Marko Ylikulju

PLANNING A SUMMATIVE USABILITY TEST FOR A MEDICAL DEVICE

PLANNING A SUMMATIVE USABILITY TEST FOR A MEDICAL DEVICE

Marko Ylikulju
Bachelor's Thesis
Autumn 2018
Information and Communications Technology
Oulu University of Applied Sciences

ABSTRACT

Oulu University of Applied Sciences
Information and Communications Technology, Option of Device and Product design

Author: Marko Ylikulju
Title of the bachelor's thesis: Planning a Summative Usability Test for a Medical Device
Supervisors: Kari Jyrkkä, Antti Kyllönen
Term and year of completion: Autumn 2018 Pages: 46 + 7 appendices

This bachelor's thesis was commissioned by MoveSole Ltd and the primary goal of this work was to create a handbook of summative usability testing for medical devices. The process started by gathering references from several web-sites, books and standards concerning usability testing and medical device designing.

The handbook was created by collecting the relevant information from the references and by following European medical device regulations. Using the collected guidelines, a summative evaluation plan of usability was created for a MoveSole StepLab smart insole system. The created plan was verified with an internal pilot test.

The author is working as a mobile software developer at MoveSole Ltd having a responsibility for a StepLab mobile application. One reason to carry out this bachelor's thesis was to accumulate usability perspective for the author in a user interface design and a software development.

Keywords: medical devices, evaluation of usability, usability testing, summative evaluation

PREFACE

I would like to thank MoveSole Ltd for hiring me as an intern and providing me the opportunity to finish my studies by commissioning this bachelor's thesis. Thanks to all my colleagues at MoveSole for making a great working environment and special thanks to Minna Lappi and Antti Kyllönen at MoveSole for your help in this work.

I would like to thank Kaija Posio at OAMK for the help with grammar and a correct format in this document. I would also like to thank my supervisor Kari Jyrkkä at OAMK for the assistance through this process.

Oulu, 27.11.2018
Marko Ylikulju

TABLE OF CONTENTS

VOCABULARY

BLE	Bluetooth Low Energy
Harm	A physical injury or damage to the health of people, or damage to the property or the environment
Hazard	A potential source of harm
Hazardous situation	A circumstance in which people, property, or the environment are exposed to one or more hazards
ISO	The International Organization for Standardization
LCU	A least competent user
Manufacturer	A natural or legal person with responsibility for designing, manufacturing, packaging or labelling a medical device
Medical device	An article, apparatus, or machine that is used in the prevention, diagnosis or treatment of illness or disease, or for detecting, measuring, restoring, correcting or modifying the structure or function of the body for some health purpose
SUS	System Usability Scale
UI	A user interface

1 INTRODUCTION

A medical device is defined by World Health Organization as “an article, apparatus, or machine that is used in the prevention, diagnosis or treatment of illness or disease, or for detecting, measuring, restoring, correcting or modifying the structure or function of the body for some health purpose” (26). Devices intended for medical use must be evaluated with a summative evaluation of usability to get a medical device approval inside the European Economic Area (10, p. 6).

A summative usability testing is a method of usability evaluation conducted in a “summation point” of the design. The primary goal of a summative evaluation of usability is to get an objective evidence that the medical device is safe to use but also, to ensure that the device offers superior usability (10, p. 6).

The required theory and regulations to plan and conduct a summative usability test of a medical device are collected into a handbook of planning and conducting a usability test of a medical device, following European medical device regulations. The studied theory was applied to create a summative usability test plan for a MoveSole StepLab system. The aim was to create the test plan, verify the plan in a pilot test and then conduct actual usability tests with a sample of test participants.

MoveSole StepLab is a mobile force measuring system consisting a smart insole and a smart device with a StepLab application. StepLab is a product of MoveSole Ltd which is a small health technology company located in Oulu, Finland. Although the first version of StepLab is not intended for medical use and therefore the summative evaluation of usability is not mandatory, the evaluation wanted to be done so that the safety and good user experience could be ensured. Also, the plan will be needed if the company decides to apply for a medical approval to next versions of the product.

2 TERMS AND DEFINITIONS OF USABILITY

In ISO 9241: Part 11: Usability: Definitions and concepts (2018) usability is defined as follows: “Usability: the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (17, p. 6). In this chapter, relevant terms to usability in general are discussed.

2.1 Usability Engineering

Usability engineering (or human factors engineering) is a user-centered design process which helps a manufacturer to anticipate possible usability problems, solve existing problems and provide a better custom satisfaction in general. Ideally, usability engineering should be a continuous process throughout the development cycle of the product. (9, p. 3–4)

In a medical device standard EN-62366-1: Medical devices – Application of usability engineering to medical devices usability engineering is defined as an “application of knowledge about human behavior, abilities, limitations, and other characteristics to the design of medical devices (including software), systems and tasks to achieve adequate usability”. A manufacturer of medical devices is obliged to establish, document, implement and maintain a usability engineering process to provide safety for the patient, user and others. (11 p. 11–13)

2.2 Components of Usability

As stated in the standard above and according to usability specialists Jeff Rubin and Dana Chisnell, a usable product should be useful, efficient, effective, satisfying, learnable and accessible (2, p. 4).

- Efficiency (the extent to which users expend resource in achieving their goals)
- Effectiveness (whether users can actually complete their tasks and achieve their goals)

- Satisfaction (the level of comfort that users experience in achieving those goals)
- Learnability (whether users can quickly become familiar with the product and make good use of all the features and capabilities)
- Accessibility (whether users can easily understand and obtain the products features) (3, p. 1).

Another widely recognized usability specialist Jakob Nielsen defines five most important components of usability that slightly differ from the attributes listed above.

- Learnability (How easy it is for users to accomplish basic tasks the first time they encounter the design?)
- Efficiency (Once users have learned the design, how quickly can they perform tasks?)
- Memorability (When users return to the design after a period of not using it, how easily can they reestablish proficiency?)
- Errors (How many errors do users make, how severe are these errors, and how easily can they recover from the errors?)
- Satisfaction (How pleasant is it to use the design?). (1.)

The model of system acceptability attributes by Nielsen shows the same attributes of usability as listed above, but also the definition of usefulness and attributes of practical acceptability are covered. This simple model shows (see the figure 1 below) that usability must trade off against many other considerations in a development project. (4, p. 25)

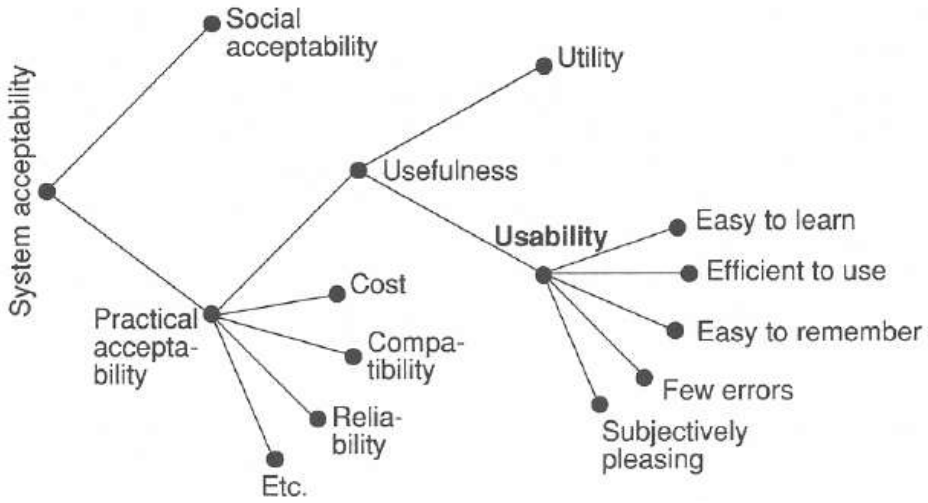


FIGURE 1. Quality Attributes of System Acceptability (4, p. 25)

2.3 Usefulness, Utility and Usability

Nielsen defines utility as a number one key of quality attributes of usefulness. Utility refers to the functionality of the design: Does the system do what users need?

Usability is the other key that along with utility makes the product useful. If the system can hypothetically do what a user wants but is extremely difficult to use, the user will most likely not use it. Or, if the system is easy to use but it cannot do what the user wants, it is not useful. Nielsen defines the terms as follows:

- "Utility = whether it provides the features you need
- Usability = how easy & pleasant these features are to use
- Useful = usability + utility". (1)

2.4 User Experience

User experience is often confused with usability and in some contexts, they can have the same meaning. But several experts recommend distinguishing the terms since usability is considered as just a part of user experience (see the figure 2 below) (5; 6).

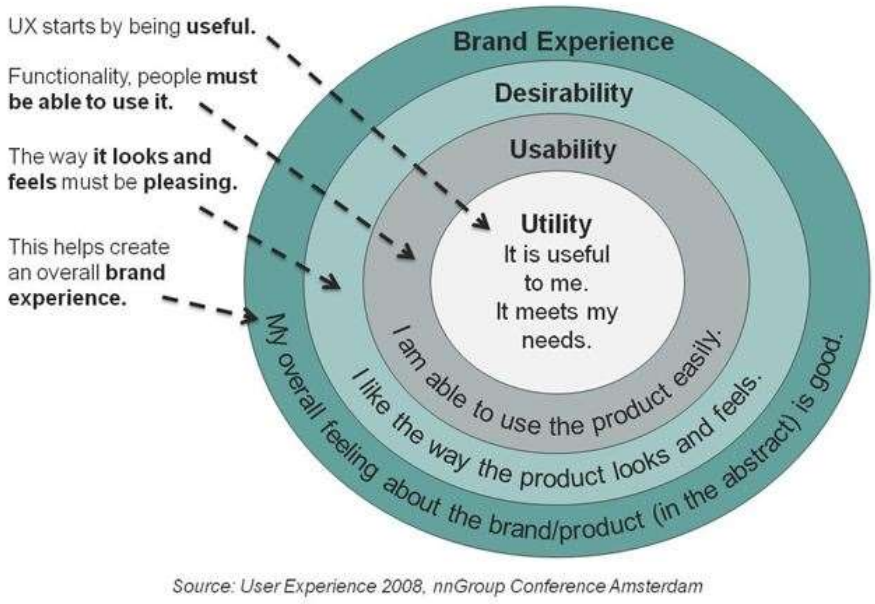


FIGURE 2. Usability is a Part of User Experience (25)

While usability focuses on actual use of a system, user experience takes it further by looking at the processes before and after the use activity, including all aspects of the end-user's interaction with the company, its services and its products (6). For example, if we are concerned about usability of an online shop, it means that we measure how effectively and easily the user can purchase the product they want, and how pleasant it is for them to use the design of the webpage. Whereas user experience also covers the promotion of the service, delivery time of the product and other actions around the webpage itself.

2.5 Matter of Context

A usability specialist Brooke emphasizes the importance of context in usability (3, p. 4). It is comprehensible that designing a UI for a tablet with a 10-inch display is different from designing a UI for a smart watch with a tiny display of 2 inches. Or developing a programming software for other software developers in comparison to creating a web-page allowing senior end users to check their bank balance.

So, depending on the goals, environmental variables, types of users and other components of the context of use, using the same system can result in significantly different levels of usability. The figure 3 below shows the components of context of use and the other outcomes along the usability such as accessibility, user experience and avoidance of harm from use. (17, p. 12)

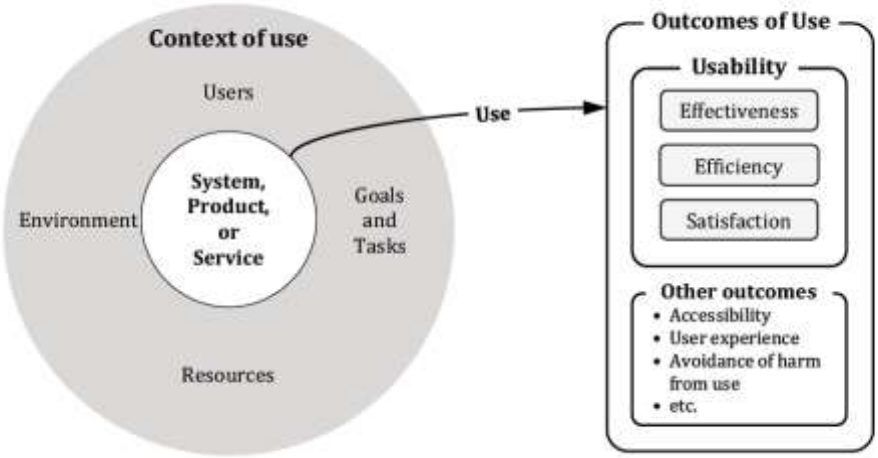


FIGURE 3. Usability in a Context of Use (17, p. 12)

In case of medical devices, usability is especially linked with safety because along with all the other factors an evaluation of usability can judge if the device is either resistant to or vulnerable to dangerous use errors that could lead to a user or patient injury or death (7, p. 2).

3 REGULATIONS AND STANDARDS

Standards and directives that regulate medical devices focusing on subjects relevant to the StepLab system are more widely discussed in Miro Kuusijärvi's bachelor's thesis (8). In this chapter, the content is focused on the standards relevant to usability and the usability evaluation of medical devices.

3.1 Standards of Usability

ISO 9241: Ergonomics of Human System Interaction is a set of standards regulating the usability of systems in general. These standards are widely cited, and they work as a guideline for UI developers and usability evaluators.

The part 11 of ISO 9241 Usability: Definitions and Concepts provides the definition of usability and makes the point that the usability is dependent on the context including the task that is being done, the background and experience of the user, and the environment in which it is being done (3).

Other relevant ISO standards and sets of standards related to usability and ergonomics are the following:

- ISO 6385: Ergonomic principles in the design of work systems
- ISO 10075: Ergonomic principles related to mental work-load
- ISO 11064: Ergonomic design of control centers
- ISO 13406: Ergonomic requirements for work with visual displays based on flat panels
- ISO 13407: Human-centered design processes for interactive systems
- ISO 14915: Software ergonomics for multimedia user interfaces
- ISO/TS 16071: Ergonomics of human-system interaction
- ISO/TR 16982: Ergonomics of human-system interaction
- ISO/TR 18529: Ergonomics of human-system interaction
- ISO/IEC 9126: Software engineering – Product quality
- ISO/IEC 11581: Information technology – User system interfaces and symbols – Icon symbols and functions

- ISO/IEC 15910: Information technology – Software user documentation process

This list contains just some of the all international standards related to usability. The manufacturer of a product must be aware of local regulations and purchase and comply with standards that are relevant to the product under development. Since a MoveSole StepLab system uses a smart device the company must comply with standards regulating human-system interaction, software engineering, and user system interfaces and symbols for example.

3.2 Usability Standards of Medical Devices

EN 62366-1: Medical devices – Part 1: Application of usability engineering to medical devices is a European standard regulated by CENELEC (European Committee for Electrotechnical Standardization). The standard is intended to identify and minimise use errors and reduce use-associated risks in use of medical devices. It focuses on optimizing usability as it relates to safety, but also on how usability relates to attributes such as task accuracy, completeness and efficiency as well as user satisfaction. (10, p. 6)

This standard offers guidelines for the UI design and software development, but it also requires the manufacturer to conduct evaluations of usability for its product. The standard defines a formative evaluation as follows: “A user interface evaluation conducted with the intent to explore user interface design strengths, weaknesses and unanticipated use errors”. A summative evaluation is defined in the standard as “A User interface evaluation conducted at the end of the user interface development with the intent to obtain objective evidence that the user interface can be used safely”. (10, p. 9–10)

The standard requires the manufacturer to conduct the usability engineering process at nine stages:

- Preparing use specification
- Identifying UI characteristics related to safety and possible use errors
- Identifying known or foreseeable hazards and hazardous situations
- Identifying and describing hazard-related use scenarios

- Selecting hazard-related use-scenarios for a summative evaluation
- Establishing a UI specification
- Establishing a UI evaluation plan (formative and summative evaluations)
- Implementing and verifying the UI
- Validating the UI

Validating the UI requires that the stages listed above are conducted and the outputs are documented in a usability engineering file. The file can be a part of the risk management file or an independent document. The usability engineering file enables an efficient auditing of the development process by containing at least references or pointers to all required documentation. (10, p. 23)

4 USABILITY EVALUATION METHODS

Usability testing is widely recognized as the most comprehensive and effective method to evaluate the usability and to reduce use-related risks. However, certain types of use errors are unlikely to occur during a usability test because of relatively small sample sizes and the nature of the testing event. Thus, usability inspection methods, such as a heuristic evaluation, a cognitive walkthrough, or other evaluation methods are recommended to be conducted in addition to the usability testing. (7, p. 35)

A StepLab system has been evaluated with healthcare professionals along the way of developing the product and a few formative usability tests have been done to the product. Earlier usability evaluation of StepLab is discussed in the chapter 6.3.

4.1 Heuristic Evaluation of Usability

Nielsen published 10 usability heuristics for user interface design in 1995. These heuristics, such as a checklist in evaluating usability of interfaces, are widely known and used. Nielsen's ten heuristics are:

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation.

Some of the issues can be easily unseen by novice users, therefore some expertise of usability evaluation may be required to do the heuristic evaluation. (11)

4.2 Cognitive Walkthrough

A cognitive walkthrough is a method in which one or more evaluators work through a series of tasks that the user would be expected to carry out and ask a set of questions from the user's point of view (14). The process starts by writing a list of the actions needed to complete a single task with the UI and doing so for all selected tasks. If the list of actions to complete one task is too long (depending on the context, ten steps is probably too much), it can be already said that there is a usability problem with that UI. (15)

If the length of the list is reasonable, going through those steps and then asking questions from the perspective of the user can reveal flaws in the design. Typically, four questions are asked:

1. Will the customer realistically be trying to do this action?
2. Is the control for the action visible?
3. Is there a strong link between the control and the action?
4. Is feedback appropriate? (15)

4.3 Formative Usability Testing

A formative evaluation of usability is not mandatory to be done, but it is highly recommended as it can offer really valuable data along the way of development of the product. Generally, it is performed as usability tests iteratively throughout the design and development process. (10, p. 9) The tests can be approached in a casual or formal manner and test planners may decide how many participants they want in each test. A usual and recommended pattern is to start from around 6 participants and increase the number of participants with each formative usability test, so that even small errors will be noticed as the design evolves. (7, p. 90)

A formative evaluation of usability does not have any formal acceptance criteria, but the manufacturer should set a quality level for the UI to be achieved so that the final summative evaluation of the usability can be conducted successfully (10, p. 35).

4.4 Summative Usability Testing

A summative evaluation of usability is a formal evaluation having formal acceptance criteria. It can be conducted in a “summation point” of the design – when the product is considered complete, production ready, and the formative evaluation of usability is done. The primary goal of the summative evaluation of usability is to get an objective evidence that the medical device is safe to use. This means that the chance of committing dangerous use errors is minimized. (7, p. 91)

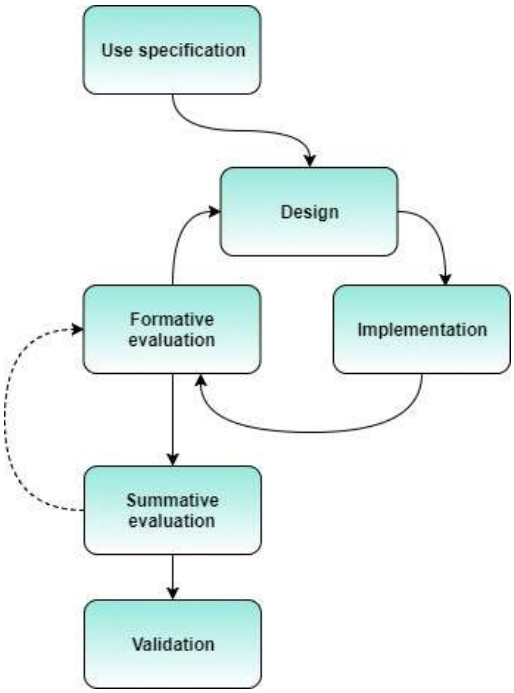


FIGURE 4. UI Evaluation Cycle

Summative usability testing is done for selected hazard-related use scenarios which adequately represent the actual conditions of use. A single test event or multiple tests may be done, depending on the nature of the product and variation of use cases. (10, p. 34) If the manufacturer discovers that some of the risk control measures in the UI are not effective, the summative evaluation becomes, in effect, a formative evaluation and the UI evaluation cycle returns to the point of formative evaluation as described in the figure 4 above (10, p. 37).

If the product passes the summative evaluation, the UI of the system can be validated as safe to use. This validation is a part of the activities involved in verifying and validating the overall medical device design.

5 USABILITY TESTING

This chapter provides a simple guideline of how to create a test plan, recruit participants, conduct a test, and analyze the data and finally write results to a usability test report. The theory in this chapter was implemented in a summative usability test of StepLab, which is discussed in the chapter 7.

Well planned usability testing covering all the relevant stages can be conducted in five weeks as shown in the figure 5 below.

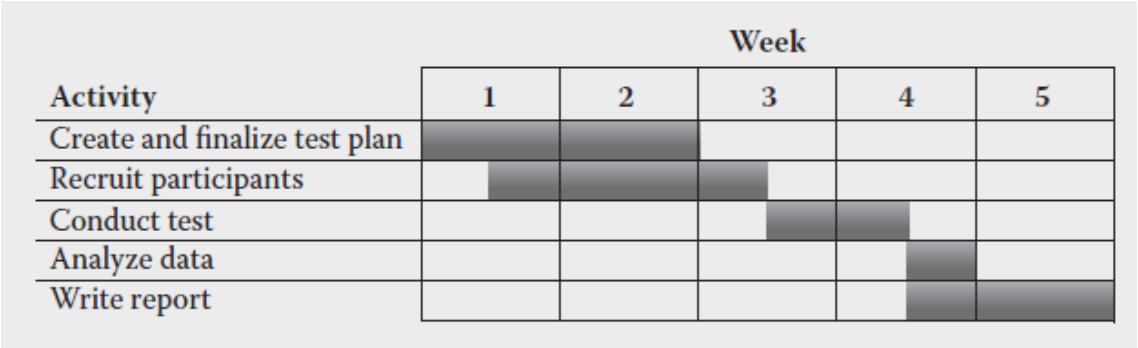


FIGURE 5. The Timetable of Arranging and Conducting Usability Tests (7, p. 41)

5.1 Creating and Finalizing the Test Plan

5.1.1 Choosing the Test Location

A summative evaluation should be conducted in adequately representative actual conditions of use. It can mean that the testing is conducted in a clinic where the system is going to be used, or in a usability lab that is built to imitate the actual environment of use. The advantage of conducting the usability testing at the clinic is that the conditions are already on place, and the participant is more likely going to do things in the same way he or she would do them normally.

On the other hand, if the evaluation is supposed to be done with 15-25 participants, video recorded and observed by a group of people, it may be easier to organize the tests in a usability lab. Instead of asking permissions, travelling to different clinics, and setting up the equipment repeatedly, equipment could be set up in the usability lab once and participants could be invited in there alternately.

5.1.2 Test Personnel

The least personnel needed to conduct a usability test is a moderator and an observer. The number of observers depends on the nature of a tested system and testing environment, but more than one is recommended to get more comprehensive notes. If the test is going to be video recorded and/or captured by a screen capturing software, such as Morae, probably one person is needed to use the recording equipment. Especially, in the formative usability tests in an early phase of the product development, a technical expert might be needed in case of something goes wrong with the product during the test.

A moderator (or a test administrator) is a person who is leading the testing event and interacting with the participant. The moderator can be in the same room with the participant (see the figure 6 below) or in another room giving instructions via speakers. The moderator should not interfere the participant or provide help unless it is necessary to continue the test because the aim of usability testing is to observe representative users interacting naturally with the device and to identify UI design strengths and shortcomings of the device. (7 p. 266) Depending on the environment and type of the test, the moderator might start the video recording, or a particular person can be named to use the recording equipment. Either way, it should be careful not to start recording before required documents are signed to have a permission to film the participants.

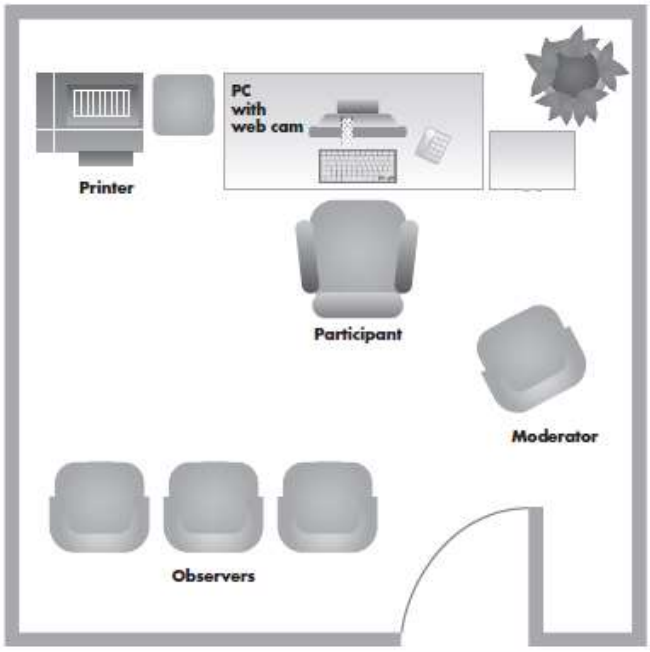


FIGURE 6. Simple Single-room Setup of a Usability Test (2, p. 102)

Observers' job is to take notes concerning the usability of the product. They are watching how easily the participant can use the product and they are looking for potential obstacles to carry out different actions. Research questions and benchmarks help the observers to look in the right direction (2, p. 71). Observers should stay as imperceptible as possible so that they do not bias to the participants' working. They may be in the same room with the participant or in another room watching the test via a screen or a one-way mirror. Observers can be developers of the product, other personnel of the company or external people hired to observe the test event.

A data gatherer (or a note taker) is like an observer but has a particular job in taking notes such as counting errors or taking time on task completion. In many cases, the observers are taking the notes and no data gatherers are named particularly.

5.1.3 Choosing the Tasks

In formative usability testing, tasks can be selected to find out if specific elements of the UI work or if they need to be refined. Tasks should be selected so that the participant will use the parts of the system in question, without telling that directly

in the task. In the end, the testing is supposed to measure the relationship of the system to the end user and expose possible usability flaws in there. (2, p. 82)

A good example of designing the tasks is introduced by a usability expert Jared M. Spool: While testing a website of Ikea.com, instead of saying to the participant "Find a bookcase", it was changed to a form "You have 200+ books in your fiction collection, currently in boxes strewn around your living room. Find a way to organize them". With the first task definition, all participants just typed "bookcase" to the search box and got the same results as everyone else. The second case revealed differences between participants as they started to navigate on the categories of the website or used search typing in different keywords such as "shelves" or "storage systems", but nobody searched on "bookcase". (12)

In summative usability testing, the ideal situation is that the participant would have to interact with the system in every conceivable way. However, in most of the cases, it is not possible because of complexity of the system and limited time of testing. Especially with the medical devices, prioritizing the tests should focus on safety-critical tasks. The selected tasks should be linked to risk management and analysis efforts. (10, p. 204)

5.1.4 Pilot Testing

When the test plan is ready, it is recommended to test it with a pilot test. Pilot testing helps the test administrators to identify the flaws in the test plan and makes the actual testing event to proceed more smoothly and effectively.

A pilot test might lead to changing the order or details of test tasks to ensure a realistic workflow. Pilot test results can also help to adjust the recording equipment to do a better job by capturing critical details. In addition to all the adjustments and improvements to the test plan and papers such as forms and questionnaires used in testing, a pilot test enables to estimate how long it takes to conduct the test with one participant. The pilot testing is also important so that the test administrators, observers and other personnel can practice their roles and the workflow in upcoming usability tests. (7, p. 244)

5.2 Recruiting Participants

5.2.1 Number of Participants

A sample size of the first formative test could be just around five participants, since it is widely recognized that just five test sessions can generate many of the findings that could result from a much larger study (10, p. 122). Then in later evaluations, it is recommended to increase the sample size to ensure that the required level of quality of UI is achieved before conducting the summative evaluation of usability. The more participants, the more precise evaluation of usability it will produce. However, conducting moderated usability tests is relatively expensive, which is the reason usability researchers are willing to estimate the required number of participants and there is even a mathematical formula for calculating the estimation (19).

According to medical device regulators, 15-25 is a reasonable number of participants in a summative usability test, presuming a reasonably homogeneous user population. It should be anticipated that almost 10% of the test participants will cancel their scheduled appointments or simply not show up. (10, p. 122–123)

5.2.2 User Profile

“A user profile is a summary of the mental, physical and demographic traits of an intended user group, as well as any special characteristics, such as occupational skills, job requirements and working conditions, which can have a bearing on design decisions” (10, p. 11).

The participants of a user group should be picked carefully so that they are in a target group of the system of users but cover different backgrounds, as well as different levels of experience in using such systems. If the intended users of the product have potential disabilities, such as diabetics often have poor visual acuity and poor sense of touch, it should be taken into account.

If the risks of using the device are not too serious, it can by good practice to include a few LCUs among participants, i.e. end users who represent the least skilled persons who could potentially use the product. If LCUs can make it through

the test, it can be inferred that the actual professionals are able to use the product as well. On the other hand, if LCUs struggle with some of the tasks, it is not necessarily an indictment, but it might reveal clues on where the UI can be simplified or improved. (2, p. 146)

If using the product requires training or if the great majority of users in the field will be trained before using the device, the training should also be applied before the usability test. In some cases, the training can be applied only to some of the test participants to get the representative sampling of the real-life situation in the test. The training before the usability testing should not be better than normal training so that it would boost the participants' performance in the usability test. (7, p. 274–275)

After providing the training to a test participant, the testing team should allow some time to pass before conducting the usability test event. A delay can be something between hours to a couple of weeks. The purpose is to ensure that the focus of the test is on the usability strengths and shortcomings of the device rather than on the participant's near-term memory. (7, p. 276)

5.2.3 Screening Questionnaire

One way to find suitable participants is to document the user profile and generate questionnaire to qualify the participants by their background and level of experience. A questionnaire can be used to exclude unqualified participants but also to find participants with a different background and level of experience to make the test more comprehensive.

5.2.4 Inviting Participants

Once the user profile is ready, it is time to invite the participants utilizing the screening questionnaire. Recruiting suitable participants may be challenging and may also require some compensation to be paid to the participants, although many participants are more likely motivated because of an opportunity to help improve the safety and usability of a new medical device. Recruiting can be done by phone, e-mail or in a face to face interview.

5.3 Conducting the Test

5.3.1 Pre-test Questionnaire

A pre-test questionnaire (or a background questionnaire) is intended to provide historical information about the participants that will help the evaluators to understand participants' behavior and performance during a test. It is similar to the screening questionnaire but just goes further by exploring previous training and experience. Sometimes the experience of using similar devices before may affect their performance negatively, sometimes positively but almost certainly it affects, and it is important to know when conducting the tests. (2, p. 162)

5.3.2 Orientation Script

An orientation script is a starting speech of the test moderator, the orientation script explains for the participants how the test is conducted and why it is recorded. Usually the participant is asked to fill in a pre-test questionnaire and sign a nondisclosure agreement and recording consent form. The speech is intended to put the participant at ease, as they can be nervous because of the testing environment and recording equipment. Reminding about the fact that the product is being tested, not the participant, might help them to relax. A list of guidelines in developing an orientation script is introduced by Rubin and Chisnell:

- Keep the tone of the script professional, but friendly
- Keep the speech short
- Write the orientation script out
- Plan to read the script to each participant verbatim
    - This is important so that all the participants are exposed to identical conditions prior to the test
- Explain why the participant is here
- Describe the testing setup
- Explain what is expected of the participant
- Allow and ask for any questions from the participant
- Refer to any forms that need to be completed and pass them out. (2 p. 161)

5.3.3 Running the Test

After the required forms are filled and orientation script read by the moderator, it is time to conduct the test itself. In a formative usability test, the moderator will encourage the participant to think aloud during the test but in a summative usability test the participant should focus on doing the tasks as they would normally do without distractions. Depending on the nature of the test, participant should say “done” or give another agreed sign after they are done with each test task or all the given tasks.

5.4 Analyzing the Data

5.4.1 Research Questions and Benchmarks

Defining research questions and benchmarks before conducting the test is crucial as it gives a purpose for the testing. Research questions help observers to look in the right direction during the test and they should consider hazard-related use scenarios and possible flaws in the design. Research questions for a software interface could be “How easily and successfully do users find the tools or options they want?” or “Is the response time a cause of user frustration or errors?” and benchmarks to measure those could be counting clicks and indications of frustration on completing a task. (2, p. 71)

Usually, the UI is measured by timings and facial expressions together with task completion times, incorrect actions, requests for assistance, task abandonments, and indications of frustration, confusion, or declining morale (10, p. 266). The benchmarks should be thoughtfully selected beforehand so that they actually help to evaluate the usability of the product and reveal the potential flaws in the UI.

5.4.2 System Usability Scale (SUS)

The System Usability Scale (see the appendix 4) has become a standard method to measure the usability of systems although it has never been through any formal standardization process (3, p. 1). It is originally introduced by a usability specialist John Brooke in 1986. The questionnaire has total 10 questions in a form of Likert scale. Each question has a 5-point answer scale from “strongly agree” to “strongly

disagree”. A half of the questions are formulated negatively and a half of them positively, thus the respondent must read questions and answer to them thoughtfully. Questions in the scale are selected carefully to capture a general agreement of extreme attitudes instead of ambiguity of answers. The selected questions are giving a global view of subjective assessments of usability. (18, p. 191)

While the original version of SUS is a standard and modifying it should be avoided to protect the validity of it, some usability specialists recommend replacing the word “cumbersome” with the word “awkward” in the statement 8, especially if the respondents are non-native English speakers (21, p. 2; 22, p. 2).

5.5 Writing the Report

A usability test can produce plenty of data to be reduced and analyzed. This data should be summarized in a usability test report. The report may contain questionnaire results, diagrams, and conclusions as well as comments from test participants about the product. A simple calculation of means and standard deviations usually reveals flaws in the usability and, a compilation of the participants’ comments often points directly to the weaknesses of the product and provides recommendations to solutions as well. (9, p. 279–280)

The report should be distributed to all members of the design team as well as to marketing personnel who are introducing the product to customers. Problems and findings cited in the reports should be resolved. Sometimes, it is reasonable not to fix a problem after considering the benefits and costs of doing so. The product should be retested until it is free of most usability problems and not significant changes are done after the last test. The next step is the production of a final prototype that will go through a summative evaluation of usability to get the objective evidence that the product is usable and safe to use. (9, p. 281)

6 OVERVIEW OF THE PRODUCT

MoveSole StepLab is a gait measuring system which consists of a smart insole(s) to the shoe and a smart device with a StepLab application (23). The StepLab product and its package content (see the figure 7 below) is described in this chapter.

FIGURE 7. A StepLab Package Content

A StepLab package includes 6 pairs of smart insoles in EU sizes ranging from 36 to 46, the smart device and a charger, a screwdriver, a user manual, and spare

parts to the smart insoles such as coin cell batteries and ankle bands to hold electronics casing in place (see the figure 8 below).

FIGURE 8. Ankle Band of a MoveSole Smart Insole

6.1 Smart Insole

The MoveSole smart insole has seven individual sensors measuring forces between the human sole and the ground. Printed intelligence production method together with unique sensor technology allows the thickness of the smart insole to be less than 3 millimeters which makes them indistinguishable underneath the feet (see the figure 9 below). The sensor data is processed in the embedded circuit of the smart insole and sent wirelessly to the smart device to be shown as steps. (23)

FIGURE 9. A MoveSole Smart Insole

6.2 Smart Device

The smart device is delivered in the package with the StepLab application installed into it. The smart device is in so called Kiosk mode, which means that the user cannot exit the StepLab application or use other features of the smart device than the application and its features. The kiosk mode was applied to ensure the stability of the application and confidentiality of user information.

The StepLab system uses the Samsung Galaxy XCover 4 smart device, with the Android 7.0 Nougat version. The minimum requirements to the smart device were BLE compatibility, the Android version 5.0 or later and a physical durability (waterproof and shockproof design).

MoveSole smart insoles are intended to be used with MoveSole StepLab application. The step data is sent from the smart insoles inside BLE data packets, decoded in the application and shown on the screen in real-time and also in a session summary and a session report afterwards.

A UI of the application was designed to be as simple as possible so that people, who do not necessarily have a lot of experience about smart devices, could still learn to use the application. Simple design also allows the user to get the relevant information at a glance.

6.3 Earlier Evaluation of Usability

The aim has been that the UI of the StepLab system would not have unnecessary settings or buttons, instead all the basic functions should work just by a maximum of a couple of clicks. This idea came clear in the early phase of product development when the product was tested in Tampere University Hospital. From the very beginning, StepLab has been tested and evaluated by medical professionals and physiotherapists including several universities and university hospitals in Finland. All the relevant feedback and suggestions are gathered to develop the usability of the system. (24)

Three formative usability test events were done with the product before this thesis was started. The first one was in Oulu in April 2017 medical students being as main participants. The second one was in Helsinki in June 2017 with a group of foot therapist students. (8, p. 52) The third formative usability test was held in the eHealth event in Oulu in April 2018 with medical and nursing students and it was moderated by the author of this thesis.

6.3.1 Improvements After Usability Evaluations

The first two usability tests of the StepLab system were ‘quick and dirty’, therefore, the documentation was done at the minimum level. The tests produced valuable knowledge of the development of the product and many improvements have been done since. The third test included more precise planning and documentation which resulted in a great score for the usability of the system, providing improvement suggestions only to a user manual (20, p. 19). That indicated to the development team that the product would soon be ready for the summative evaluation of usability.

One finding in earlier evaluations of usability was that a StepLab application would need better navigation tools. Multiple test participants were being stuck in some screen of the application and they did not realize they could go back by using capacitive buttons of the smart device. (8, p. 53) A current UI of the application has back button on every view that has option to go back to previous view.

Also, a title of a current view is shown in a standard toolbar to help in navigating. (see the figure 10 below)

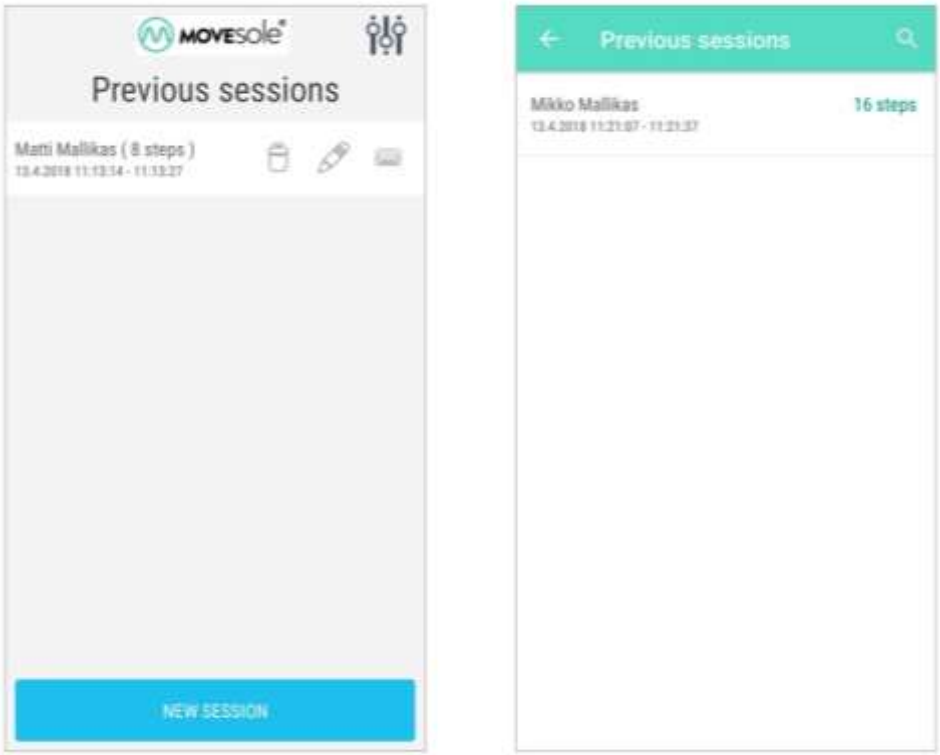


FIGURE 10. A Standard Android Toolbar Was Added to the Application to Help in Navigation

Another example of findings is related to colors of the UI. Previous design utilized color scale from green to red in expressing force distribution in measurement results. That resulted in users thinking that red color was somehow indicating wrong or bad style of walking, whereas they thought green would be right or good style of walking. (8, p. 53–54) Because the system cannot tell user what the correct way to walk is, but it is rather individual trait and must be evaluated by professionals, the colors of the force distribution scale were changed into different shades of blue. (see the figure 11 below)

FIGURE 11. The Session Summary View of a Previous and a Current UI

One finding was not found in formative usability tests but when providing demo products for potential customers. That was, users had problems to find smart insoles in the application. Either they did not understand that they needed to activate the smart insoles by walking or they did not use enough force to activate them. To improve a usability of the product, a threshold to activate insoles from sleeping mode was lowered and a hint text was added to a smart insole selection window in the application (see the figure 12 below).

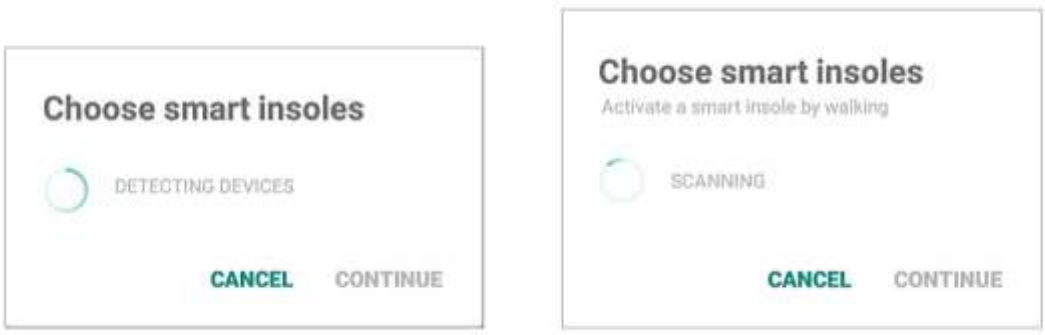


FIGURE 12. A Hint Text was Added to a Choose Smart Insoles Window

7 SUMMATIVE USABILITY TEST PLAN OF STEPLAB

This thesis was intended to produce a plan of summative evaluation of usability for the StepLab system and conduct a pilot test along the plan. This chapter discusses a process of planning the test based on the theory considered in this thesis. A final test plan document can be found in the appendix 7. For confidential reasons, some parts of the test plan were omitted from the test plan document in this thesis.

7.1 Recruiting the Participants

Since the system uses an Android based smart device, the participants should have different levels of experience in using smart devices. The user profile should contain both people that are used to operate with an Android and people that are used to operate with an iOS or other mobile operating systems. Also, different levels of experience in a gait analysis and healthcare in general are required as well as variation in participants ages.

The user profile was defined by creating a table of attributes (see the table 1 below). The aim is to find a relatively diverse sample of participants by asking screening questions and filling in the table. The more diversity in the qualified participants, the more reliable the usability test will be.

TABLE 1. Screening Question Answer Table

Profession or field of study	Physiotherapist	Foot therapist	Doctor
Does walking tests or gait analysis	Once in month or less	More than once in a month	Weekly, more often
Age	< 30 years	31 - 45 years	> 46 years
Experience in healthcare	0-3 years	3-10 years	> 10 years
Familiar with Android devices?	Does not own smart device	Uses smart device with other OS daily	Uses Android smart device daily

Participants are recruited through known contacts of the company in Finnish hospitals and clinics. Primary ways of contacting are e-mails and phone calls. It is known that resources of the company are limited and recruiting a diverse sample of required 15-25 participants may be challenging. Adding realism, the actual testing can be conducted with around 10 participants with an adequately diverse user profile.

The users will receive the StepLab system and they are free to read a user manual and to try the product on their own or with their patients before the testing event. No training is provided by the manufacturer in this case. The tests will be conducted at clinics of the recruited participants with their actual patients. This ensures that the test is done in the actual conditions of use as stated in the chapter 5.1.1. Permissions to film the participant and the patient will be needed. Thus, forms of consent can be found in appendices (see the appendix 2; 3).

7.2 Defining the Tasks

The primary operation functions of the StepLab system are listed in a usability specification document. From those functions, hazard-related use scenarios were chosen as summative usability test tasks. Tasks should not point directly on actions that participants are expected to do, but they should rather provide a realistic scenario of where the product could be used.

A four phased test case was defined for the testing. Participants will be given time to train with the product and an accessibility to the user manual. One of the tasks in the test is open for different use cases of the product. They can be, for example, operability of an off-loading insole or analyzing walking symmetry. The test task list is intentionally left out from this document.

7.3 Defining Test Personnel

Because the StepLab summative usability tests are conducted at clinics of the recruited participants, test personnel are limited to two persons in a room. One of them is operating the video camera and making notes on an observer form. The other one is acting as a moderator and also as an observer. The moderator will use the orientation script (see the appendix 5). The script is supposed to keep the conditions identical between participants prior to the test.

7.4 Defining Research Questions

Research questions were chosen according to the defined test tasks and hazard-related use scenarios listed in a risk management file of the StepLab system. Potential phases for use errors or other issues are listed in the table 2 below.

TABLE 2. Research Questions

Initialization	• Identify obstacles to replacing the battery of a smart device • Identify obstacles to wearing smart insoles to the patient

	• How easily do users find the smart insoles on the smart device? • Identify obstacles to set alert threshold levels on a measurement
Measurement	• Identify obstacles to complete the measurement of symmetrical walking • Identify obstacles to complete the measurement with an alert threshold set
Results	• How well do users understand the session summary data? • How well do users understand the report data? • Identify obstacles to store pdf report files on a computer • Identify obstacles to delete all sessions from the smart device?

7.5 Setting Benchmarks and Collecting Results

The benchmarks were set to reflect to the research questions. Observers will tally errors and indications of frustration during the session and fill in a benchmark table on an observer form.

The observer form (see the appendix 3) contains the benchmark table, a table of task completion times, and a field for user comments and other notes. The observers are encouraged to write down all relevant comments from participants concerning the usability of the system. Completion times can be checked from video recordings after the tests have been run through.

After the test, the participant will be asked to fill in a SUS form (see the appendix 4) and they are interviewed in a post-test interview (see the appendix 6). These results along with the observer form and pre-test questionnaires data are used to analyze test results and make conclusions about the usability of the product. A SUS score is calculated as instructed in the chapter 5.4.2. Number of errors,

comments, task completion times and other measures will be visualized in a test report and the conclusions can be made after.

7.6 Pilot Testing

A pilot testing was done to verify the test plan and to get an estimation of how much time the test will take. But also, to practice roles of test personnel for the actual testing event. The first round was conducted as an internal test where employees of MoveSole Ltd were acting as test participants. After the internal test, detected flaws in the test plan were fixed before the test plan document was approved by a project manager.

7.6.1 Changes in an Orientation Script

In a pilot test, a test administrator accidentally jumped over one part of an orientation script going straight from an introduction to the task one without providing a participant information form to the participant. Lines of the script were numbered in order to clarify the order.

It was noticed in the pilot test that in the orientation script the participant was only asked to sign the participant information form, but the script did not mention that the form also has several questions about the participant's background. Question was reformed respectively. Also, the following comment was added to the orientation script after the pilot test: "In a SUS questionnaire you are asked to mark your immediate response to each item, rather than thinking about items for a long time".

7.6.2 Changes in Other Forms

Another issue noticed during a pilot test was that an observation form was missing a benchmark question if a participant has problems to delete a session on a smart device. The question was added in an updated plan. A few other, mostly just spelling corrections were made to test tasks and documents translated in Finnish accordingly.

The pilot test generated a conversation about a post-test questionnaire and its open-ended questions. It was mentioned by a test participant that many people

are afraid of answering open-ended questions in writing or they do not remember some issues they were facing during the test and therefore the observations can miss from test results. That is why the post-test questionnaire was reformed to a post-test interview with mock-up pictures of a StepLab application (see the appendix 6). A moderator will show the mock-up pictures and ask questions from the participant and the participant can point out the phases where they were facing issues or success during a test.

8 CONCLUSION

The objects of this bachelor's thesis were to create a handbook of usability testing for medical devices and plan and conduct a summative usability test for a MoveSole StepLab smart insole system. The primary objectives of this thesis were met but due to the shortage of time, the test plan of StepLab was not put into practice by conducting usability tests with actual test participants and patients. The tests will be carried out as soon as suitable participants have been recruited.

As a result, this bachelor's thesis provides a comprehensive guide of planning and conducting a usability test for a medical device taking European medical device regulations into account. Also, the summative usability test plan of StepLab was created and the operability of the plan was verified in an internal pilot test. The test plan was written in English but all the appendices of the document were also translated into Finnish. MoveSole can use the plan to carry out usability tests for the StepLab product and this bachelor's thesis can work as a handbook if the plan needs to be adapted to other products later.

The research that was done for this bachelor's thesis provided a significant amount of knowledge about usability testing and about other usability evaluation methods to the author. It has also made clear the importance of usability in a product development. The gained knowledge will be valuable in a development of a StepLab application.

REFERENCES

1. Nielsen, Jakob 2012. Usability 101: Introduction to Usability. Date of retrieval: 11.9.2018. https://www.nngroup.com/articles/usability-101-introduction-to-usability/.

2. Rubin, Jeff – Chisnell, Dana 2008. Handbook of Usability Testing, Second Edition: How to Plan, Design, and Conduct Effective Tests. Indianapolis: Wiley Publishing Inc., p. 4, 71, 82, 102, 146, 161, 162.

3. Brooke, John 2013. SUS: A Retrospective. Date of retrieval: 11.9.2018. http://uxpajournal.org/wp-content/uploads/pdf/JUS_Brooke_February_2013.pdf.

4. Nielsen, Jakob 1993. Usability Engineering. San Diego: Academic press, p. 25.

5. Norman Don – Nielsen, Jakob. The definition of user experience. Date of retrieval: 11.9.2018. https://www.nngroup.com/articles/definition-user-experience/.

6. Robier, Hannes 2015. Usability versus user experience. Date of retrieval: 11.9.2018. https://www.linkedin.com/pulse/usability-versus-user-experience-hannes-robier.

7. Wiklund, Michael – Kendler, Jonathan – Strochlic, Allison 2011. Usability testing of medical devices. Boca Raton: CRC Press, p. 2, 35, 41, 90, 91, 244, 266, 274–276.

8. Kuusijärvi, Miro 2017. Medical Application Design. Oulu: Oulu University of Applied Sciences, Degree Programme in Information Technology. A thesis.

9. Wiklund, Michael 1995. Medical device and equipment design: usability engineering and ergonomics. Buffalo Grove: Interpharm Press, Inc., p. 3–4, 279–281.

10. ISO 62366-1:2015. 2015. Medical devices - Part 1: Application of usability engineering to medical devices. Helsinki: Finnish Standards Association.

11. Nielsen, Jakob 1995. 10 Usability Heuristics for User Interface Design. Date of retrieval: 27.9.2018. https://www.nngroup.com/articles/ten-usability-heuristics/.

12. Spool, Jared M. 2005. Seven Common Usability Testing Mistakes. Date of retrieval: 27.9.2018. https://articles.uie.com/usability_testing_mistakes/.

13. Morville, Peter 2004. User Experience Design. Date of retrieval: 27.9.2018. http://semanticstudios.com/user_experience_design/.

14. Cognitive Walkthrough. Date of retrieval: 27.9.2018. https://www.usabilitybok.org/cognitive-walkthrough.

15. Travis, David 2010. The 4 questions to ask in cognitive walkthrough. Date of retrieval: 27.9.2018. https://www.userfocus.co.uk/articles/cogwalk.html.

16. Guo, Frank, 2012. More Than Usability: The Four Elements of User Experience, Part I. Date of retrieval: 27.9.2018. https://www.uxmatters.com/mt/archives/2012/04/more-than-usability-the-four-elements-of-user-experience-part-i.php.

17. ISO 9241-11:2018. 2018. Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. Helsinki: Finnish Standard Association.

18. Jordan, Patrick W. – Thomas, Bruce – Weerdmeester, Bernand A. – McClelland, Ian L. 1996. Usability Evaluation in Industry. London: Taylor & Francis Ltd, p. 191.

19. Sauro, Jeff – Lewis James R. 2016. Quantifying the User Experience, 2nd Edition. Cambridge: Morgan Kaufmann, p. 6.

20. Pikkarainen, Niklas 2018. Usability Evaluation Report. Oulu: MoveSole Ltd. A research report, p. 19.

21. Finstad, Kraig 2006. The System Usability Scale and Non-Native English Speakers. Date of retrieval: 2.10.2018. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.365.2352&rep=rep1&type=pdf.

22. Bangor, Aaron – Kortum, Philip – Miller, James 2009. Determining What Individual SUS Scores Mean: Adding and Adjective Rating Scale. Date of retrieval: 2.10.2018. http://uxpajournal.org/wp-content/uploads/sites/8/pdf/JUS_Bangor_May2009.pdf.

23. MoveSole. Date of retrieval: 5.10.2018. http://www.movesole.com/en/.

24. Kaikkonen, Eero 2018. CEO, MoveSole Ltd. Interview 8.10.2018.

25. Nilsson, Richard 2010. Usability vs. User Experience (UX). Date of retrieval: 12.11.2018. http://www.neospot.se/usability-vs-user-experience/.

26. World Health Organization. Date of retrieval: 20.11.2018. http://www.who.int/medical_devices/definitions/en/.

APPENDICES

Appendix 1 Patient Consent Form

Appendix 2 Participant Information Form

Appendix 3 Observer Form

Appendix 4 System Usability Scale

Appendix 5 Orientation Script

Appendix 6 Post-test Interview Questions

Appendix 7 StepLab Summative Usability Test Plan

Patient Consent Form

I consent to be video and audio recorded during this test event by MoveSole Oy. I waive any right to inspect or approve the finished recording. All records created during this test event are for the use of MoveSole Oy only and are stored and maintained in a private network within the premises of MoveSole Oy, upholding information security.

Yes ☐ *No* ☐

I agree to participate in the study conducted by MoveSole Oy.

I understand that participation in this usability study is voluntary and I agree to immediately raise any concerns or areas of discomfort during the session with the study administrator(s).

Date and place: _____

Signature: _____

StepLab Summative Usability Test

Name: _____

Participant proficiency

Proficiency in these categories is rated 0-5 depending on your knowledge of the specific subject. The numbered values represent the following grades:

5 – Expert
4 – Advanced
3 – Intermediate
2 – Novice
1 – Fundamental awareness
0 – None

These grades are used to help the test administrators analyse your test results based on your skill level.

Proficiency with technological devices:
How experienced are you with commercially sold technological devices such as computers?

5 ☐ 4 ☐ 3 ☐ 2 ☐ 1 ☐ 0 ☐

Proficiency with mobile device use:
How experienced are you with mobile technology, e.g. smart phones, tablets, and other handheld smart devices?

5 ☐ 4 ☐ 3 ☐ 2 ☐ 1 ☐ 0 ☐

Which mobile operating systems are you familiar with? *Android* ☐ *iOS* ☐ *Other* ☐

Proficiency with medical equipment:
How experienced are you with items and tools used in healthcare, e.g. stethoscopes, orthopedic insoles?

5 ☐ 4 ☐ 3 ☐ 2 ☐ 1 ☐ 0 ☐

Proficiency with measuring devices:
How experienced are you with devices used to measure humans, e.g. ECG devices, blood pressure meters, blood sugar meters?

5 ☐ 4 ☐ 3 ☐ 2 ☐ 1 ☐ 0 ☐

Proficiency with lower-limb issues:
How experienced are you with detecting, treating and/or preventing lower-limb issues?

5 ☐ 4 ☐ 3 ☐ 2 ☐ 1 ☐ 0 ☐

Participant Information

Name: _____

	<25	***26-35***	***36-45***	***46-55***	***>56***
Age:	☐	☐	☐	☐	☐

	Female	***Male***	***Other***
Sex:	☐	☐	☐

Profession: _____

I consent to be video and audio recorded during this test event by MoveSole Oy. I waive any right to inspect or approve the finished recording. All records created during this test event are for the use of MoveSole Oy only and are stored and maintained on a private network within the premises of MoveSole Oy, upholding information security.

Yes	*No*
☐	☐

I agree to participate in the study conducted by MoveSole Oy.

I understand that participation in this usability study is voluntary and I agree to immediately raise any concerns or areas of discomfort during the session with the study administrator(s).

Date and place: _____

Participant signature: _____

Observer Form 1/2

Participant name: _____

Yes	No	Case to be evaluated	Notes
☐	☐	Participant has issues replacing the battery	
☐	☐	Participant has issues helping the patient wear the smart insoles	
☐	☐	Electronics casing or ankle band chafes the participant's skin	
☐	☐	Insole does not fit in the shoe even though the sizes are the same	
☐	☐	Participant does not know how to start a session	
☐	☐	Participant has problems to choose the right insoles on the smart device	
☐	☐	Participant does not know if the insole is measuring during a session	
☐	☐	Participant has problems to understand the alert function	
☐	☐	Participant accidentally stops the session too early	
☐	☐	Participant does not understand or misunderstands the data shown during a session	
☐	☐	Participant does not understand or misunderstands the data shown in the session summary or report	
☐	☐	Participant has problems to transfer data to a PC	
☐	☐	Participant has problems to delete sessions on a smart device	
☐	☐	Participant required administrator's assistance during the test event	
☐	☐	Software bug	

Observer Form 2/2

Participant name: _____

Task completion times

	Start	End	Time	Notes
Task 1				
Task 2				
Task 3				

Notes

Date and place: _____

Observer's signature: _____

System Usability Scale

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

StepLab Summative Usability Test Protocol

1. Greet the test participant, introduce the test team and make the participant feel comfortable through small-talk.

   "The purpose of this test is to evaluate the usability of this product, not to test you"

2. Hand out a participant information form to the test participant:

   "The participant information form is used to evaluate the usability of the system between different user groups. You are asked to answer these questions and sign the form to give us a permission to video record this session. All the videos and information are only for internal use of the company and only to help evaluating the usability of the system."

3. Let the participant fill in the information form in peace, and answer to any questions that they may have.
4. Check the form and start the video recording.

   "This text is read aloud from a script to ensure that each participant gets the same instructions each time. You will be asked to fill in a question form and give other feedback concerning usability after the test has been done."

5. Hand out a task list to the test participant

   "You are now asked to carry out these four tasks, one at the time in a respective order. If you don't have any questions, you can start working on task one. Then you can invite the patient in and continue with tasks two and three"

6. Wait until the task 1 has been done and the participant invites the patient in.
7. Hand out a patient consent form to the patient:

   "You are asked to sign this form to give us a permission to video record this session. All the videos and personal information are only for internal use of the company and only to help evaluating the usability of the system."

8. Wait until tasks 2 and 3 have been done.
9. To the patient:

   "How did the smart insole feel in your shoe? Did it cause discomfort?"

10. Thank the patient.
11. Let the participant finish with the patient.
12. Wait the participant to carry out task 4.

"Now you are asked to evaluate the usability of this system on these forms. All comments about any part of the system are valuable. In a SUS questionnaire you are asked to mark your immediate response to each item, rather than thinking about items for a long time"

13. Hand out a system usability scale form and a post-test questionnaire to the participant and ask to fill in them.
14. After the participant has finished filling in forms, open free discussion. Comments, feed-back, and suggestions are welcome.
15. End the test event, thanking the participant. Stop video recording.

StepLab Summative Usability Test

Post-test interview

Participant's name: _____

1. *When was the first time you used the StepLab system?*

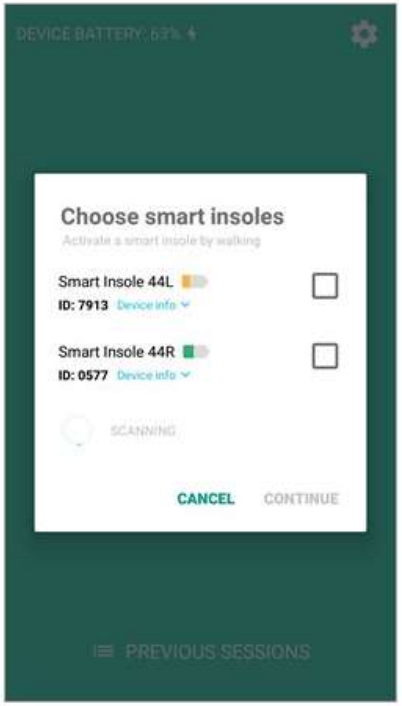
2. *How many times have you used the system before?*

Following the mock-up pictures of the application can you point out…

3. *If you find the system complicated to use at some phase?*

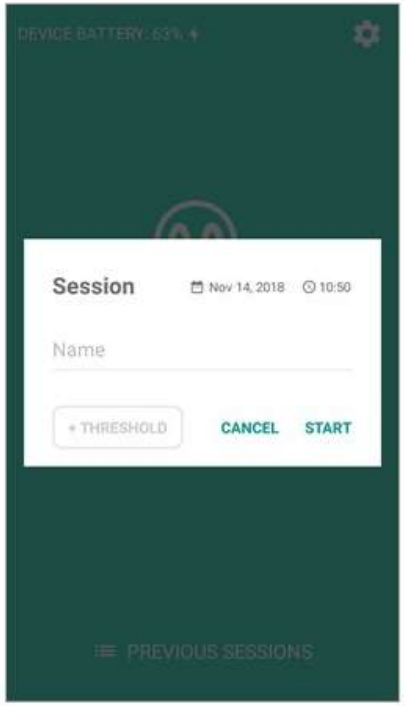
4. *Where did you find the system really simple/smooth to use?*

Task 1

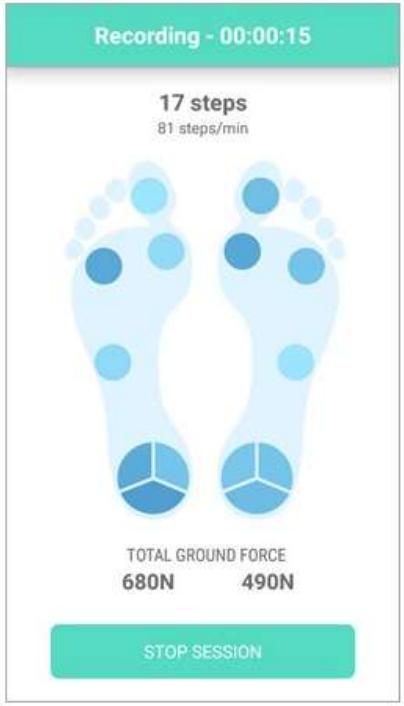


1. Checking Battery Levels
2. Changing Batteries

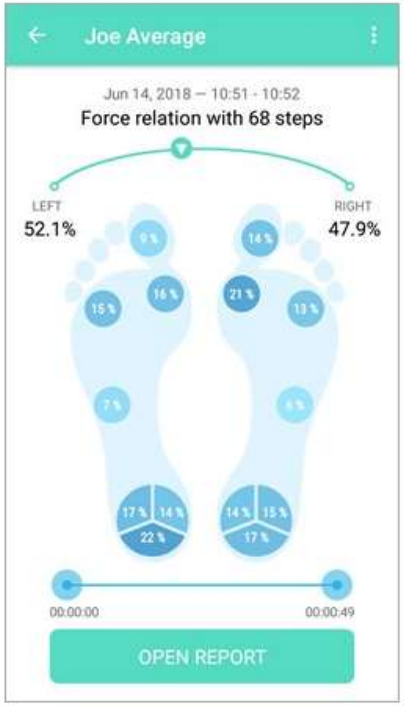
Task 2



3. Wearing Smart Insoles/ Helping the Patient to Wear Smart Insoles
4. Starting a New Session

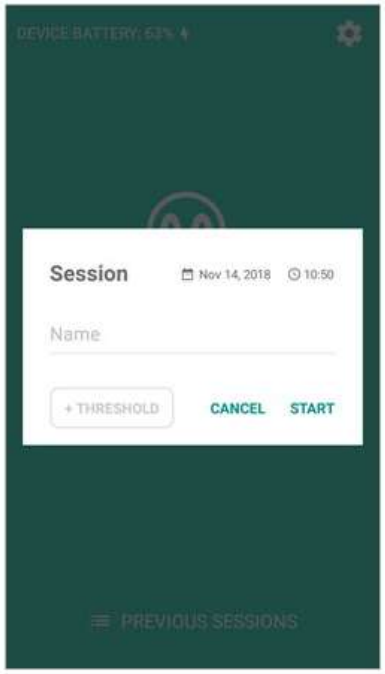


5. Recording Phase
6. Understanding the Data Shown during a Session

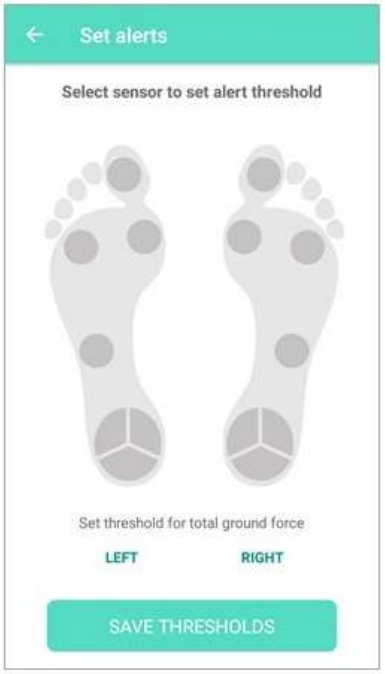


7. Understanding the Data Shown After a Session

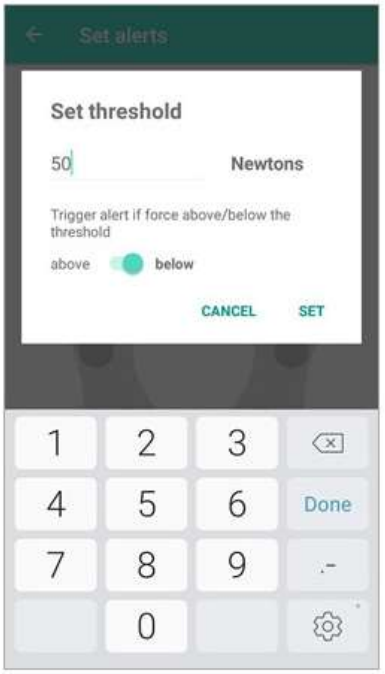
Task 3



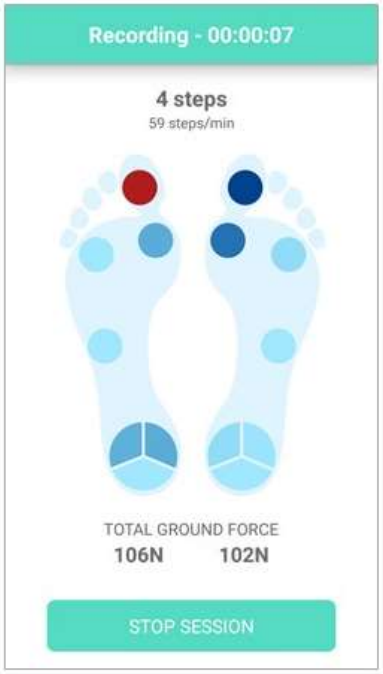
8. Finding an Alert Setting Menu



9. Setting an Alert for a Correct Sensor



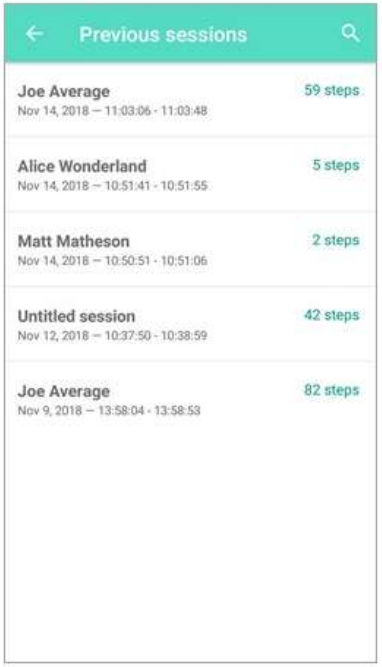
10. Setting an Intended Alert Threshold



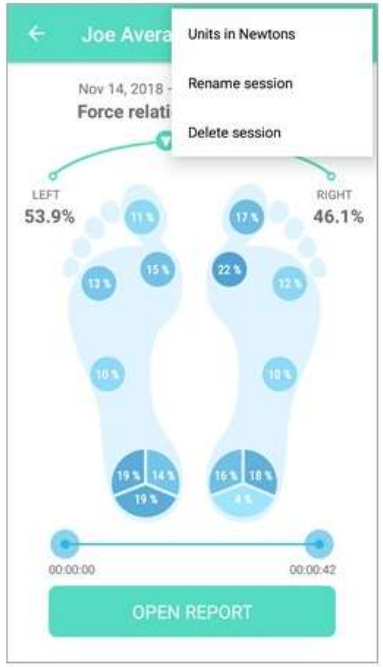
11. Understanding the Alert Feature and its Functionality

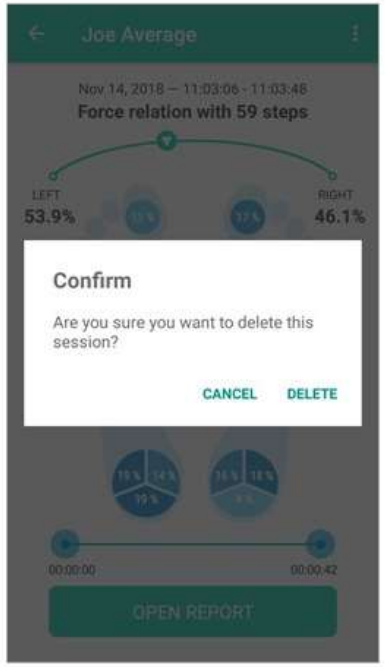
Task 4

12. Storing Recorderd Session Reports on a PC



13. Finding the Previously Recorded Sessions on a Smart Device





14. Deleting a Session

Table of contents

1 Testing Plan

This summative usability test will include a screening questionnaire, a pre-test questionnaire, a test task list, a SUS questionnaire and a post-test questionnaire with open ended questions.

2 Main Purpose of the Test

The main purpose of the test is to provide an objective evidence that the StepLab system is safe to use by ensuring that using the StepLab system does not cause harm to the patient, the user or others. Also, the purpose is to ensure that StepLab meets the device regulators' expectations.

3 Test Methods

The tests are recorded on Full HD video with sound so that they can be played back and inspected on a later date. The videos will be used to document any significant comments about the usability of the device and to examine the pleasantness of using StepLab.

4 Test Environment and Equipment

This test will be conducted in clinics of recruited participants to ensure the actual environment of use.
StepLab system with:

- All available sizes of Smart Insoles
- A smart device
- Spare batteries
- A screw driver
- Ankle bands
- A user manual

Also included:

- An orientation script (for moderator)
- Pre-test questionnaire forms
- Post-test questionnaire forms
- System Usability Scale (SUS) forms
- Observer forms
- A camera
    - Full-HD video and voice recording required
    - SD card with a memory of 30 min * number of participants

5 Test Administrators

A moderator is the person leading the test and interacting with the participant. The moderator will explain the instructions and hand out to and collect the forms from the participant.

An observer(s) is (are) supposed to observe and take notes without biasing the test in any way. The observer will count errors and indications of frustrations or other expressions by the participant. In this test, at least 2 observers are needed to take the notes and one of them will start the video camera when the test starts. A technical expert is needed if something goes wrong with the system and for example a software of smart insoles must be reinstalled.

This summative usability test is conducted by 2 people. One of them is using the video camera, observing and taking notes. The other one is administrating the test, observing and handling technical issues if needed.

6 Test Participants

6.1 Sample Size

A sample size of this summative usability test should be minimum 10 participants but the target is 15-20 participants. More important than the number or participants is that the requirements of the user profile are met, meaning that different levels of expertise in using smart devices as well as different levels of expertise in the gait analysis and feet ulcers are covered by the user population.

6.2 Participant Characteristics

An Intended user base of StepLab defines the most critical test participant groups as healthcare professionals, such as:

- Orthopedists
- Physiotherapists
- Foot therapists

All these user groups should be covered with at least 3 participants of each profession. Researches and students in the medical field are also important test participant groups. The StepLab system is relatively simple to use and just doing measurements with the device should not cause any harm to the patient, the user or any people around. Basically, anyone can use the system but viewing the results and doing an analysis may require some health care experience.

6.3 Recruitment

The recruitment will be handled by the person who is responsible for customer relationships in the company. Recruiting can be handled via phone, e-mail or live interviews. A screening questionnaire (Appendix 1) is used in recruitment to find suitable participants.

7 Test Reliability

A pilot test will be held by the test personnel using several employees of MoveSole as test participants. The pilot will be an exact simulation of the actual test event, except its purpose is not to test usability but rather test the testing process, used forms and equipment and to practice test personnel's roles in the event.

Each device used in the test will be checked and tested before the beginning of each test occurrence to ensure that every test goes as planned without obstacles.

The test result reliability is upheld by having a high participant count and by ensuring that the tests are recorded on video for analysis after the test has been finished. The diversity of user profile is ensured by using screening questions when recruiting participants. Screening questions and the answer table are in the appendix 1.

To ensure that the test is conducted in the actual conditions of use, they are held in clinics of the recruited participants with their actual patients. This requires that the permission to film the test event is asked from both the participant and their patient. The consent forms can be found from the appendices 2 and 5.

8 Test Schedule

One test event with each participant should not take more than 30 minutes. The recruitment process is started in October 2018 and the tests are conducted when the suitable participants are found. The tests will be conducted in several parts in different locations with different participants.

9 Test Procession

9.1 Before the Beginning of the Test

The participant will receive the StepLab system and they are free to read the user manual and try the product on their own and with their patients before the testing event. No training is provided by the manufacturer. If the customer is providing training to their employees, it should be provided normally and noted in the test report.

9.2 Tasks for the Test Participant

The test event includes four separate tasks. Through the tasks the usability of the system is evaluated and measured. The first task is measuring the readability of the user interface of the system and the usability of its maintenance. The second task is left open for different use cases and it is assumed that the participant knows what they want to measure with the smart insoles. The third task is used to test the usability of an alert functionality on a smart device. The fourth task is intended to measure the usability of the smart device in case of maintaining the security of patient information.

9.3 Ending the Test

The actual test is finished when the participant says “done” in the end of the last task. After that, they will be asked to fill in the SUS form and a post-test questionnaire. Also, the participant is encouraged to provide vocal comments about the usability of the product.

10 Test Personnel Instructions

10.1 Collected Data

Observers are filling in the observer form (Appendix 4) which includes the benchmarks based on defined research questions. A data of the benchmarks is collected during the test and any additional notes concerning the usability of the product should be written down in the *notes* field. Indications of frustration or vocal comments must be written down in the observer form. Task completion times can be checked from the video recordings afterwards.

10.2 Time-critical Tasks

Test tasks of StepLab systems are not critical to safety what comes to the speed of the task performance. However, task times are measured for each task and ease-of-use and satisfaction are evaluated through the task times along with the other benchmarks.

11 Analysis

11.1 Participant Information Form

Each participant will fill in a form where they agree to being filmed during the test procedure. Besides the name, age, gender and profession, the participant information form (Appendix 5) gives information about the expertise in technological and medical device usage and lower limb injuries. The expertise in technological and medical device usage is separated in five different questions, which give a good basis of the participants’ expertise background.

11.2 System Usability Scale (SUS) Analysis

System Usability Scale (SUS) forms are collected as a 10-item scale as the global view of subjective assessments of usability is formed. For the analysis, there is a global form for interpreting the scores:

- For the odd numbered questions, the result value will be subtracted by 1.
- For the even numbered questions, the result value will be subtracted from 5.
- All the new values in range 1-4 will be summed and multiplied by 2.5.

For the result, there will be one value in range of 0-100, which is not percentage, but it is easy to interpret.

11.3 Results and Analysis

A participant information form (Appendix 5), a post-test questionnaire (Appendix 8), an observer form (Appendix 4) and a System Usability Scale (Appendix 6) are collected as results. The results will be analyzed and visualized in the test report document. The results will be segregated according to a user group. The results are considered separately from a validation and commercial point of view.

As a validation point of view the focus will be on the use errors and the use safety and success related post-test discussion of participants. The analysis of checklist will be done with descriptive statistics. The results of a post-test checklist will be presented as a graph and responded with conclusions and recommendations for improvements.