribution between the different health literacy levels.

4.2 Usability test findings and results

All 8 participants managed to successfully complete the tests at least once within the time frame of 1.5 hours that was allocated for each usability test. For the majority of the tests the participants were able to perform the given tasks with no additional assistance although there were a couple of instances where the participants struggled and were not able to proceed forward or go back to a previous state. In these cases, the participants requested help from the facilitator who was constantly present and who then helped them overcome the obstacles that they were facing. The following subsections discuss the results and main findings that emerged during the usability tests, based on the different data collection methods.

4.2.1 Usability issues

The test recordings including the participants thinking out loud were transcribed and the transcriptions were supplemented with notes taken by the test facilitator while he observed the tests. The notes taken during the observation include issues encountered by the participant but which the participant did not necessarily pay attention to, mainly because the issue went unnoticed by them. For example, when a view expanded as a consequence of a participant's action, some answering options were pushed out of sight. In this case, the participant assumed they were done with all the questions since the remaining answering options were no longer visible and the participant decided to proceed to the next view. Issues such as this which happened during the usability test were caught by the observer as he had a better understanding of the service and had more time to familiarize himself with the symptom checkers being tested.

The transcripts and observation notes were then coded which resulted in a total of 52 different usability issues and irregularities identified across the two services that were tested. These usability issues negatively affected the participants' abilities to navigate the services' webpages, fill in the symptom checkers according to the provided symptom vignettes, learn about what is causing their health problems, and decide what to do next. After the individual usability issues were identified, they were categorized into 13 high-level groups. These groups have been described in Table 4 along with how many participants encountered issues related to each group. A more detailed description of the individual usability issues identified can be found in Appendix E.

Table 4. Categorization of usability issues and irregularities.

Category	**Number of individual findings**	**Number of participants who encountered the issue (out of 8 participants)**
Visibility of information	6	8
Clarity of answering options	11	8
Guiding the user's focus	3	6
Long, uncommon, and compound words	5	4
Input methods	4	3
Allowing the user to revert their actions	3	2
Consistency of external links	3	3
Intuitive and consistent interactions	6	7
Allowing the user to express themselves freely	2	7
Informing the user of errors	3	3
Icons, symbols, and abbreviations	2	3
Navigation	3	5
Responsiveness	1	1

In some cases, the usability issues caused participants to enter differing answers and ultimately receive different recommendations from the services even though the participants were given the same vignettes and therefore should have answered the questions in a similar manner. The differences were caused by mainly two aspects. First, some answering options were not chosen due to their insufficient visibility. And second, the other cause was due to the instructions given to the participants which stated that they could come up with their own answers when asked about something that was not in the provided vignette. The differing answers were not given much weight in this study as assessing the accuracy of the symptom checkers was not part of the scope, although it is something interesting and worth looking into in future studies.

4.2.2 Task completion times

The usability issues identified in Table 4 hindered the participants' ability to finish the given tasks, meaning filling in the symptom checkers, in an efficient manner. Some issues had a bigger impact than others. Most notably, navigation and figuring out where to find the correct options had the biggest influence on the task completion times. Table 5 summarizes the average mean time it took the participants to use each symptom checker the first time they interacted with them. The time was started when the instructions were read and understood and the participant was guided to the front page of the symptom checker and then stopped when the symptom checker was completed and the participant was content with the end result. Each symptom checker was tested twice if the time allowed for it to measure the learnability of the systems. During the second test the participants were given a different set of symptoms to use so the path taken would not be exactly the same. The order in which the participants tested the symptom checkers was also alternated to avoid possible learning bias, meaning that half of the participants tested Service A first and the other half tested Service B first before switching services.

Table 5. Task completion times for Service A and Service B the first time the services were tested.

User group	Service A	Service B
Mildly intellectually disabled MEAN	14.9 min	16 min
Older adults MEAN	14.5 min	8 min
TOTAL MEAN	**14.7 min**	**12 min**

In addition to the task completion times depicted in Table 5, six participants had time to test Service A for a second time, three participants from each user group. The average completion time for the mildly intellectually disabled participants during Service A's second test was 7.3 minutes and for the older adults 7.5 minutes, bringing the average of all participants to 7.4 minutes. For Service B, five participants had time to test the service for a second time. Two mildly intellectually participants averaged a task completion time of 7.8 minutes and the three older adults averaged 7.5 minutes, bringing the average of all participants who tested Service B for the second time to 7.6 minutes.

The participants who had time to test services two times, finished the tasks considerably faster on their second try. This is to be expected given that during the second test they were more familiar with the service and had more knowledge about where certain options are located on the webpage. The second attempt also yielded less usability issues compared to the first time the participants tried the services. Service A's times show that both user groups took on average half as long to complete the task on their second try. Participants from both groups experienced difficulties in navigating the Service A's webpage and in finding the beginning of the symptom checker. One participant took 15 minutes just to find the symptom checker's first question. Service A also contains a lot of text a user has to read before the beginning of the questionnaire and some participants mentioned in the interview that they skipped most, if not all, of the text on their second time. The participants said they assumed the introduction and instructions were the same as the first time so they did not read them thoroughly again. This might happen in a real situation as well where a more experienced user who is already familiar with the service can shortcut and only pays attention to the parts that will help them achieve their goal with the service whether it is to read the diagnosis of their symptoms or book an appointment for example. Similar learning was noticed in Service B as the time it took for the mildly intellectually disabled participants to finish the task during the second try was halved as well. In this case, the participants credited the swiftness of the second try to the fact that they knew better what to expect from the service and were mentally prepared to process all the presented information.

There were also some external factors which might have affected the reported task completion times. This was a fictitious situation in which the participants were given a piece of paper with the symptoms listed on it. Participants were then instructed to come up with answers to the symptom checkers' questions if what was asked of them was not on the provided vignette. This presumably might have affected the task completion times in the sense that participants spent time reading the provided vignettes and coming up with their own answers as opposed to a real situation where a user knows what symptoms they are experiencing, in which case it might be easier for them to answer questions about their condition. In other words, the task completion times in Table 5 should be used as a reference point rather than accurate values representing the actual time it takes a user to fill in a symptom checker. Nevertheless, the recorded times of Service A are comparable to the recorded times of Service B and both services took on average as much time during the second test.

4.2.3 Perceived usability

The SUS questionnaire was given to the participants to fill in right after they had tested one of the services for the first time. According to Brooke (1996), the SUS questionnaire provides a "quick and dirty" way of measuring a system's perceived usability on a scale from 0 to 100, covering the perceived effectiveness, efficiency, and satisfaction of said system as stated in the ISO definition of usability. These scores can be used to compare the usability of different systems which might otherwise be difficult to compare given that the perceived usability might vary from user to user.

The participants had no problems in answering the SUS questionnaire's 10 questions on their own. The questions were clear to the participants and the answering options providing a Likert scale from 1 to 5, where 1 was "Strongly disagree" and 5 "Strongly agree", were an understandable way of inputting data. The answers were then scored as suggested by Brooke (1996). Each question, or item, was given a score contribution and then each score contribution was added and their sum was multiplied by 2.5. The score of the items 1, 3, 5, 7, and 9, the positively worded items, is the scale point minus 1, and the score of the items 2, 4, 6, 8, and 10, the negatively worded items, was calculated as 5 minus the scale position. The end result is a score between 0 and 100. However, this score does not represent a percentage result but instead it gives more of a reference point and a quick way to compare two different systems. The average score, or the 50^{th} percentile is around 68 points meaning that a score over 68 points is above average and a score lower than 68 points is below average (Brooke, 2013). The mean SUS results for each service are depicted in Table 6.

Table 6. SUS score means and standard deviations for Service A and Service B.

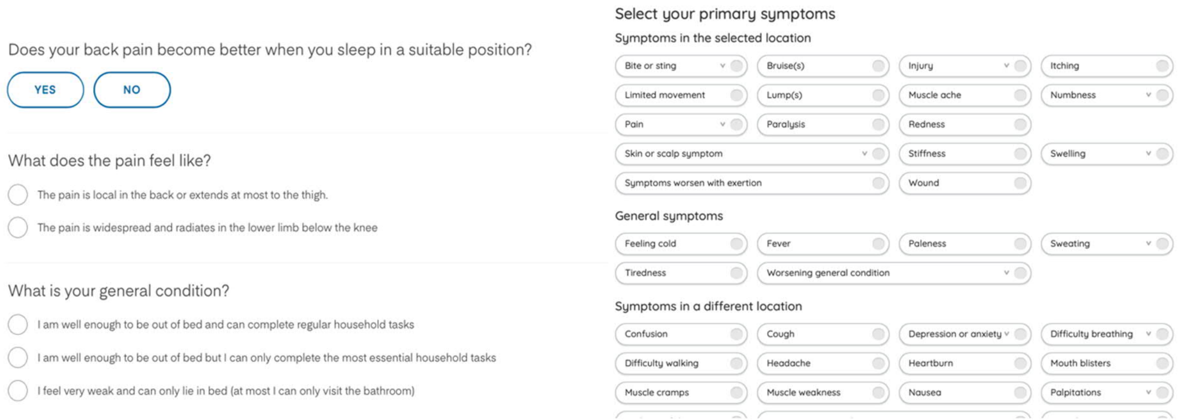
User group	**Service A**	**Service B**
Mildly intellectually disabled MEAN	73.8	65.0
Mildly intellectually disabled SD	± 20.7	± 13.8
Older adults MEAN	70.6	70.6
Older adults SD	± 7.4	± 19.6
TOTAL MEAN	**72.2**	**67.8**
TOTAL SD	**± 15.9**	**± 17.2**

The mean score for Service A was 72 and for Service B it was 68. These scores are just short of what Bangor et al. (2009) would say is a good score after having analyzed around 3 500 SUS questionnaires over a decade. However, the scores are still categorized as acceptable according to Bangor et al. The SUS scores in Table 6 indicate that the older adults rated both services on average the same. On the other hand, a notable difference can be seen in the scores given by the mildly intellectually disabled participants. They rated Service A 14% higher than Service B, meaning that they perceived Service A to have a better usability than

Service B. This being said, there was quite a difference in scores even within the user groups. For example, the highest and lowest evaluation given by the mildly intellectually disabled for Service A differ by a fair amount given that the highest score was a perfect 100 while the lowest score was merely a 45. So, while one participant rated the service perfect, another participant encountered mayor issues with the usability, a reflected in the suboptimal score. A similar swing in scores was found in Service B's SUS results where the older adults' scores were 45 at the lowest and 100 at the highest. Such large ranges hint to the fact that while some design choices are preferred by some users, others might have completely differing likings or might not find the design choices as familiar and intuitive to use, making it challenging for the services' designers to suit everyone's preferences.

4.2.4 Participants' views on the symptom checkers

The participants had varying opinions about the services. Based on the interviews, five out of the eight total participants said they preferred Service A over Service B. These statements are supported by the SUS scores which were also in favor of Service A by a small margin. The interviews revealed that the main reason for this was the way in which the questions and answering options were presented. The participants who preferred Service A also found it cognitively easier to use as the questions within the symptom checker were presented one at a time and the answering options were clearly stated out. These participants described Service B's answering options as cluttered and somewhat messy in the way they were presented. In particular, participants with mild intellectual disabilities stated that they found it challenging to focus when reading through all the answering options because it made their eyes wander around the webpage jumping from one option to another which made them unintentionally skip over some answering options and in some cases led to an incomplete response and possibly the wrong diagnosis.



Does your back pain become better when you sleep in a suitable position?

What does the pain feel like?

○ The pain is local in the back or extends at most to the thigh.
○ The pain is widespread and radiates in the lower limb below the knee

What is your general condition?

○ I am well enough to be out of bed and can complete regular household tasks
○ I am well enough to be out of bed but I can only complete the most essential household tasks
○ I feel very weak and can only lie in bed (at most I can only visit the bathroom)

Select your primary symptoms

Symptoms in the selected location

General symptoms

Symptoms in a different location

Figure 3. Side-by-side view of how the questions and answering options are displayed to the user in Service A (left) and Service B (right).

While the presentation style of the questions favored one symptom checker over the other, many participants mentioned that it was quite challenging to find the first question in Service A. The participants stated that Service B went straight to the point and its progression was easy to follow but that navigation in Service A was not as effortless. Many participants spent a considerable time familiarizing themselves with the webpage and some complained that there was too much to read and go through even before locating the symptom checker's first question. The mildly intellectually disabled reported that they felt mentally tired and almost exhausted by the amount of text they had to read in the introduction and instructions. The older adults did not have much of an issue with the amount of text. They found the navigation and technical aspects more challenging as some of them had issues with remembering that a webpage can typically be scrolled up and down to reveal more information. They mentioned that the information and options that were hidden and not readily visible on the webpage, were at times non-existent to them since they did not always remember to scroll down the webpage to see whether there was more content to be revealed outside of their view.

The free-form line of questioning led to both positive feedback as well as some improvement suggestions made by the participants based on their perceived usability of the tested services. The positive feedback revolved mostly around the symptom checkers' function and provided value as most of the participants had no previous experience with online symptom checkers. All of the participants liked the concept of symptom checkers and said they were happy to know that services such as these exist. They appreciated the convenience of being able to handle these situations from their home as well as the advice of what to do next. The participants felt like they were being heard and taken care of in cases where the symptoms were severe and the symptom checker instructed the participant to contact urgent care as soon as possible. This also gave some participants the impression that using these services would give them a faster response than using the more conventional ways of calling the health center or physically going there. They felt getting an initial opinion or diagnosis on the matter can be beneficial, especially when it is difficult for the patients themselves to assess the severity of their displayed symptoms and they might not know whether to seek medical attention right away or not.

Some other comments and topics the participants themselves brought up during the interview were about the cognitive load and the lack of sufficient human presence. According to the participants, the services should cognitively strain the user as little as possible because one only uses these services when they are ill or have something wrong with them and in those situations, it might be challenging to fill in questionnaires and read a lot of text. To assist this, a couple of older participants wished to have some sort of teaching about the services. They mentioned there are many electronic health applications that they need to use occasionally so being able to follow instructions or even practice in a demo environment can increase their knowledge so they can confidently operate these systems on their own, without needing to rely on someone else to do it for them. Another potential solution that came up during the interviews was to present the user with some sort of a roadmap when the user lands on the webpage. According to one participant, this would help the user to outline how the system will proceed and what the webpage consists of. This would help them know what is coming and what to expect from the service, making its use more effortless and less cognitively straining.

Some participants mentioned that they felt a lack of human presence in the process. Especially the older adults would have liked to know that a real human has seen their request and is attending to their needs. They also mentioned that this would increase the services' trustworthiness. The participants did not question the symptom checkers' trustworthiness during the tests but some did voice their concerns about online health applications in general. After asking them to further elaborate, they said that hearing advertisements regarding these services or perhaps having a family member or a healthcare professional refer them to these sites would increase their trust. Being able to trust a service, especially services handling sensitive data such as medical information, allows the user to utilize the service with a peaceful mind and they dare fill in the requested information.

In addition to the freely flowing conversation, a set of specifically structured questions were also asked as proposed by the service providers involved in the study. These questions are listed in Table 7 and relate to the usefulness and possible future directions of the symptom checkers.

Table 7. Summary of the interview's questions and answers.

Question	**Summarized answers**
Would you use the service again in the future?	8 participants said yes.
Was the summary and the instructions of what to do next clear enough?	4 participants said yes, 4 participants said no, there was also a 50-50 distribution between the two user groups.
Would you actually follow the instructions given?	8 participants said yes.
Given the option, would you use the service using your phone?	2 participants said yes, 6 participants said no

The interviews revealed that the end result and instructions of what to do next were not as clear to the participants as they could be. Some mentioned that they had not even noticed that the symptom checkers offered instructions and advice of what the user should do next based on the symptoms they reported. The participants who did notice it suggested that the section could be emphasized a bit more as it is an important part of the service and the main reason why they would use the symptom checker in the first place. The participants were also asked if they would like to use the symptom checkers on their phone but the majority said they preferred a computer. The participants who were against the idea were mainly older adults and their unanimous reasoning was their weakened eyesight. They opted for the computer as computers typically have a bigger screen making the buttons and the overall text bigger and more accessible. Some added that mobile versions of webpages tend to be somewhat different than their desktop counterparts meaning that the user would need to learn the webpage layout and navigation tricks for a second time.

4.3 Guidelines

The objective of this study was to propose some guidelines in an attempt to improve the usability of online symptom checkers and thus make the services more available to a wide range of users. Based on the information gathered and analyzed, a total of 13 guidelines are proposed. These guidelines are derived from the summary of the usability issues and irregularities presented in Table 4 and supplemented with corroborating findings collected using multiple data collection methods. The following sections present each proposed guideline in greater detail while providing some examples encountered in the conducted usability test. Appendix E details the complete list of usability issues found, the service in which they appeared, and the number of participants who encountered the issues.

1. **Keep any relevant information visible at all times.** The webpage's main functionality and options should be visible immediately. "Hiding" information should be avoided as sometimes the user does not notice that the webpage continues "outside the screen" and hence does not realize they have to scroll up or down or open collapsed buttons to reveal more information. The user should not be expected to remember everything they have read.

   For example, in Service A, most of the participants clicked on the only visible button on the front page and ended up working on the wrong symptom checker. They started to do the COVID-19 symptom checker even though they had been instructed to skip anything related to COVID-19. The participants did not notice that they could scroll down on the webpage to reveal other symptom checkers as in Service A the user has to first select the kind of symptom checker they want to fill in before being presented with the appropriate questions. The participants also did not know which symptom checker they had chosen after landing on the symptom checker's page. In Service B on the other hand, when the user is asked to write on a text field and further describe their symptoms, the detailing questions about the symptom can be found in the text field as placeholders. When the users click on the text field and start answering, the questions disappear and out of the four presented questions, none of the participants answered to more than two of them, most of the participants only gave an answer to the first question since it was the only one they remembered after they started typing.

2. **The options provided to the user should be clear and understandable**. The user should be aware of the possible actions they can take at any given moment and they should also be able to tell the difference between the options.

   For example, in Service A some participants did not notice that the links to the different symptom checkers on the front page were actual clickable links and therefore did not know how to advance and start one of the symptom checkers. The links could be made to look more like selectable options or the text above the list could instruct the user to "select one option from the list below" for example. Secondly, in Service B many participants found the order of the symptoms and other answering options quite confusing and messy. They felt like there was a lot to process at once and that they had to read a lot in each of the steps. This made the participants' eyes wander a lot and jump inconsistently from option to option, which led to completely missing and ignoring some alternatives which they were supposed to choose. When asked if they had noticed that the choices were in alphabetical order, all the participants said that they had not noticed it.

3. **Make it clear to the user where they should be focusing at the given moment**. Whenever there is a lot of information, it is beneficial to the user if they are indicated what to focus on. The relation between the user's actions and the webpage's response should also be clear to the user as they might not notice that clicking something opened or changed something elsewhere. By showing the user's current location also makes them feel more in control if, for example, they were moved to a different view as a consequence of their click.

   For instance, in Service B many participants did not notice that by selecting some symptoms, additional options would be displayed so that they could choose a more detailed explanation of their symptom. These additional dialogues were mostly ignored as the participants did not see connection between their selection and the newly presented options. Closing the dialogue also shifted the view and some participants were not sure anymore if their selection was recorded properly.

4. **Avoid long or uncommon words and difficult compound words**. In the Finnish language, compound words are widely used and can place an additional burden on the user. Using a simpler and more common vocabulary, or opting to use plain language can make reading the text easier and more fluent and does not strain the user as much.

   Some participants struggled to tell similarly sounding words apart and some of the words were mixed and confused with other words. This was observed in both services. Depending on the situation and the word in question, the participants chose a wrong answer because they failed to read and understand the word in question properly, for example, when the words sounded similar to others like "heikotus/heikkous" and "kuume/kuumotus" (translations: "feeling weak/weakness", ”fever/warmth").

Similarly, both services contained certain words that were not familiar to the mildly intellectually disabled participants and they asked for help in interpreting them, for example, “käypä” and “pääasialliset” (translations: “current”, “primary”). Furthermore, one participant was dyslectic and found challenging to read and understand some of the longer compound words, such as, “laaja-alaisena”, “alaraajat”, “hengitysvaikeuksia”, and “leuanalustassa” (translations: “widespread”, “lower limbs”, “trouble breathing”, “under the jaw”).

5. **Offer simple input methods**. Keeping the input methods simple improves the readability and learnability of the system, can reduce the number of errors a user makes, and lessens their cognitive load.

   For example, in Service B some participants were confused when choosing the duration of the symptoms and therefore clicked on the wrong unit (three hours instead of days or weeks). One participant even took out a calculator and calculated three times 24 hours as the participant had first chosen hours as the unit and the symptom duration given was three days. The participant was not sure how to change their answer from hours to days so they calculated the value on their calculator and inputted 72 hours into the symptom checker instead of changing the unit to days and giving their answer as three days.

6. **Allow the user to revert their actions and return to the previous page or state**. This makes the user feel more in control of the system which makes for a better user experience. Users will make mistakes and the system should support recovering from them.

   For example, in Service A one participant chose the wrong symptom checker and did not know how to return back to the list of symptom checkers since there was no button on the webpage with this option. Similarly, in Service B one participant did not know how to remove their selection once they realized that they had chosen a wrong option.

7. **Make sure that external links also follow the original webpage’s conventions**. Links should convey to the user that they are in fact links and they should also state where the user is taken when the link is clicked. If a webpage follows a certain convention or guidelines, for example it is easy to read, one should attempt to link webpages that are also easy to read whenever possible. The user expects consistency and might experience excessive stress if faced with something completely new without being mentally prepared for it.

   For example, in Service A after completing the whole symptom checker and inputting the location information, one participant clicked on the link that should take them to the opening hours and contact information of the nearest health center. However, the webpage contained a lot more information that was promised and the information the participant was looking for took a while to find since they had to scroll down and look for it for a considerable amount of time. This information was also not very visible on the linked webpage. In Service A, there was an instance where a couple of participants clicked on the “Duodecim symptom checker” link on the symptom checker webpage thinking that the link would take them to the intended symptom checker. However, this is not the case as said link only provides some additional information about the symptom checkers of Service A. The participants spent some time looking for the symptom checker on the other webpage and when their patience started to run out, they were guided back to the original webpage by the test’s facilitator and were explained that the link did not take them where they thought it did.

8. **The pages and the interactions in them should be intuitive and logical**. The user has expectations of how webpages generally work. Meeting these expectations makes the user experience more pleasant and lessens the user’s cognitive load. The webpage should also be consistent. If something worked in a certain way previously, the user will expect it to work the same way again.

   For example, in Service A after one participant had identified themselves with a test bank account, they thought that they had booked an appointment with a doctor because, as the participant put it, “why else would you have to identify yourself”. The participant got the impression and assumed that once they identified themselves and logged in on the webpage, they would automatically get booked an appointment with a healthcare professional. However, this was not the case and there was no appointment booked at that point. Meanwhile in Service B, choosing certain symptoms opens an additional dialogue providing the user with more options to choose from but the button that closes this dialogue has the word “Done” on it which suggests that the section has been completed and that clicking it would advance the user to the next phase. However, clicking the button simply closes the subsection which is not the interaction the participants were expecting. Only one participant actually closed the opened subsection by clicking the button, others left it open and navigated around it.

9. **Allow the user to express themselves with their own words**. Providing the user with the option of a free text field and allowing them to tell about everything else that did not come up during the structured symptom checker gives the user a better sensation of being heard and empowers them.

   In both services, almost all the participants would have liked the option of a text field where they could list and describe other symptoms and things that they experiencing that they were not asked in the questionnaire such as preexisting medical conditions. Some of the participants were left wondering whether the rest of their symptoms were not relevant or whether they needed to complete the symptom checker again but this time with the symptoms that were not included in the first questionnaire. In cases where the participants missed a couple of questions during the questionnaire, providing them with an additional opportunity would have allowed them to include everything they thought was relevant.

10. **Tell the user when errors occur and how to recover from them**. Error messages should be clear and explain to the user what caused them and what the user could do to solve them. In a way, "holding the user's hand", especially in unexpected and possibly stressful situations, makes for a better user experience and increases the user's trust in the system as this may provide the feeling that they will not be left facing the issue alone.

    For example, in Service A some participants had not answered all the questions as requested by the symptom checker but they still tried to proceed by clicking the "Continue" button. This prompted an error text saying "Please answer all mandatory questions". However, the error text is not clearly visible on the screen and appears to be cut off if the user does not scroll down enough. This caused a couple of situations where the error message was displayed but the participants did not notice it and hence did not know why clicking the "Continue" button had no visible effect and did not allow them to proceed. Furthermore, one participant encountered an unexpected error in Service A and the webpage only displayed the text "Something went wrong". There was no explanation of what had happened or how to recover from the encountered error if the user wanted to continue using the webpage.

11. **Use icons, symbols, and abbreviations with consideration and give them explanations**. Not everyone interprets symbols and icons the same way and sometimes using them may confuse or mislead the user more than provide assistance. Giving symbols, icons, and abbreviations an explanation of what they mean in the present context brings clarity to the user and improves understandability.

    For example, one participant was not familiar with the abbreviation "tms." (meaning: "etc.") that was used in Service A. In Service B, some participants did not understand that the red asterisks after the final questions meant that the section was mandatory to fill in. Instead, they got confused why they could not proceed since to the participants there was no visible reason stopping them from moving forward.

12. **Navigating the webpage should be easy and effortless**. The information and different options the user can choose should not require many clicks to reach. Excessive scrolling should also be minimized as the user may easily miss information that requires scrolling.

    In Service A, some older adults clicked and dragged the scroll bar with their mouse in order to scroll down the webpage. Since new questions kept appearing underneath the question the participants had just answered, the participants were required to repeatedly move the mouse back and forth between the scroll bar on the right side of the webpage and the questions in the middle of the webpage. This movement was repeated after each question. One way to simplify this, could be to move the webpage automatically after a question has been answered.

13. **Webpages should be responsive**. Sometimes people can be impatient, so when they click on a button, the webpage should respond immediately and give visual feedback informing the user that the interaction was recorded. If the response takes longer than expected to process, an indicator should be displayed telling the user that it will take a moment. If the system does not react to the user's actions visibly, the user will get frustrated and think that maybe they are the problem. Keeping the user informed of what the webpage is doing gives them a sense of being in control of the situation and that is reflected on the user experience.

Table 8. Summary of the proposed guidelines.

#	**Guidelines**
1	Keep any relevant information visible at all times.
2	The options provided to the user should be clear and understandable.
3	Make it clear to the user where they should be focusing at the given moment.
4	Avoid long or uncommon words and difficult compound words.
5	Offer simple input methods.
6	Allow the user to revert their actions and return to the previous page or state.
7	Make sure that the external links follow the original webpage's standards
8	The pages and the interactions in them should be intuitive and logical.
9	Allow the user to express themselves with their own words.
10	Tell the user when errors occur and how to recover from them.
11	Use icons, symbols, and abbreviations with consideration and give them explanations.
12	Navigating the webpage should be easy and effortless.
13	Webpages should be responsive.

5 Discussion

5.1 Answering the research questions

The aim of this research was to evaluate the usability of currently available online symptom checkers from the perspective of vulnerable user groups and then, based on a usability test, present some guidelines that will help future implementations fit the needs of a wide range of users. To help answer this, two research questions were presented. The research questions were "How usable are current online symptom checkers for older adults and the mildly intellectually disabled?" and "How to design symptom checkers for all users?". The following subsections will answer the two supporting research questions by reflecting on the findings of both the empirical part of the study as well as the literature review and discussing what can be learned from them.

5.1.1 Usability of current online symptom checkers

The two symptom checkers were well-received by the participants of this study and each participant stated they would use the services again in the future in case they fell ill for unknown reasons. However, the usability tests conducted suggest that there is still room for improvement. The different data collection methods revealed different usability issues that were present in the services being tested. Some issues were found either in one or in the other service, but many of the issues can be generalized as they were present in both symptom checkers.

The main issues encountered by the mildly intellectually disabled and the older adults were the way the information was displayed and the navigation on the page respectively. As discussed in Section 2.2, mildly intellectually disabled individuals may have limitations in their cognitive functioning and skills, meaning that it may be challenging to them to process a lot of new information (Patel et al., 2018; Setchell et al., 2021). This was apparent in some participants as they visibly struggled when faced with a lot of text to read and a lot information to process at once during the tests. The interviews revealed that the mildly intellectually disabled found it difficult to advance structurally from one section of the page to the next one and the vast amount of text in some cases made them exhausted, making it harder to focus on what was relevant at the given moment.

Meanwhile, the older adults' issues were to some extent more technological as they occasionally found it challenging to navigate on the webpages and interact with the different interface elements as they intended. Most noticeably, buttons and other options that were located off screen, either behind an additional button or that required scrolling down the webpage to be revealed, often went unnoticed as what was not visible or outside the view was regarded as non-existent. The older participants did not seem to notice or be aware that the webpages often extended further that what was been displayed at the moment on the screen if they would just scroll down the webpage. This, sometimes insufficient visibility of relevant information, led to participants from both user groups spending a considerable amount of time learning where the options they were looking for were located. This is reflected in the task completion times where, during the second try, the participants took on average half the time to complete the tasks compared to the first time they were testing the symptom checkers. Learnability played a role in this, as it was to be expected that during the second time the participants would take less time as they became more familiar with the webpage. However, to avoid situations where a new user takes 15 minutes just to find the beginning of the symptom checkers as was the case with one participant, certain steps can still be taken to make the participant's first experience with the service more effortless.

The usability issues had occasionally rather severe consequences. The wandering eyes due to the somewhat overwhelming user interfaces in addition to the unclear answering options, caused some of the participants to not see and hence miss important options which meant that their answers were sometimes incomplete. In some cases, this even led to a wrong recommendation altogether. These incomplete answers and wrong recommendations were mainly caused by the symptom checkers' usability issues that the participants did not even know were present but instead were noted by the tests' facilitator while observing the participants.

The participants' comments regarding the symptom checkers' usability as well as the observations made by the test facilitator are supported by the SUS scores which were acceptable but far from excellent as presented in Section 4.2.3. The Omaolo symptom checker's average score of 72 and Klinik Access' average score of 68 suggest that the usability of both services can be improved to better fit the participants' needs. The results can be compared to two studies conducted in Germany which show similar results as the perceived usability of the symptom checkers Ada and Rheport was measured resulting in SUS scores ranging between 72 and 78 (Knitza et al., 2020; Knitza et al., 2021). However, it should be noted that the user groups involved in the present study may be more demanding or may need certain special consideration in order to be included to the same extent as a user from the general population causing them to score the services slightly lower, thus presenting the services in a more negative light than actually is the case.

5.1.2 Designing symptom checkers for all users

Digital health services, and by extension online symptom checkers, should be accessible by the whole population, regardless of the individual's skill, age, and abilities. This is essential to remember when operating in the public sector, which the tested symptom checkers usually serve. Each person is different and hence might have different preferences or needs when it comes to using digital health services. To accommodate to a wide userbase's needs simultaneously, there are some additional aspects that should be considered.

As previously mentioned in Section 2.1.1, the concept of universal design sets some parameters which guide designers towards a more inclusive design. The seven principles (Connell et al., 1997) mentioned provide general rules and encourage good practices when it comes to designing products and services for a wide audience with varying needs. Acknowledging users with the widest range of capabilities during the design process, conceivably results in an outcome which addresses many users' needs better compared to a situation where they are ignored instead. However, since the seven principles of universal design were initially crafted for architecture, not all of the principles or accompanying guidelines may be directly applied in the domain of digital health services. For instance, accommodating for varying hand and grip sizes is relevant when thinking how a person opens and closes doors but it is less relevant when designing webpages. Nonetheless, many of the points emphasized in the seven principles are germane to other fields outside architecture. Based on the existing literature as well as the findings of the conducted usability tests, this thesis presents 13 guidelines in addition to the pre-established practices as an effort to improve the usability of digital health services. The 13 guidelines are designed to act as general reminders of what needs to be considered when designing services for a wide range of users while keeping in mind the overall usability of the systems, which in this case study were the online symptom checkers. As with following the concepts of universal design, designing inclusive and accessible systems improves the usability of said system for all users.

As mentioned above, the guidelines presented in this study along with their more detailed descriptions and examples (see Section 4.3), help designers avoid and navigate around usability issues that emerged during the usability tests. Following these guidelines, designers would pay more attention to aspects and parts of the systems where usability issues were present when testing with vulnerable user groups. For example, the previous section discussed how the most pressing usability issues were caused by the relevant information's visibility as well as the clarity of the different answering options. Based on these highlighted observations, guidelines such as “Keep any relevant information visible at all times” and “The options provided to the user should be clear and understandable” were formed as an effort to bring more attention to them and emphasize what needs to be acknowledged when designing for vulnerable user groups. Similarly, the guideline “Avoid long and uncommon words and difficult compound words” encourages to use plain language. Adopting the use of plain language not only would make the webpages more understandable and easier to process by the mildly intellectually disabled users, but it could make the user experience cognitively easier for the rest of the users as well.

5.2 Evaluation of the study

The presented guidelines are drawn from the results of the conducted usability tests and they are accompanied by examples of usability issues the participants encountered during the tests. It is to be noted that the proposed guidelines are, to an extent, similar and comparable to those found in existing literature. For example, Nielsen's 10 usability heuristics (1994c) and the 7 principles of universal design (Connell et al., 1997) share common similarities with this study's presented guidelines. This suggests that the usability issues found in this study are common, supporting and giving validity to the test results. Nevertheless, as this study focuses on the usability of namely online symptom checkers and it does so from the perspective of vulnerable user groups, the presented set of guidelines are more describing and appropriate for the goal that this study is trying to achieve and thereby emphasizes some different aspects such as the previously mentioned adoption of plain language for example.

When discussing about vulnerable user groups, there might be a presumption that vulnerable users are more demanding and laborious to deal with as they may need some special attention. However, since the guidelines derived from the findings of this study are analogous to existing heuristics and principles as mentioned above, it suggests that vulnerable users face the same issues as any other user. That being said, the impact which even the minor usability issues have on the user can be described as emphasized since when the participants involved in this study encountered some seemingly minor issues, at times the problems felt too great to overcome. In other words, even though the encountered usability issues may be the same, the impact they have on the user's experience may be greater and more severe when the user is someone with a mental disability or of older age.

Another aspect to be considered is that when asking for verbal feedback, the person providing the feedback may speak of their experience in a more positive way than what actually was the case. During the study, there was no reason to believe the participants were insincere when testing the services and providing feedback. In fact, all the participants were encouraged to criticize the services and give honest feedback since the objective was to find how the services could be improved. Still, some participants who encountered rather major issues during the tests gave positive feedback stating that everything went smoothly and that they were content with how things had transpired. For this reason, this study incorporated triangulation. The methods used in this study were seen appropriate to accomplish the goal of this study and they all served a purpose adding value to the end result. As previously discussed, different data collection methods yield different results and hence this study gathered and analyzed data from both the participants' perspective as well as that of the facilitator giving a more rounded result and conceivably reflecting the results in a more truthful manner.

The COVID-19 pandemic also affected the results to some extent because even though the participants were instructed to skip and ignore any section or question related to COVID-19, the participants still encountered some issues due to the presence of pandemic specific questions and the way in which the interfaces were structured (to emphasize the symptom checker related to COVID-19) in order to make it more visible and accessible. In one service more than in the other, this influenced the results in the sense that many participants felt misled at first since the service seemed to push all users, no matter what their visit concerned, towards filling in the questionnaire regarding COVID-19. This meant that other symptom checkers on the webpage were significantly less visible and therefore harder to find. These elements that would not have been present if it were not for the COVID-19 pandemic, created some new usability issues and hence lengthened the task completion times during the participants' first test try. As the participants were more familiar with the interface on their second try, they knew how to navigate around the COVID-19 symptom checker saving a considerable amount of time.

Using standardized vignettes proved to be useful for a couple of reasons. First, it allowed the simulation of a real-world situation without the need of using the participants' personal medical information. Participants can see the collection of their own personal data stressful and uncomfortable as shown in a study conducted by Johnson & Clarke (2003). The ethical concern placed on the study was also relieved as the amount of sensitive data that needs to be handled is minimized. The use of standardized vignettes also allowed the study to simulate a real situation in the sense that the used symptoms and conditions were chosen with the user groups in mind. Common conditions were selected and no participant was given a condition that would be too unrealistic to them. A couple of participants even mentioned that they found the symptoms to be relatable and they had no issues in understanding the vignettes or imagining they were experiencing said symptoms. In addition, having control over the vignettes and symptoms entered into the systems as well as using standardized vignettes allows possible further studies to recreate the steps taken in this study and compare the results, for example, when the service has been updated. The possibly differing results might indicate that the changes made to the system have been positive providing proof of improvement. Since there were two different user groups in this study, differences between the user groups and their representatives could also be analyzed given that, at least in theory, the participants who had the same vignettes should have answered the questions in a similar manner. However, as was discussed in Section 4.2.1, in a couple of cases the participants' answers or the symptom checkers' end results were not the same. Further investigation might be required into the reason why the participants make different choices even though they are given precisely the same premise.

5.3 Limitations of the study

One limitation to be considered when evaluating the study is the relatively small sample size. For this study, a total of eight participants were recruited with half of them being from one user group and the remaining half from the other. Although this allowed the collection of qualitative data, it also meant that only a small group of people's voices were heard. Similarly, the user groups involved in this study do not represent all users who could be considered vulnerable users. Other vulnerable user groups, such as the unemployed, immigrants, and people with disabilities other than intellectual, were also considered during the planning phase but were not included in this study as the scope needed to be narrowed down.

In addition to the limited sample size, all of the usability tests and interactions with the participants were conducted through online means due to the COVID-19 pandemic. This too imposed certain limitations on the study. For instance, additional time was spent solving communicational problems that emerged regarding the screen-sharing and navigating from one service to the other. These nuisances, even though minor, might have tired some participants more than intended, and therefore might have led to tests where the participants were not fully concentrated in the task at hand. Fortunately, in most instances the participants had a support person on standby ready to help with technical problems which lessened their impact.

Conducting the tests remotely also caused the interactions between the facilitator and the participants to lose some humanness. The participants often had only one computer screen which meant that, when they had questions during the test, they could not see the facilitator who was speaking to them and helping them since the participants had the symptom checker open on their screen and the video call window minimized. Just hearing a voice speaking to them but not being able to see the speaker’s face might have added a level of secrecy or insecurity causing the participant to be more nervous. Similarly, a couple of participants also voiced that they were a bit nervous to think aloud while conducting the study as they were asked to do. Allowing the facilitator to be present when conducting the tests would make the conversation between the participant and the facilitator easier and more effortless for all parties.

Another limitation of the study was the involvement of sensitive data. In order to minimize the handling of sensitive data, standardized symptom vignettes were used so that the participants would not need to enter their own health information into the services. This simulation differs from a real situation where the user would use actual data concerning their health and have a better understanding of what symptoms they had and would therefore be able to answer the questions more accurately and in more detail. Instead, in this study the participants had to refer to a paper printed for them in order to see the symptoms. The used vignettes were also simplified in order to be more understandable. This meant that the participants had to spend some time coming up with their own response whenever they were asked something that had no straight answer in the provided vignette.

5.4 Ethics

Since this study revolved around vulnerable users and sensitive data, certain additional aspects were considered when planning the tests. When the study’s participants were first contacted, they were briefly informed what the study was about and were given a consent form to sign as well as a privacy statement to read through. The privacy statement explained how the collected data such as the video and screen recordings would be handled and for what purpose they were used. This also included how the data would be stored and for how long. Before the test, the participants were asked if they understood everything and if they had any questions regarding the privacy statement or anything else that might concern them. In addition, the participants were given a more detailed description of the whole test including the different phases of the session and about their duration. Finally, they were explained how they could stop the test at any time and for any reason if they wished to do so.

All of the data collected in this study has been handled respectfully and as intended and has been stored and used as the participants were told. Participant anonymity was kept in mind and addressed throughout this study. For example, some questions which may have led to identifying the participants were discussed in a more general manner and the data was excluded from the table describing each participant's characteristics. The used privacy statement, consent form, and data management practices were approved by the Ethical Review Board of the Aalto University.

5.5 Future work

The future suggestions mainly address limitations discovered in this study. As the sample size in this study was relatively small, having a bigger sample size would allow to get a more well-rounded picture of the usability of the tested services and perhaps address more usability issues than encountered by the participants in this study. Other users may also give more weight to different usability issues.

Additional benefits could result if other user groups would be involved. This study only addressed two user groups deemed as vulnerable users but future research could augment the scope and address the needs of other user groups and then compare the results. In addition to focusing on other user groups, other health services could also be studied. This study involved two online symptom checkers but other digital health services may highlight additional issues or may have already overcome the issues encountered in this study's services. Studying and comparing different services will allow service providers to learn from one another and further improve their system's usability.

Lastly, the applicability of the presented guidelines could be studied in different contexts or in other domains. The guidelines presented in this study are the result of the data collected and analyzed from the conducted usability tests. Therefore, they are designed to be generalizable and applicable to different services. Future research may focus on further developing and iterating the guidelines to more accurately assess how easy they are to follow and how big of an improvement they actually cause in a digital health service's usability.

6 Conclusion

The aim of this research was to study how usable existing online symptom checkers are from the point of view of vulnerable user groups and what to consider when developing these services in the future. For this, usability tests were conducted on two of the most prevalent symptom checkers in Finland, Omaolo symptom checker and Klinik Access. A total of eight participants were recruited for this study, four people with mild intellectual disabilities and four older adults. These user groups were selected since they can be considered to be at risk of digital exclusion, which then may lead to social exclusion. This may happen as health services become progressively more digitized restricting some people's access to them, leading to possible situations where vulnerable users' needs are not taken into account during a service's design phase, thereby resulting in designs that are not as inclusive and accessible to the whole population. By developing usable services for people of all ages, abilities, and skill levels, everyone can benefit, as stated in the principle of universal design.

Various data collection methods were used in this study. A combination of the think-aloud method, observations, questionnaires, and semi-structured interviews was used to better understand the research problem. Through observing the participants while they completed the tasks and by asking them to think aloud during the usability tests, a number of usability issues were found. These usability issues negatively affected the participants' ability to fill in the symptom checkers in an efficient manner. To evaluate the perceived usability of the services being tested, the participants were asked to fill in a System Usability Scale (SUS) questionnaire. The Omaolo symptom checker received an average score of 72 and Klinik Access averaged a score of 68. These scores are acceptable but leave room for improvement. Both, the results from the SUS as well as the interviews, suggest that the mildly intellectually disabled participants preferred the Omaolo symptom checker over Klinik Access as it presented the user with less information at once and the different options were more clearly displayed, thus using said service was seen as cognitively less straining. The older adults on the other hand, highlighted the visibility of relevant information as one of the main issues they encountered during the usability tests. At times, the information was either hard to find or accessing it required interactions that were not so intuitive for all the participants, leading to situations where the hidden information was unreachable and, therefore, the symptom checker could not be completed properly.

Based on the gathered data and analyzed results, this study presents 13 guidelines that address issues found in the usability tests such as the insufficient visibility of relevant information and the clearness of the different options in the services. These guidelines are an effort to make online symptom checkers usable by people with the widest range of capabilities and therefore minimize service quality gaps between the different user groups, no matter their age or capabilities. These guidelines are also an attempt to guide future design work and to act as a reminder of which features and design choices make for a more usable whole. Learning from the usability issues revealed in this study and applying this acquired information in the future will improve the chances to take every user into consideration and ultimately result in universally more usable online symptom checkers.

References

Aboueid, S., Meyer, S. B., Wallace, J. R., Mahajan, S., Nur, T., & Chaurasia, A. (2021a). Use of symptom checkers for COVID-19-related symptoms among university students: a qualitative study. *BMJ Innovations*, *7*(2).

Aboueid, S., Meyer, S., Wallace, J. R., Mahajan, S., & Chaurasia, A. (2021b). Young Adults' Perspectives on the Use of Symptom Checkers for Self-Triage and Self-Diagnosis: Qualitative Study. JMIR Public Health and Surveillance, 7(1), e22637.

Alfredsson Ågren, K., Kjellberg, A., & Hemmingsson, H. (2020). Digital participation? Internet use among adolescents with and without intellectual disabilities: A comparative study. New Media & Society, 22(12), 2128–2145. https://doi.org/10.1177/1461444819888398

Arain, A. A., Hussain, Z., Rizvi, W. H., & Vighio, M. S. (2016, July). Evaluating usability of M-learning application in the context of higher education institute. In *International conference on learning and collaboration technologies* (pp. 259-268). Springer, Cham.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114-123.

Bas-Sarmiento, P., Poza-Méndez, M., Fernández-Gutiérrez, M., González-Caballero, J. L., & Falcón Romero, M. (2020). Psychometric Assessment of the European Health Literacy Survey Questionnaire (HLS-EU-Q16) for Arabic/French-Speaking Migrants in Southern Europe. *International journal of environmental research and public health*, *17*(21), 8181.

Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998?. In *International Conference on Human-Computer Interaction.* (pp. 143-151). Springer, Cham.

Bevan, N., Kirakowskib, J., & Maissela, J. (1991). What is usability. In *Proceedings of the 4th International Conference on HCI.*

Boot, F. H., Owuor, J., Dinsmore, J., & MacLachlan, M. (2018). Access to assistive technology for people with intellectual disabilities: a systematic review to identify barriers and facilitators. *Journal of Intellectual Disability Research*, *62*(10), 900-921.

Brooke, J. (2013). SUS: a retrospective. *Journal of usability studies*, *8*(2), 29-40.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

Centre for inclusive Design. (2020). Report: Research reveals Inclusive Design can expand customer reach fourfold. Centre for inclusive Design. accessed 20 May 2021. <https://centreforinclusivedesign.org.au/index.php/services/reports/2020/03/31/report-research-reveals-inclusive-design-can-expand-customer-reach-fourfold/>.

Chambers, D., Cantrell, A. J., Johnson, M., Preston, L., Baxter, S. K., Booth, A., & Turner, J. (2019). Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. *BMJ open*, *9*(8), e027743.

Chan, W., Woo, J., & Hui, E. (2013). The role of empowerment in the management of chronic diseases in the elderly. In *Aging in Hong Kong* (pp. 183-210). Springer, Boston, MA.

Collins, J. C., & Collet-Klingenberg, L. (2018). Portable electronic assistive technology to improve vocational task completion in young adults with an intellectual disability: A review of the literature. *Journal of Intellectual Disabilities, 22*(3), 213-232.

Connell, B. R., Jones, M. L., Mace, R. L., Mueller, J. L., Mullick, A., Ostroff, E., Sanford, J., Steinfeld, E., Story, M., Vanderheiden, G. (1997). *The Principles of Universal Design, Version 2.0,* Raleigh, N.C.: Center for Universal Design, North Carolina State University.

Copeland, C., Morreale, P., & Li, J. (2018, July). m-Health application Interface design for symptom checking. In *10th International Conference on e-Health.*

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research (3rd ed.): Techniques and procedures for developing grounded theory.* SAGE Publications, Inc. https://www.doi.org/10.4135/9781452230153.

Dalal, N. P., Quible, Z., & Wyatt, K. (2000). Cognitive design of home pages: an experimental study of comprehension on the World Wide Web. *Information Processing & Management, 36*(4), 607-621.

DigiFinland. (2020). Omaolo service. DigiFinland. accessed 30 July 2021. <https://digifinland.fi/en/our-operations/omaolo-service/>.

DigiIN. (2021). Project info. DigiIN. accessed 1 August 2021. <https://digiin.fi/en/consortium-and-partners/>.

Dumas, J. S., & Redish, J. (1999). *A practical guide to usability testing.* Intellect books.

Easterbrook, S., Singer, J., Storey, M. A., & Damian, D. (2008). Selecting empirical methods for software engineering research. In Guide to advanced empirical software engineering (pp. 285-311). Springer, London.

Eronen, J., Paakkari, L., Portegijs, E., Saajanaho, M., & Rantanen, T. (2019). Assessment of health literacy among older Finns. *Aging clinical and experimental research, 31*(4), 549-556.

Feldman, M. A., Owen, F., Andrews, A. E., Tahir, M., Barber, R., & Griffiths, D. (2016). Randomized control trial of the 3Rs health knowledge training program for persons with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 29*(3), 278-288.

Frielink, N., Oudshoorn, C. E., & Embregts, P. J. (2020). eHealth in support for daily functioning of people with intellectual disability: Views of service users, relatives, and professionals on both its advantages and disadvantages and its facilitating and impeding factors. *Journal of Intellectual & Developmental Disability*, 1-11.

Gassmann, O., & Reepmeyer, G. (2008). Universal Design–Innovations for All Ages. In *The silver market phenomenon*(pp. 125-140). Springer, Berlin, Heidelberg.

Geukes, C., Bröder, J., & Latteck, Ä. D. (2019). Health literacy and people with intellectual disabilities: What we know, what we do not know, and what we need: A theoretical discourse. *International journal of environmental research and public health, 16*(3), 463.

Gkouskos, D., & Burgos, J. (2017). I'm in! Towards participatory healthcare of elderly through IOT. *Procedia computer science, 113,* 647-652.

Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., Hinder, S., ... & Shaw, S. (2017). Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of medical Internet research*, *19*(11), e367.

Harrington, C. N., Ruzic, L., & Sanford, J. A. (2017, July). Universally accessible mHealth apps for older adults: Towards increasing adoption and sustained engagement. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 3-12). Springer, Cham.

Hesse, B. W., & Shneiderman, B. (2007). eHealth research from the user's perspective. *American journal of preventive medicine*, *32*(5), S97-S103.

Hesse, B. W. (1995). Curb cuts in the virtual community: Telework and persons with disabilities. In *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences* (Vol. 4, pp. 418-425). IEEE.

Holden, R. J. (2020). A simplified system usability scale (SUS) for cognitively impaired and older adults. In *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* (Vol. 9, No. 1, pp. 180-182). Sage CA: Los Angeles, CA: SAGE Publications.

International Organization for Standardization. (2018). Ergonomics of human-system interaction – Part 11: Usability: Definition and concepts (ISO Standard No. 9241-11:2018). Accessed 13 July 2021. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>.

Irish, J. E. (2020). Increasing participation: Using the principles of universal design to create accessible conferences. In *Journal of Convention & Event Tourism* (Vol. 21, No. 4, pp. 308-330). Routledge.

Johnson, B., & Clarke, J. M. (2003). Collecting sensitive data: The impact on researchers. *Qualitative health research*, *13*(3), 421-434.

Jordan, P. W. (1998). Human factors for pleasure in product use. *Applied ergonomics*, *29*(1), 25-33.

Jovanić, M., Zdravković, M., Stanisavljević, D., & Jović Vraneš, A. (2018). Exploring the importance of health literacy for the quality of life in patients with heart failure. *International journal of environmental research and public health*, *15*(8), 1761.

Kadir, S. A., & Jamaludin, M. (2013). Universal design as a significant component for sustainable life and social development. *Procedia-Social and Behavioral Sciences*, *85*, 179-190.

Kehitysvammaliitto. (2021). Kehitysvammaisuus. Kehitysvammaliitto. accessed 16 July 2021. <https://www.kehitysvammaliitto.fi/kehitysvammaisuus/>.

Kim, H. (2017). A Study on Usability Improvement of Mobile Healthcare Services. *International journal of advanced smart convergence*, *6*(2), 72-81.

Klinik Healthcare Solutions. (2021). *Klinik Healthcare Solutions*. Accessed 1 June 2021. <https://klinikhealthcaresolutions.com>.

Knitza, J., Mohn, J., Bergmann, C., Kampylafka, E., Hagen, M., Bohr, D., ... & Welcker, M. (2021). Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. *Arthritis Research & Therapy*, *23*(1), 1-7.

Knitza, J., Mohn, J., Bergmann, C., *et al.* (2020). AB1346-HPR Real-world effective ness and perceived usefulness of symptom checkers in rheumatology: interim report from the prospective multicenter better study. *Annals of the Rheumatic Diseases* 2020; 79:1961.

Kujala, S. (2003). User involvement: a review of the benefits and challenges. *Behaviour & information technology*, *22*(1), 1-16.

Kujala, S., Hörhammer, I., Hänninen-Ervasti, R., & Heponiemi, T. (2020). Health Professionals' Experiences of the Benefits and Challenges of Online Symptom Checkers. In *MIE*(pp. 966-970).

Larson, S. A., Anderson, L., & Dlojanac, R. (2005). Health promotion for persons with intellectual and developmental disabilities. *Access to Health Care. Washington: American Association on Mental Retardation*, 129-84.

Laubheimer, P. (2015a). Preventing User Errors: Avoiding Unconscious Slips. Nielsen Norman Group. accessed 30 July 2019. <https://www.nngroup.com/articles/ slips/>.

Laubheimer, P. (2015b). Preventing User Errors: Avoiding Conscious Mistakes. Nielsen Norman Group. accessed 30 July 2019. <https://www.nngroup.com/articles/ user-mistakes/>.

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013, April). UMUX-LITE: when there's no time for the SUS. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2099-2102).

Liu, V. D., Sellgren, L., Kaila, M., & Koskela, T. (2021). Sähköisten oirearvioiden käytettävyys. *Finnish Journal of eHealth and eWelfare*, *13*(2), 100-112.

Lorini, C., Lastrucci, V., Mantwill, S., Vettori, V., Bonaccorsi, G., & Florence Health Literacy Research Group. (2019). Measuring health literacy in Italy: The validation study of the HLS-EU-Q16 and of the HLS-EU-Q6 in Italian language. *Annali dell'Istituto superiore di sanita*, *55*(1), 10-18.

Lussier-Desrochers, D., Normand, C. L., Romero-Torres, A., Lachapelle, Y., Godin-Tremblay, V., Dupont, M. È., ... & Bilodeau, P. (2017). Bridging the digital divide for people with intellectual disability. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *11*(1).

Mace, R. (1988). Universal design: housing for the lifespan of all people. *The Center for Universal Design, North Carolina State University*.

Maramba, I., Chatterjee, A., & Newman, C. (2019). Methods of usability testing in the development of eHealth applications: a scoping review. *International journal of medical informatics*, *126*, 95-104.

Marston, H. R., Genoe, R., Freeman, S., Kulczycki, C., & Musselwhite, C. (2019). Older adults' perceptions of ICT: Main findings from the technology in later life (TILL) study. In *Healthcare* (Vol. 7, No. 3, p. 86). Multidisciplinary Digital Publishing Institute.

Martin, C., Hope, S., Zubairi, S., & Scotland, I. M. (2016). The role of digital exclusion in social exclusion. *Ipsos MORI Scotland, CarnegieUK trust*.

Mehta, A., Singh, A., Deshmukh, S., & Tathawade, P. (2020). The New Era of Digital Transformation in Healthcare: "Emerging Technologies for Value-Centered Marketing in Healthcare Ecosystem". European Journal of Molecular & Clinical Medicine, 7(11).

Meyer, A. N., Giardina, T. D., Spitzmueller, C., Shahid, U., Scott, T. M., & Singh, H. (2020). Patient Perspectives on the Usefulness of an Artificial Intelligence–Assisted Symptom Checker: Cross-Sectional Survey Study. Journal of medical Internet research, 22(1), e14679.

Moran, K. (2019). *Usability Testing 101*. Nielsen Norman Group. accessed 1 June 2021. <https://www.nngroup.com/articles/usability-testing-101/>.

Morse, K. E., Ostberg, N. P., Jones, V. G., & Chan, A. S. (2020). Use Characteristics and Triage Acuity of a Digital Symptom Checker in a Large Integrated Health System: Population-Based Descriptive Study. Journal of medical Internet research, 22(11), e20549.

National Academies of Sciences, Engineering, and Medicine. (2015). *Mental disorders and disabilities among low-income children*. National Academies Press.

National Disability Authority. (2021). What is Universal Design? Centre for Excellence in Universal Design, National Disability Authority. Accessed 14 July 2021. <http://universaldesign.ie/what-is-universal-design/>.

Nielsen, J. (2012). Usability 101: Introduction to Usability. Nielsen Norman Group. accessed 14 July 2021. <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>.

Nielsen, J. (1994a). 10 Usability Heuristics for User Interface Design. Nielsen Norman Group. accessed 19 May 2021. <https://www.nngroup.com/articles/ten-usabilityheuristics/>.

Nielsen, J. (1994b). Estimating the number of subjects needed for a thinking aloud test. *International journal of human-computer studies*, *41*(3), 385-397.

Nielsen, J. (1994c). Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 152-158).

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

North, F., Ward, W. J., Varkey, P., & Tulledge-Scheitel, S. M. (2012). Should you search the internet for information about your acute symptom?. *Telemedicine and e-Health*, *18*(3), 213-218.

Omaolo. (2021). *Omaolo*. accessed 1 June 2021. <https://www.omaolo.fi/>.

Oudshoorn, C. E., Frielink, N., Nijs, S. L., & Embregts, P. J. (2020). eHealth in the support of people with mild intellectual disability in daily life: A systematic review. *Journal of Applied Research in Intellectual Disabilities*, *33*(6), 1166-1187.

Parnell, T. A. (2014). *Health literacy in nursing: Providing person-centered care*. Springer Publishing Company.

Patel, D. R., Cabral, M. D., Ho, A., & Merrick, J. (2020). A clinical primer on intellectual disability. *Translational pediatrics*, *9* (Suppl 1), S23.

Patel, D. R., Apple, R., Kanungo, S., & Akkal, A. (2018). Intellectual disability: definitions, evaluation and principles of treatment. *Pediatr. Med*, *1*(11).

Petrie, H., & Bevan, N. (2009). The Evaluation of Accessibility, Usability, and User Experience. *The universal access handbook*, *1*, 1-16.

Raspa, M., Fitzgerald, T., Furberg, R. D., Wylie, A., Moultrie, R., DeRamus, M., ... & McCormack, L. (2018). Mobile technology use and skills among individuals with fragile X syndrome: implications for healthcare decision making. *Journal of Intellectual Disability Research*, *62*(10), 821-832.

Raviselvam, S., Wood, K. L., Hölttä-Otto, K., Tam, V., & Nagarajan, K. (2016). A lead user approach to universal design–Involving older adults in the design process. *Studies in health technology and informatics*, *229*, 131-140.

Rizzo, A. A., Kim, G. J., Yeh, S. C., Thiebaux, M., Hwang, J., & Buckwalter, J. G. (2005, July). Development of a benchmarking scenario for testing 3D user interface devices and interaction methods. In *Proceedings of the 11th International Conference on Human Computer Interaction, Las Vegas, Nevada, USA*.

Ruzic, L., & Sanfod, J. A. (2017). Universal design mobile interface guidelines (UDMIG) for an aging population. In *Mobile e-Health* (pp. 17-37). Springer, Cham.

Scherer, M. J., & Craddock, G. (2002). Matching person & technology (MPT) assessment process. *Technology and Disability*, *14*(3), 125-131.

Seaman, C. B. (1999). Qualitative methods in empirical studies of software engineering. *IEEE Transactions on software engineering*, *25*(4), 557-572.

Seifert, A., Hofer, M., & Rössel, J. (2018). Older adults' perceived sense of social exclusion from the digital world. *Educational Gerontology*, *44*(12), 775-785.

Seifert, A., Cotten, S. R., & Xie, B. (2020). A double burden of exclusion? Digital and social exclusion of older adults in times of COVID-19. *The Journals of Gerontology: Series B*, *76*(3), e99-e103.

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. bmj, 351.

Setchell, J., Barlott, T., & Torres, M. (2021). A socio-emotional analysis of technology use by people with intellectual disabilities. *Journal of Intellectual Disability Research*, *65*(2), 149-161.

Shackel, B. (2009). Usability–context, framework, definition, design and evaluation. *Interacting with computers*, *21*(5-6), 339-346.

Shogren, K. A., Wehmeyer, M. L., Lassmann, H., & Forber-Pratt, A. J. (2017). Supported decision making: A synthesis of the literature across intellectual disability, mental health, and aging. *Education and Training in Autism and Developmental Disabilities*, *52*(2), 144-157.

Sinkkonen, I., Kuoppala, H., Parkkinen, J. & Vastamäki, R. (2006). Käytettävyyden psykologia. Helsinki: Edita Prima.

Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018). *This is service design doing: applying service design thinking in the real world* : a practitioners' handbook (First Edition.). O'Reilly Media, Inc.

Storms, H., Claes, N., Aertgeerts, B., & Van den Broucke, S. (2017). Measuring health literacy among low literate people: an exploratory feasibility study with the HLS-EU questionnaire. *BMC public health*, *17*(1), 1-10.

Story, M. F., Mueller, J. L., & Mace, R. L. (1998). The universal design file: Designing for people of all ages and abilities.

Sørensen, K., Pelikan, J. M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., ... & Brand, H. (2015). Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *European journal of public health*, *25*(6), 1053-1058.

Sørensen, K., Van den Broucke, S., Pelikan, J. M., Fullam, J., Doyle, G., Slonska, Z., ... & Brand, H. (2013). Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). *BMC public health*, *13*(1), 1-10.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., ... & Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC medical research methodology*, *10*(1), 1-10.

Thompson, J. R., Bradley, V. J., Buntinx, W. H., Schalock, R. L., Shogren, K. A., Snell, M. E., ... & Yeager, M. H. (2009). Conceptualizing supports and the support needs of people with intellectual disability. *Intellectual and developmental disabilities*, *47*(2), 135-146.

Tsai, C. H., You, Y., Gui, X., Kou, Y., & Carroll, J. M. (2021, May). Exploring and Promoting Diagnostic Transparency and Explainability in Online Symptom Checkers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Tukiliitto. (2021). Tietoa kehitysvammasta. Tukiliitto. accessed 16 July 2021. <https://www.tukiliitto.fi/tietoa-kehitysvammasta/>.

Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour & information technology, 22(5), 339-351.

Van Teijlingen, E. R., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, (35). http://sru.soc.surrey.ac.uk/SRU35.html.

Vasalou, A., Ng, B. D., Wiemer-Hastings, P., & Oshlyansky, L. (2004). Human-moderated remote user testing: Protocols and applications. In 8th ERCIM Workshop, User Interfaces for All, Wien, Austria (Vol. 19). sn.

Ware, P., Bartlett, S. J., Paré, G., Symeonidis, I., Tannenbaum, C., Bartlett, G., ... & Ahmed, S. (2017). Using eHealth technologies: interests, preferences, and concerns of older adults. *Interactive journal of medical research*, *6*(1), e3.

Wass, S., & Safari, M. C. (2020). Photovoice—Towards Engaging and Empowering People with Intellectual Disabilities in Innovation. *Life*, *10*(11), 272.

Watfern, C., Heck, C., Rule, C., Baldwin, P., & Boydell, K. M. (2019). Feasibility and acceptability of a mental health website for adults with an intellectual disability: qualitative evaluation. *JMIR mental health*, *6*(3), e12958.

Westerinen, H. (2018). *Prevalence of intellectual disability in Finland.* [H. Westerinen]. accessed 5 August 2021. <https://helda.helsinki.fi/bitstream/handle/10138/250864/Prevalen.pdf?sequence=1&isAllowed=y>.

Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, *15*(1), 45-55.

Yin, R.K. (2002) Case study research: Design and Methods. (3rd ed.). Thousand Oaks, Calif: Sage Publications.

Zapata, B. C., Fernández-Alemán, J. L., Idri, A., & Toval, A. (2015). Empirical studies on usability of mHealth apps: a systematic literature review. *Journal of medical systems*, *39*(2), 1-19.

Appendix A: Background information survey

The background information questionnaire has been translated from Finnish to English for the purposes of this thesis.

Background information

The purpose of this questionnaire is to gather background information about the participants. Try to answer all questions as accurately as possible. Your answers will be anonymized so they cannot be connected to you.

1. Gender
   __ female
   __ male
   __ other / I don’t want to answer ___________

2. Age
   _____ years

3. How many times have you visited a doctor in the last two years?
   __ 0 – 2
   __ 3 – 5
   __ 6 – 9
   __ 10 +
   __ other ___________

4. How many medical conditions diagnosed by a doctor do you have?
   __ 0 – 1
   __ 2 – 3
   __ 4 – 5
   __ 6 +
   __ other ___________

5. Have you used digital health services before? (For example, Omakanta) If you answered yes, how many times?
   __ No
   __ Yes, ___________

6. How often do you use digital devices? For example, a computer, a smartphone, a tablet
   __ I don’t use digital devices
   __ A couple of times a week
   __ Multiple times a day
   __ other ___________

Appendix B: Health literacy survey

The health literacy survey is in English here for the purposes of this thesis, as applied by Lorini et al. (2019). The used Finnish translation is based on the translation by Eronen et al. (2019).

Health literacy survey

Health literacy refers to an individual's ability to access, understand, and then apply resources to make decisions regarding one's health. Try to answer all questions as accurately as possible. Your answers will be anonymized so they cannot be connected to you.

On a scale from very easy to very difficult, how easy would you say it is to	Very easy	Fairly easy	Fairly difficult	Very difficult	I don't know
1. Find information on treatments of illnesses that concern you?					
2. Find out where to get professional help when you are ill?					
3. Understand what your doctor says to you?					
4. Understand your doctor's or pharmacist's instructions on how to take a prescribed medicine?					
5. Judge when you may need to get a second opinion from another doctor?					
6. Use information the doctor gives you to make decisions about your illness?					
7. Follow instructions from your doctor or pharmacist?					
8. Find information on how to manage mental health problems like stress or depression?					
9. Understand health warnings about behavior such as smoking, low physical activity and drinking too much?					
10. Understand why you need health screenings?					
11. Judge if the information on health risks in the media is reliable?					
12. Decide how you can protect yourself from illness based on information in the media?					
13. Find out about activities that are good for you mental well-being?					
14. Understand advice on health from family members or friends?					
15. Understand information in the media on how to get healthier?					
16. Judge which everyday behavior is related to your health?					

Appendix C: SUS

The SUS questionnaire is in English here for the purposes of this thesis, as applied by Brooke (1996). During the usability test, a Finnish translation was used. The word "system" was also replaced by the corresponding service names, "Omaolo symptom checker" and "Klinik Access" and the participant was reminded of the service's name they had just tested.

System Usability Scale (SUS)

On a scale from 1 to 5, would you agree or disagree with the following statements?	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think I would like to use this system frequently.					
2. I found the system unnecessarily complex to use.					
3. I thought the system was easy to use.					
4. I think that I would need the support of a technical person to be able to use this system.					
5. I found the various functions in this system were well integrated.					
6. I thought there was too much inconsistency in this system.					
7. I would imagine that most people would learn to use this system very quickly.					
8. I found the system very cumbersome to use.					
9. I felt very confident using the system.					
10. I needed to learn a lot of things before I could get going with this system.					

Appendix D: Symptom vignettes

Symptom vignettes used in this study, their severity rating, and how they were assigned to the participants. The order in which the participants tested the services is also included.

Symptom vignettes

Requires emergent care

Meningitis – Imagine your head has been hurting and you have had a fever for the past 3 days. You have a stiff neck. Your eyes are sensitive to light and a normal amount of light makes you squint your eyes and your head hurt even more.

Deep vein thrombosis – Imagine your right leg has been hurting and swollen for the past 5 days. Your right leg is also more sensitive and redder compared to your left leg. When you touch the back of your knee it hurts even more. You have previously visited the doctor due to your high blood pressure and you have had complications with your heart in the past. Lately you have been laying down a lot while recovering from pneumonia.

Pneumonia – Imagine you have had a wet cough and a fever for the past 3 days. You have a fever of 38.3°C. You also have a high blood pressure and shortness of breath making it difficult for you to breath so you have to breath quite fast. You also feel more tired than normal.

Requires non-emergent care

Influenza – Imagine you have felt quite weak and you have had a fever for the past 2 days. Now you have a fever of 38.1°C, your head hurts, and you have a cough. The symptoms appeared suddenly. Your coworkers have been sick lately. You have not received the influenza vaccine yet.

Back pain – Imagine you were shoveling snow 3 weeks ago and your lower back has hurt since. You find it difficult to raise your right foot when you walk. Your toes are also a bit numb.

Self-care appropriate

Acute bronchitis – Imagine you have had a cough for the past 12 days. You feel like there is a lot of sputum in your throat and your throat is a bit sore. You have not had a fever. Lately you have not met anyone sick with similar symptoms.

Mildly intellectually disabled participants:

Participant 1 – influenza & meningitis – 1. Omaolo, 2. Klinik Access
Participant 2 – acute bronchitis & back pain – 1. Klinik Access, 2. Omaolo
Participant 3 – meningitis & influenza – 1. Klinik Access, 2. Omaolo
Participant 4 – back pain & acute bronchitis – 1. Omaolo, 2. Klinik Access

Older adult participants:

Participant 5 – back pain & pneumonia – 1. Klinik Access, 2. Omaolo
Participant 6 – pneumonia & back pain – 1. Omaolo, 2. Klinik Access
Participant 7 – deep vein thrombosis & back pain – 1. Omaolo, 2. Klinik Access
Participant 8 – back pain & deep vein thrombosis – 1. Klinik Access, 2. Omaolo

Appendix E: Usability issues

A list of the usability issues found during the usability tests, the service in which they appeared, and the number of participants who encountered the issues.

Category	Encountered usability issue	Service	Number of testers who encountered the issue
Visibility of information	Most of the participants clicked on the only visible button on the front page and ended up doing the wrong symptom checker. They started to do the COVID-19 symptom checker even though they had been instructed to skip anything COVID-19 related. The participants did not notice that they could scroll down on the page to reveal other symptom checkers. They also did not know which symptom checker they had chosen after landing on the symptom checker's page.	Omaolo	6
	Some participants had difficulties in locating the beginning of the symptom checker. The first question, which is not really a question ("I think I need treatment and I want to fill out a symptom checker"), has a much smaller font and is much less emphasized than the rest of the text on the page so some participants did not see it.	Omaolo	3
	After submitting the symptom checker, one participant did not scroll down the page and therefore did not notice that the symptom checker can be continued after the first submission by choosing one's municipality and by identifying oneself.	Omaolo	1
	Some participants forgot that in the beginning of the symptom checker they had been instructed to answer each question so they left some questions unanswered and when they got to the end and tried to submit the symptom checker they were asked to go back and answer the remaining questions since all questions were mandatory.	Omaolo	2
	When the user is asked to write on a text field and further describe their symptoms, the detailing questions about the symptom are situated in the text field as placeholders. When the users click on the text field and start answering, the questions disappear and out of the four presented questions, none of the participants answered to more than two of them, most of the participants only gave an answer to the first question since it was the only one they remembered after they started typing.	Klinik Access	8

	When the user is asked to write on a text field and further describe their symptoms, the detailing questions about the symptom are situated in the text field as placeholders. When the users click on the text field and start answering, the questions disappear and out of the four presented questions, none of the participants answered to more than two of them, most of the participants only gave an answer to the first question since it was the only one they remembered after they started typing.	Klinik Access	8
Clarity of answering options	Many participants had trouble choosing between the two options in the beginning of the symptom checker, “I think I need treatment and I want to fill out a symptom checker” and “My symptoms are suitable for self-care and I only want self-care instructions”, since they could not tell how the options differ from one another.	Omaolo	4
	Some participants did not know that they could choose multiple options as answers to a question so they only chose one option even though multiple conditions applied in their case. A similar question clarifies that one is able to choose multiple options but some questions did not include the clarification.	Omaolo	3
	Half of the participants had difficulties with choosing the right symptom checker from the list of available symptom checkers. The participants thought that choosing a symptom checker would be a big commitment and they were afraid of choosing the wrong symptom checker.	Omaolo	4
	Some of the answering options have multiple symptoms in one. This led to situations where the participants would read only half of the statement and then decide that this was not the option they were looking for even though the latter half of the answer did indeed include what they would have wanted to select. For example, “intense temporal pain and problems with vision”, the participants did not finish reading the whole sentence and hence did not choose the correct answer.	Omaolo	2
	Some participants did not notice that the links to the symptom checkers on the front page were actual clickable links and therefore did not know how to advance and start one of the symptom checkers. The links could be made to look more like selectable options or the text above the list could instruct the user to “select one option from the list below” for example.	Omaolo	3
	A couple of participants encountered a question that was formulated in a complex manner and the explanation in parentheses confused the participants even more, “Have you experienced any recent progressive loss of strength in a lower limb or both lower limbs (you cannot support yourself on the lower limb or limbs)”, which led to a situation where the participants misunderstood the question and chose the opposite of what they said out loud they would like to choose.	Omaolo	2

When choosing the municipality from a list of predefined options, a couple of participants did not find the option they were looking for and therefore did not know how to advance. The users could be given the option “Another” which would then explain to the user that the service is not used supported in municipalities not found in the list.	Omaolo	2
Some participants had a hard time choosing the right option in the beginning of the questionnaire between “Ailment or illness” and “Other health center issue”. They were not sure how the options differed and which of the two was the one they were supposed to choose.	Klinik Access	2
One participant complained that there were too many questions and that they all started to look the same. “I get asked a hundred times when and how the symptoms started”. The same question is repeated for each of the symptoms the participant had previously selected but this was not clear to the participant and they felt like answering once was enough. This resulted in the participant giving the same answer to all the remaining questions without reading them properly through and without realizing that each text field is dedicated to each of the previously selected symptoms.	Klinik Access	1
Many participants found the order of the symptoms and other answering options confusing and messy. They felt like there was a lot to process at once and that they had to read a lot in each of the steps. This made the participants’ eyes wander a lot and jump inconsistently from option to option, which then lead to missing and ignoring some options they were supposed to choose completely. When asked if they had noticed that the answering options were in alphabetical order, all the participants said that they had not noticed it.	Klinik Access	5
Some participants did not notice that when choosing the “location” for the symptoms, there was also an option of “No specific location” which could have been selected. The participants did not realize that the offered options were for the same question but instead though that the two options were for different questions. Some participants selected the “No specific location” right after they had just selected the right “location” from the list which then removed their previous selection. Another participant was not sure which option to pick from the list since they had fever as the main symptom and the “No specific location” which is the closest option wasn’t among the other options in the dropdown list.	Klinik Access	3

Guiding the user’s focus	Many participants did not notice that when they selected some symptoms, more options were revealed to them so that they could choose a more detailed explanation of their symptom. These additional dialogues were mainly skipped as the participants did not seem to make the connection between their selection and the newly presented options. Closing this dialogue also shifted the view and some participants were not sure anymore if their selection was recorded properly.	Klinik Access	5
	One participant wondered how many steps and questions they still had left. The indicator at the bottom of the page was not clear or visible enough for them and the indicator also kept changing values depending on the participant’s answers.	Klinik Access	1
	One participant thought they were ready when they had completed only one view and did not understand that the there are more questions once they press “Continue”, here the progress bar stating “3/6” did not seem to be clear or visible either.	Klinik Access	1
Long, uncommon, and compound words	Some participants struggled to tell some similarly sounding words apart and they were mixed and confused with other words. Depending on the situation and the word in question, the participants chose a wrong answer because they failed to read and understand the word properly. For example, “kymmenys/kysymys” (translation: “tenth/question”).	Omaolo	3
	A couple of the words encountered were not familiar to some participants and the participants asked for help in interpreting them. For example, “käypä” (translation: “decent”).	Omaolo	2
	One participant was dyslectic and found challenging to read and understand some of the longer compound words. For example, “laaja-alaisena”, “alaraajat”, “hengitysvaikeuksia”, “leuanalustassa” (translations: ”widespread”, ”lower limbs”, ”trouble breathing”, “under the jaw”).	Omaolo	1
	Some participants struggled to tell similarly sounding words apart and they were mixed and confused with other words. Depending on the situation and the word in question, the participants chose a wrong answer because they failed to read and understand the word properly. For example, “heikotus/heikkous”, “kuume/kuumotus”, “vastaamme/vastaanotamme” (translations: “feeling weak/weakness”, “fever/warmth”, “we answer/we receive”).	Klinik Access	3
	A couple of the words encountered were not familiar to the participants and they asked for help in interpreting them. For example, “pääasialliset” (translation: “primary”).	Klinik Access	2

Input methods	One participant did not remember the name of some medicament so when a question was presented to them asking to name the medicament they were taking, the participant went back to the previous question and changed their answer so they wouldn't need to think of a name. When asking users to name things with possibly complex names such as drugs, implementing a suggestive input or providing a ready list to choose from might make the users' interaction with the system easier.	Omaolo	1
	The participants who encountered images of mouth that they had to choose from were not sure how to make their selection. Two pictures were presented to them and then they were asked to select the more accurate one but the participants had to scroll further down to notice the checkboxes. Initially the participants thought that they were supposed to click on the images and when nothing happened, they were not sure if their selection had gone through or not.	Omaolo	2
	When selecting the age, giving the user more options that they may need confuses them and some participants clicked on years first, then months, and then after realizing their error clicked back to years.	Klinik Access	2
	Many participants were confused when inputting the duration of the symptoms and therefore chose the wrong unit (for example, 3 hours instead of days or weeks). One participant even took out a calculator and calculated three times 24 hours as the participant had first chosen hours as the unit and the symptom duration given was three days. The participant wasn't sure how to change their answer so they calculated the value and inputted 72 hours.	Klinik Access	3
Allowing the user to revert their actions	One participant unintentionally navigated to the identification portal (suomi.fi - tunnistus) and did not know how to return to the previous page as there was no button on the page offering this option.	Omaolo	1
	One participant chose the wrong symptom checker and did not know how to return back to the list of symptom checkers since there was no button on the page with this option.	Omaolo	1
	One participant did not know how to remove their selection once they realized that they had chosen a wrong option.	Klinik Access	1
Consistency of external links	Some of the "internal links" which take the user from one symptom checker to another within the page took the participant from the demo version to the real version of the page.	Omaolo	1

	Some participants clicked on the “Duodecim symptom checker” link on the symptom checker page thinking that said link would take them to the intended symptom checker. However, this is not the case, that link just provides some additional information.	Omaolo	2
	After completing the whole symptom checker and inputting the location information, one participant clicked on the link that should take them to the opening hours and contact info of the nearest health center but the page contains a lot more information that was promised and the information the participant was promised took a while to find since they had to scroll down and look for it for a considerable amount of time. This information was also not very visible on the page.	Omaolo	1
Intuitive and consistent interactions	User thinks they have booked an appointment after identifying using bank account because “why else would you have to identify yourself”	Omaolo	8
	User confesses that didn’t read the introduction parts during their second try, just assumed/expected the content to be the same	Omaolo	8
	Some symptoms open an additional dialogue providing the user with more options to choose from but the button that closes this dialogue has the word “Done” on it which indicates that the section has been completed and that clicking it would advance the user to the next phase. However, clicking the button simply closes the subsection which is not the interaction the participants were expecting. Only one participant actually closed the opened subsection by clicking the button, others left it open and navigated around it.	Klinik Access	7
	One participant clicked on the “None of these” option offered after other options which then cleared all the previous selections. This does not happen in the following section as the same button is situated in the beginning, making the page also a bit inconsistent.	Klinik Access	1
	The “None of these” option was offered in the beginning of the page but once the participant had read the rest of the options, they had already forgotten the first option so when they got to the end and no option matched their situation, they tried to proceed without selecting any of the options. Fortunately, when the proceed button was clicked a clear error message appeared asking the participant to choose at least the first option.	Klinik Access	1
	A couple of participants encountered a situation where they clicked on an option and accidentally dragged the mouse when doing so. This caused the page to select multiple options even though it was not the participants’ intention.	Klinik Access	2

Allowing the user to express themselves freely	Almost all the participants would have liked the option of a text field where they could tell about other symptoms and things that they experiencing that they were not asked in the questionnaire.	Omaolo	7
	Almost all the participants would have liked the option of a text field where they could tell about other symptoms and things that they experiencing that they were not asked in the questionnaire.	Klinik Access	7
Informing the user of errors	With one participant, an unexpected error happened and the page only displayed the text “Something went wrong”. There was no explanation of what had happened or how to recover from the encountered error if the user wanted to continue using the page.	Omaolo	1
	The error text “Please answer all mandatory questions” is not clearly visible and gets cut off if the user does not scroll down enough which caused a couple of situations where the error message was displayed but the participants did not notice it and hence did not know why clicking the continue button had no visible effect and did not allow them to proceed.	Omaolo	2
	After not answering to all the mandatory questions the service does not tell the user clearly which questions still need to be answered. It is up to the user to scroll back up the list of questions and manually locate all the missing answers. The unanswered questions are highlighted but in the case of one participant, they stopped scrolling after reaching the first highlighted question. Then they scrolled back down and when they tried to continue, the page informed them that they still had some questions unanswered. This made the participant frustrated and a bit ashamed that they had missed yet another question.	Omaolo	1
Icons, symbols, and abbreviations	One participant was not familiar with the abbreviation “tms.” used.	Omaolo	1
	Some participants did not understand that the red asterisks after the final questions meant that the section was mandatory to fill in. Instead, they got confused why they could not proceed since to the participants there was no visible reason stopping them from continuing.	Klinik Access	3
Navigation	One participant became exhausted before even reaching the start of the symptom checker since they had to read a lot of text before the symptom checker itself even began.	Omaolo	1

	Some participants clicked and dragged the scroll bar with their mouse in order to scroll down the page and since new questions kept appearing underneath the question the participants had just answered, the participants were required to repeatedly move the mouse back and forth between the scroll bar on the right side of the page and the questions in the middle of the page. This movement was repeated after each question and could be removed if for example the page moved on its own after the question has been answered.	Omaolo	3
	A couple of participants did not notice the “Continue” button in the bottom right corner and asked for help since they did not know how to proceed on the page.	Klinik Access	2
Responsiveness	One participant tried to change their selection but the page did not respond in a reasonable time which then led to the participant getting frustrated.	Klinik Access	1