Usability Evaluation

Maisa health portal

Phase 4 - Evaluation Plan

Jehan Khattak

Ondřej Brém

Shanshan Hou

Tomas Villikka

Katyayani Singh

Table of Contents

1. Objectives

Our evaluation plan for the Maisa portal aims to tackle the main usability issues that were revealed after the team conducted the Heuristic Evaluation in phase 1. We also want to address the overall user experience of the portal. Our study will be based on planning user-testing methods that match our area of focus and the context of the Maisa Portal.

Our focus area that we decided to tackle is the menus and navigational structure of the service. The menu structure includes the top menu, the “Quick Links” sidebar menu and the overall navigation within the service functionalities themselves.

Each method has been chosen with a clear objective for each. This has been summarized below:

#	Evaluation Method	Objective of Evaluation	Measurement/Metric
1	Moderated Usability Testing	To allow users to identify usability problems in tasks like appointment scheduling and navigation with supervision.	The feedback from the final interview and recorded sessions. The observed difficulty or ease with which the tasks were performed.
2	Unmoderated Remote Usability Testing	To allow users to freely identify usability issues by completing tasks, without any supervision.	The feedback interview from the participants. As well as the number of successfully performed tasks.
3	Eye-tracking	To understand the cognitive behaviour of users when using the navigational structures. To collect quantitative data on usability testing.	Using the software (Begaze) results. Heatmaps and calculated averages (e.g. time to find wanted navigation parts) for each session.
4	UX Questionnaire	To evaluate the user experience after using the navigation menus to complete tasks.	Quantitative results from the scaled questionnaires.

2. General User Groups

For our testing plan we took time to carefully consider the best user group and keep their needs and contexts in mind before choosing the appropriate methods. Since the Maisa portal aims to be a convenient online healthcare solution we want to focus our study on patients who are using the portal.To successfully plan our usability testing we looked into the characteristics and context of use for our targeted user group:

- Our user group would be patients who are fairly familiar with technology and may have had exposure to other online booking systems within Finland.
- These patients may choose to use the portal while at home, work or while on the go. The least amount of hassle and efficiency are major factors with this user group.
- It is also important to note that with this user group there may be a sense of urgency and unease if there is a medical issue at hand. This is contextual detail that we will take into account while planning the usability tests.

This specific user group was chosen for our study because of a few factors. Firstly, this is a common user group that covers a majority of users for Maisa. Which means that we should find the maximum number of pitfalls and possible solutions by using this user group. We also felt that the patients would be the category of users we would aim for since their use of the navigational structures determines heavily the success of the portal itself. Therefore this target group was kept in mind when coming up with the recruiting plan for each evaluation method.

3. Suggested Methods

We searched through and discussed quite a long list of methods, many of which can be found at NN groups website[1] or in an online method list curated by Philosophical faculty of Masaryk University in Brno[2]. From this list, combined with the aforementioned objectives, we picked four basic methods or method groups that we will introduce further and explain their usage in *Evaluation Execution*.

3.1. Moderated Usability Testing[3]

As the name suggests, during moderated (facilitated) testing, the facilitator gives instructions and task scenarios to the participant. The participant provides behavioral and verbal feedback about the interface while he performs those tasks.[4] Moderated testing is in most cases used as a qualitative method. As such it focuses on collecting insights and findings about how people use the product or service. The best use of this method is discovering problems in products' user experience.

[1] https://www.nngroup.com/
[2] https://100metod.cz/ (unfortunately doesn't have english version)
[3] COCKTON, Gilbert. Usability Evaluation. The Encyclopedia of Human-Computer Interaction [online]. The Interaction Design Foundation, 2014 [cit. 2016-02-19]. ISBN 9788792964007. Dostupné z: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/usability-evaluation
[4] https://www.nngroup.com/articles/usability-testing-101/

3.1.1. Advantages of Moderated Usability Testing

1. Direct access to the user during the test - facilitator can dig deeper with questions to better understand why the participant lost or what don't they understand.
2. Better control over how the test is going as the moderator can intervene when the participant gets stuck for reason unrelated to the test objective
3. The facilitator may change, skip, and reorder tasks as needed.
4. Facilitator can keep the participants more focused and stay on topic.
5. Participants may be reminded to think aloud if they forget.

3.1.2. Disadvantages of Moderated Usability Testing

1. Scheduling can be an issue
2. Participants may not be comfortable sharing negative feedback to someone face to face.
3. The presence of a facilitator may introduce behaviour that is not normal for the user - they try to please the facilitator or look better in their eyes.

3.1.3. How to Apply this Method

Moderated testing can be used both in person and remote with slight modifications to the process and attention from the facilitator. The main difference is in preparation of the test. For remote testing the technical support needs to be perfected because it's way more difficult to improvise and fix issues when the participant can't even connect to the test. The general process for both would be following:

1. State the objective of the test - which parts of UI, flow or structure will be tested and what hypothesis is being validated (it can be purely discovery based but it's good to know that beforehand)
2. Define who are the users of that part of your service and then recruit participants that fit that description.
3. Based on the objectives of the test prepare the tasks that will be conducted by the participants.
4. Get the facilitator familiar with the tasks.
5. Perform the test with participants - following all the necessary steps to make the process run smoothly without compromising the results
6. Perform a final feedback interview with the participant.
7. Analyze notes and potential recordings from the sessions with all participants to discover root causes for their struggles.

This process is usually performed with around 5 participants per evaluation objective. With more the issues start repeating too much and the returns of additional cost are diminishing.

3.2. Unmoderated Remote Usability Testing

Unmoderated remote testing involves testing the system without the physical or temporal proximity of evaluators to the test subjects. Instead, a software application or a test document guides the test subject through various tasks and prompts the user with follow-up questions. Test subjects are free to take the test on their own computers at any time convenient to them. [5]

3.2.1. Advantages of Unmoderated Remote Usability Testing

1. Unmoderated testing is much faster than moderated testing because the researchers do not have to schedule separate evaluation sessions with each test subject.
2. Because the researcher and test subjects do not have to meet physically, the test can be sent out to a large number of potential participants to get a robust dataset of usability problems.
3. The test subjects undertake the study from their home, which comes very close to mimicking the behaviour of actual Maisa users. Since most Maisa users will be accessing the system from their homes and on their own devices, it makes sense to collect data on usability issues encountered in the absence of external moderators and evaluators.
4. Experts warn against using Unmoderated Remote Usability testing for early prototypes. But in the case of Maisa, the portal is already developed and this limitation does not apply.
5. Research has shown that participants express negative feedback more freely in the absence of external scrutiny from researchers. [6]

3.2.2. How to Apply this Method

In order to perform Unmoderated Remote Usability Testing, we intend to complete the following steps:

1. Selecting a testing software based on the study objectives defined above.
2. Preparing a list of tasks, instructions and follow-up questions for the subjects.

---

[5] Whitenton, K. (2019). Unmoderated User Tests: How and Why to Do Them. Internet access: https://www.nngroup.com/articles/unmoderated-usability-testing/

[6] Tullis, Thomas & Fleischman, Stan & Mcnulty, Michelle & Cianchette, Carrie & Bergel, Marguerite. (2002). An empirical comparison of lab and remote usability testing of Web sites.

3. Pilot testing the list of instructions, tasks and questions with some people to identify shortcomings.
4. Sending the list of tasks, instructions and questions to recruited participants.
5. Analyzing the data obtained from the users.

3.3. Eye-tracking

Eye-tracking in usability testing offers additional tools for usability evaluators in addition to the existing methods. As a method, it gives more insights into the user's thinking process and what captures the attention in more detailed levels. Eye-tracking is considered somewhat a difficult method to be used in usability testing[7]. Furthermore, eye-tracking adds one more level of planning, analyzing, and time-consumption steps as compared to normal usability testing. The method follows the user's eye movement with , for example, eye-tracking glasses. It produces a video of the user's vision field with an added dot. The dot moves rapidly and represents precisely the user's eye focusing point. It is, however, important to use a method which isn't visible for the user in order not to disturb the test results[8].

3.3.1. Advantages and Disadvantages of Eye Tracking

This method is broadly used in medical devices and telemedicine systems. Compared with other usability evaluation methods, eye tracking is more of a supportive method. It provides quantitative and objective data from participants' behaviors[9], especially obtaining information-seeking behavior.[10] Unlike cognitive walkthrough and protocol analyses, which also focus on cognitive behavior, eye tracking is time-saving and effective. With the development of VR and Machine Learning, the gathered data can be more reliable than ever before. As we consider focusing on the navigational structure as a whole, through eye tracking we can directly know how users react to the current navigation and locate where the problems are. However, problems such as calibration, education and devices cost ethical issues still need to be fixed.

[7]Pernice, K. & Nielsen, J. (2019). How to Conduct Eyetracking Studies. Internet access: https://media.nngroup.com/media/reports/free/How_to_Conduct_Eyetracking_Studies.pdf

[8] Elbabour, F., Alhadreti, O., & Mayhew, P. (2017). Eye tracking in retrospective think-aloud usability testing: is there added value?. Journal of Usability Studies, 12(3), 95-110. https://dl-acm-org.libproxy.aalto.fi/doi/10.5555/3190862.3190864

[9] Olsen, Anneli & Smolentzov, Linnea & Strandvall, Tommy. (2010). Comparing different eye tracking cues when using theretrospective think aloud method in usability testing. 45-53. 10.14236/ewic/HCI2010.8. https://www.researchgate.net/publication/221436817_Comparing_different_eye_tracking_cues_when_using_theretrospective_think_aloud_method_in_usability_testing

[10] King A J, Cooper G F, Clermont G, et al. (2020) Leveraging Eye Tracking to Prioritize Relevant Medical Record Data: Comparative Machine Learning Study[J]. Journal of Medical Internet Research,22(4): e15876.https://www.jmir.org/2020/4/e15876/

3.3.2. How to Apply this Method

Eye-tracking for our usability testing is an additional method for supporting analyzing phase. Thus, the usage of eye-tracking glasses follows the process of moderated lab testing. The glasses are fairly close normal glasses and shouldn't be a too big disturbance for the user[11]. The software (Begaze) can be used to analyze results. For instance, heatmaps or calculated averages (e.g. time to find wanted navigation parts) can be created from the test session.

3.4. UX Questionnaire

Here we introduce questionnaires as a method for evaluating user experience. After the participants have completed experiencing the product, it is common to use a survey or questionnaire to understand the user's experience. The content often contains close-ended questions and open questions. The former is presented as a rating for an experience or attitude, and the second encourages the participants to express their subjective feelings in order to help researchers obtain more information. According to our research, questionnaires are the most common way of evaluating a telemedicine system, which appears in 69% of relevant research. The most popular types of questionnaires include SUS, IBM ease-of-use, TAM-2, SUMI, NASA TLX. System Usability Scale (SUS) evaluates a wide variety of products and services. Technology Acceptance Model (TAM) is the most widely applied model of user acceptance and usage[12]. IBM computer usability satisfaction questionnaires were created by James R. Lewis, including two after-scenario questionnaire(ASQ and PSQ) and two overall satisfaction questionnaires (PSSUQ and CSUQ)[13]. Here, we would use SUMI as the main method in the questionnaire, which is more relevant to user experience but not usability.

3.4.1. Advantages and Disadvantages of UX Questionnaires

Besides avoiding negative feelings, evaluating fulfillment is also an important topic. Questionnaires can bring both qualitative and quantitative answers, depending on how you manage it. The questionnaires we mentioned above all have a specific and strict matrix for results analyzing. The standardized measurements provide objectivity, quantification (researchers are able to use statistical methods to analyze) and economy (both in time and money). However, response biases such as acquiescence bias, social desirability bias, and recency bias might lead to an unreliable result.

---

[11] https://www.aalto.fi/sites/g/files/flghsv161/files/2018-10/abl_manual_eyetrackinglasses_02_11_2017.pdf

[12] Venkatesh, Viswanath & Davis, Fred. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science. 46. 186-204. (PDF) A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies

[13] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57-78. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=093CCED22F5F2079226DCCEC03840472?doi=10.1.1.584.6610&rep=rep1&type=pdf

4. Ethical and Legal Issues

For each of the methods it is important to consider the ethical and legal issues for the success of the entire evaluation. To help with that we have prepared sample consent forms that follow the GDPR guidelines, this can be found in the Appendix. For more information on how to use the consent forms check the instructions for each evaluation method in Evaluation Execution.

4.1. Moderated Usability Testing

The ethical issues come mainly from two areas. Firstly the personal data of the users that need to be properly taken care of. Those can be identification, video and audio recordings of the test, data collected for paying the participants for their time and so on. This is quite well described by privacy regulations such as GDPR. The second area is the ethics of exposing the participants to experimental services that could disturb the participants, or lead to negative effects due to not being able to finish given tasks. Dealing with those is two fold; legally it needs to be covered by consent from the participant but on top of that precautions need to be taken to not to cause harm to the participant. This is especially important in our case where we deal with the healthcare system which can cause stress to the participants.

4.2. Unmoderated Remote Usability Testing

Unmoderated testing software that implicitly records the activities of test subjects raises an ethical concern of data privacy and security. Especially when dealing with sensitive information like patients' health data, the possibility of a data breach may reduce the participants' trust in the evaluation system. Therefore, it is necessary to reassure the subjects of complete confidentiality, anonymity and data security. This can be done through a consent form signed by both parties and by using evaluation systems that are proven to be secure.

4.3. Eyetracking

Ethically, the researchers need to inform and make sure participants understand what would happen to them. Only in this situation , they would feel safe and comfortable to get on with the session. However, as a high-technological method, one obvious ethical problem is that participants have no idea about what kind of information it would gather. In addition, from this technology itself, it has great power in privacy mining. Excessive utilization in eye tracking might be treated as surveillance. Therefore transparent consent forms are essential.

4.4. UX questionnaires

Considering the data gathering from the UX questionnaire would be analyzed by software provided by the third party. Moreover, the raw data includes more information than the responses to the questionnaire.It is important to use the ID feature, so you are able to link the score data to all information from participants. It is common to make an ID mark on the paper to

help connect responses with specific participants without releasing personal information. If you use external software resources, be sure to confirm that the participant's personal information is not included when uploading data.

5. User Study and Feedback

Before going ahead with the execution of the evaluation methods we took our four proposed evaluation methods and planned a small study to see how a potential user would react. In order to conduct the user study we first made instructions for each type of evaluation method. We then carried out small scenario based evaluations with potential users. Since our testing is limited with no access to the portal we used screenshots and a mockup that we created on Figma.The aim of this mini study was to involve potential users for gathering important feedback to support the development of the usability testing.

5.1. Methods used in the User Study

This ministudy was conducted with one participant for the moderated test and one participant for the eye tracking, each representing a new user of Maisa. The pilot was conducted over Zoom by introducing to the participant the instructions and tasks. Both the Moderated Usability Testing and Eye Tracking methods were conducted in the same method for the piloting, this was due to the inability to use the actual devices needed for eye tracking. Afterwards a short discussion was conducted over the screenshots of the Maisa portal to see how the tasks may be carried out.
For Unmoderated Testing the mini study was conducted remotely with one participant. One Google form (containing Pre-Task Questions, Task Questions, Post-Task Questions) was sent to the participant along with a Figma Prototype of the Maisa Portal and a set of screenshots in the form of a PDF. The screen and audio of the user were not recorded in the ministudy. Finally for the UX Questionnaire the pilot was carried out by taking our potential user and giving them the instructions for this testing method. For the evaluation questions we chose the SUMI(Software Usability Measurement Inventory) as the main method for user experience evaluation. We used this form as an example: http://sumi.uxp.ie/en/. The mini study was executed through Figma.

5.2. Main Feedback Received

While we received feedback for each study the overall feedback can be summarized with the following:

- **Incomplete Instructions:** For each of the instructions initially written we found that certain information was missing. This included:
    - explaining what is the Maisa portal and what is its purpose.
    - GDPR related info was not included in the instructions -how long and with whom the personal information will be shared.
    - What is the aim of the test?
    - How long will the test be?

- **Descriptive Tasks:** While asking the users to carry out the tasks we found that:
    - Context/situation based tasks seem to be clear to understand and easy to think of what is being asked to do.
    - Some questions were duplicated and some users felt very confused about terminologies used.
- **Anonymity Concerns:** Some users were apprehensive about being uniquely identified via her Google Form response when asked for their feedback.
- **Short Answers:** Because the users were not promised any incentive, they had low motivation to answer the Google Form in detail. Hence most answers were quite short, with no scope for giving greater insight into the user's psyche.
- **PDF/Figma Limitations:** While conducting the user study we found that:
    - It is important to note that while using Figma, when the user cannot find the target object and clicks at random, Figma will pop up the blue box to indicate other optional areas. This also helps the subject to complete the task and affects the accuracy of the performance to some extent.
    - Testing with screenshot pdf's is rather confusing and difficult.

5.3. Incorporation of Feedback and Limitations

In order to make the most of the evaluation methods the feedback was analyzed and the following changes were incorporated into the final evaluation plan:

- **Rewriting Instructions**: Each instruction set was written so that the participants would have a clear idea of what kind of test was going to take place. Open invitations were given to back out whenever the user did not feel at ease.
- **Addition of the GDPR agreements:** This was added to lower the apprehension users may have regarding anonymity.
- **Understanding the limitations:** Since we did not have access to the Maisa portal while planning the evaluation plan it is important to note that the actual evaluation should not be carried out with images, PDF's and even the Figma Prototype, as this can hamper the success of the evaluation. Furthermore we understood that not all of the evaluation methods can be handled remotely.

Overall our user study helped us locate the main issues with our evaluation techniques and gave us some insights into refining them. For this mini-study we were however not able to test the eye-tracking, due to the limitations around the equipment. Moreover we realized that our final plan cannot be conducted in its entirety, remotely. Thus our final evaluation methods were updated in the execution section below.

6. Evaluation Execution

Before any of the methods is to be executed with participants the context of the portal and accompanying service should be presented and explained by the moderator/facilitator. A brief suggestion how it can be done:

“Maisa is a portal used in Vantaa’s social and health care services.It provides citizens a way to communicate with professionals, book appointments, see research results and deal with other related issues. In this study we mainly focus on it’s healthcare part that should make it easier to get needed care.”

6.1. Moderated usability testing

6.1.1. Requirements

- 5+1 participants
- Moderator/facilitator - person who runs the test with participants.
- Observer(s) - someone from the research/design team and potentially someone from the developer team to write down findings.
- Testing environment - actual working instance of the portal with demo data in it. (including demo accounts so that participants don’t need to use theirs)
- Testing space - usability lab or improvised lab where participant and facilitator can sit at a computer without being disturbed/distracted by the observer(s), which means the observers need to be separated from the place where the test is happening. (observation is usually done through one way mirror or set of cameras)
- Instructions for moderator/facilitator and participants.
- Task list (provided in the appendix)
- Screening questionnaire (provided in the appendix)
- Consent forms (provided in the appendix)
- [optional] Video and audio recording equipment - it is good to have recordings but it is advisable to write down notes during the test as going over hours of footage later is a very painful process.

In terms of equipment it is recommended to use whatever devices you and your team are familiar with using. It goes both for the recording devices and for the testing computer.

6.1.2. Execution Steps

- Test run the system
- Recruiting of participants
- Screening questionnaire/interview
- Introduction to the test, explaining what it is about
- Signing the consent form

- The actual test session with the task list.
- Post test debrief with participant - to figure out their overall impressions and takeaways
- [optional] Debrief with the team to write down notes that individually weren't noticed

Time Estimate: Doing these steps with 5 participants will likely take up the whole day if we assume one hour duration of testing session per participant which is reasonable considering logistics and buffer time. Be prepared for that and make sure to account for necessary break time.

6.1.3. Participants

For moderated usability testing, which can be enhanced by the use of eye tracking explained separately later, the main aim is to gain qualitative view into the users understanding of the system that is being tested. As such the study is conducted on a relatively small number of participants. The general recommended number in the industry is 5 participants for an usability study.[14] With adding more the added gain in insight is negligible. If more insight is needed then it is advised to do so with the next iteration of the design after findings from the study have been incorporated by the design team.
The one extra participant mentioned above is for a test run of the study. It is always recommended to have an extra person to use for a test run so that time durations can be estimated, facilitators can familiarize themselves with the tasks and where potential issues arise. Based on our recommended target user group participants should be recruited from very new users of the system. This way the study avoids being biased by users that already remember all the wrong patterns that are potentially present in the navigation design.

6.1.4. Moderator Instructions

- Familiarize yourself with the tasklist.
- Try following the tasks beforehand to know where potential issues are.
- Prepare the testing environment and all the equipment that you will need during the test.
    - Space
    - Testing computer
    - Testing instance of the system
    - Recording Hardware
    - [optional] Eye Tracking hardware and software
- When the test is under way don't interrupt the participant and don't give leads on how to continue.
- Make sure the participant knows it's the product that is being tested not them!
- Let the participant know they can end the test whenever they want.
- Let the participant know about the GDPR related consents and let them sign it.
- During the test observe the participant behaviour, feelings and comments.

[14] Nielsen, J., 2012. How Many Test Users In A Usability Study?. [online] Nielsen Norman Group. Available at: <https://www.nngroup.com/articles/how-many-test-users/> [Accessed 16 May 2020].

- Don’t ever make the participant feel bad by showing lack of compassion, patience or any other way that could suggest that they are doing something wrong!
- Present the tasks to the participant as a situation/story.
- Ask complementary questions if they seem fit, but be considerate not to make the participant uncomfortable.
- Conduct the feedback interview after the task part of the test is conducted.

6.1.5. Participant Instructions

- The product is being tested, not you, it’s okay if you don’t know or don’t understand something.
- WIth this test the aim is to make the Maisa portal better for you and other potential users. We want to learn as much as possible about what we could improve.
- Please think out loud - whenever you are thinking about the next step or don’t understand something, say it out loud so that we know there is a problem with the product.
- If you are unsure about how to achieve some of the tasks do as you think and comment on it.
- Please be honest in your comments and feedback. Both good and bad are helpful. We need to know what is not understandable to make the product better.
- The test will last approximately [fill number based on the test run with actual test environment available]

6.1.6. Task list

The task list is the same for all the methods and is attached in the appendix section.

6.1.7. Pre and Post Evaluation Questions

Below is the list of potential questions to be asked by the facilitator before and after the test:

- Pre test
    - How would you approach [insert the task that is gonna be tested later] now? Try to answer the question with whatever information you have so far.
- Post test
    - If you want to, you can now try out the rest of the features of the portal.
    - How did the overall experience of the Maisa portal feel?
    - Can you describe what was most difficult to understand/achieve?
    - Compared to your current way of doing [insert whatever was tested] how does using the entire portal change your experience?

6.1.8. Result analysis[15]

If you have done the notetaking properly during the test itself you already have a big list of insights in some form, if not you have the recordings we recommended as an option. Now the question is how to proceed. In the end you want to improve the design of the service but before that can be done the insights need to be sorted out and some sense in them needs to be found. The design team can then take that and figure out changes that need to be made to improve the usability and UX of the service.

To make sense of the long list of insights use a whiteboard or big table and pack of post-its of different colors. The colors will help you determine which insight comes from which participant. Also find a simple and memorable way to mark which task was the insight related to. Take your list of insights and with the optional help of the recordings turn them into post-its such that each post-it has one isight that is clear to understand but brief enough to be read at glance. Write the mark of the task on each post-it as well. Use color to differentiate between participants.

Once this phase is done it is time to sort things out, cluster the post-its according relations and see what is repeating. There are obviously many ways to cluster things and there is no one correct answer to that. It is recommended to cluster the insights in a few different ways to improve the understanding. This process will bring understanding of what are the issues that the participants had during the test.

These clustered insights and findings then form the basis for new design decisions for the improved version of the service. As a qualitative method there is no need to turn findings into statistical data. The results are used as a base for designers to make design changes and managers to potentially make business decisions.

6.2. Unmoderated remote usability testing

6.2.1. Requirements

- 30 + 1 participants, assuming a dropout and/or failure rate of 30%, we would still have data from 21 participants to gather insights
- Moderator/facilitator - person who runs the test with participants
- Evaluator(s) to analyze the gathered data
- Testing environment
    - Actual working instance of the portal with demo data in it. (including demo accounts so that participants don't need to use theirs).
    - A web-portal/Google Form to convey test tasks, ask supplementary questions, accept users' responses and store them into a database/ spreadsheet.
    - [optional] A web-portal or forms can be replaced with an online remote usability testing service like lookback.io, usertesting.com etc. These services tend to be

[15] Clouston, M., 2018. Analysing Usability Testing Data. [online] Medium. Available at: <https://uxdesign.cc/analysing-usability-testing-data-97667ae4999e> [Accessed 16 May 2020]

expensive but can significantly improve the workflow of setting up an online test and analysing the gathered data. [16]

- Instructions for the moderator and participants. (provided below)
- Task list (provided in the appendix)
- Screening questionnaire (provided in the appendix)
- Consent forms (provided in the appendix)
- [optional] Video and audio recording browser extension- As note-taking is not possible in this method, a browser extension may be provided that can record the users' audio and video to provide additional insights. However analysing hours of footage for several participants will be a cumbersome process. So this step should be taken only if time and resources permit.

6.2.2. Execution Steps

- Test run
- Recruiting of participants
- Screening questionnaire
- Introduction to the test, explaining what it is about via remote communication methods including email or video conferencing.
- Digital consent signing
- The actual test session
- Sending the participant a Thank-You note
- Data analysis

Time Estimate: Given that participants return the responses within the stipulated deadline, this method can take up to a week to setup, implement and analyze. It is good to give users a period of 2-3 days during which they can undertake the test any time.

6.2.3. Participants

For unmoderated remote usability testing, the main aim is to gain quantitative evidence for hypotheses formed as a result of other methods of testing. In order to gain additional qualitative insights, researchers may choose to record the testing sessions and conduct analysis later. As this method tends to lean towards quantitative rather than qualitative, the recommended number of participants is 20 in order to get statistically significant results.[17] We have recommended a participant size of 30 to account for dropout rates and potential failure to complete the test/ completing the test incorrectly.

The one extra participant mentioned above is for a test run of the study. It is always recommended to have an extra person to use for a test run so that time durations can be

[16] Whitenton, K. 2019. Tools for Unmoderated Usability Testing. Available at:<https://www.nngroup.com/articles/unmoderated-user-testing-tools/> [Accessed 16 May 2020]

[17] Neilson, J. 2012. How Many Test Users in a Usability Study. Available at:<https://www.nngroup.com/articles/how-many-test-users/> [Accessed 16 May 2020]

estimated, evaluators can familiarize themselves with the tasks and where potential issues arise.

6.2.4. Moderator Instructions

- Unmoderated remote usability tests should be used when the main focus of the study is a few specific elements, rather than an overall review. Remote studies are great for gathering data on an element or widget or for seeing the impact of a relatively minor change.
- Typically, an unmoderated test should be only 15–30 minutes in duration—comprising approximately 3–5 tasks—because the dropout rate tends to increase if a test takes longer.
- Make a list of metrics or information to be collected from the users.
- Choose a service or tool that allows those metrics to be recorded or collected via a user test.
- Prepare a set of pre-task and post-task questions for the users.
- Prepare a list of instructions to install/ set up the testing tool and undertake the test.
- The written instructions need to stand on their own in the case of an unmoderated remote usability test. Every instruction, task, and question needs to be fine-tuned to eliminate the potential for misunderstanding.
- Provide the users with a consent form and ask them to sign it digitally and return via email. Make sure the consent form lists exactly what information will be recorded and how.
- Provide users with an email address or phone number to contact someone for assistance, in case they need it.
- Be available by email (if not by phone) as much as possible to help with any potential user questions.
- Be aware that some sessions may be less valuable or unusable for the study if a user runs into problems, skips tasks, or fails to complete what was asked.
- No-show rates for any remote study can be higher than for in-person studies. The quality of an unmoderated session cannot be determined until the evaluator has analysed it. Therefore, it is better to add a few more users than you think you need in order to accommodate such problems.
- Send the users a thank-you note to express appreciation for their time.

6.2.5. Participant Instructions

- To conduct the test, you need to complete the 5 tasks below and answer some questions before, during and after performing the tasks.
- The test will take a maximum of 30 minutes.
- Please complete the test in a single sitting.
- The answers will be collected via a web-portal/ web service/ Google forms.
- First, answer the pre-task questions.
- Then perform the tasks and answer questions in the form related to each task.

- Then answer the Post-Task questions.
- The product is being tested, not you, it's okay if you don't know or don't understand something.
- If you are unsure about how to achieve some of the tasks do as you think and comment on it.
- Please be honest in your comments and feedback. Both good and bad are helpful.

6.2.6. Task list

The task list is the same for all the methods and is attached in the appendix section.

6.2.7. Pre and Post Evaluation Questions

Pre-Task Questions

- Demographic information like age, gender, occupation, job title.
- Do you have any previous experience with Maisa? (Yes/ No)
- Do you suffer from any chronic illnesses that require frequent visits to a doctor? Specify.
- How do you manage your appointments currently?
- How do you record your medical history currently?

Task Questions

- Were you able to complete the task? (Yes/ No/ Partially)
- Did you face any difficulties in finding the correct option to click? Describe.
- Did you face any difficulties in performing the task? Describe.

Post-Task Questions

- Which task did you struggle with most and why?
- Did you face any difficulties in performing the task? Describe.

6.2.8. Result analysis

Some of the quantitative metrics that can be directly obtained from this evaluation method are

- Task completion: If there is a task that multiple participants failed to complete, it could directly translate to a major usability issue.
- Task completion time: This can be used to assess if a response is admissible. For instance, if participants have taken way too long between two tasks, then that probably implies an interruption to the test session. Such an interruption renders the response inadmissible.

In order to effectively analyze results, there should be more than a single evaluator responsible for analyzing the data. Start the data analysis only after all participants have submitted their response. Assign a random identifier to each test data and randomly assign each evaluator the order of analyzing each dataset.

Each evaluator should identify usability problems based on the participants' responses and/or video and audio recordings. In order to collate the findings of multiple evaluators, each of them should assign a severity rating to each usability issue- critical, serious or cosmetic. Once these ratings are in place, the evaluators should sit together and come to a consensus on each of the usability issues identified and its severity rating. This can be done by using a 'worst-case' schema, i.e. a problem is critical even if a single evaluator classifies it so.
This will result in a list of usability problems identified, giving designers and developers a focussed scope for areas of improvement in the Maisa system.

6.3. Eye Tracking

6.3.1. Requirements

- Since being additional method in moderated usability testing, the requirements follows those mentioned in the section 6.1.1.
- Specific to eye tracking, participants can't have hard contact lenses. However, soft lenses or lens correction set (comes with the eye tracking glasses) can be used.

6.3.2. Moderator Instructions

- You should inform the participants of using eye tracking before they arrive at the test place.
- Explain for the participants that the glasses are part of the usability testing and those are for recording the eye movements. However, you should be careful not to explain in a too detailed level since participants might come too aware of their eye movements.
- You should explain that the data is only collected when the participant is watching the screen.
- Eye tracking method is only used in moderated testing. Before starting the moderated usability testing, set-up the eye tracking glasses.
- The Begaze software is used for the analyzing and more detailed instructions for the analysis is explained in Result analysis.
- For setting up and calibrating the eye tracking glasses follow the instructions: https://www.aalto.fi/sites/g/files/flghsv161/files/2018-10/abl_manual_eyetrackinglasses_02_11_2017.pdf

6.3.3. Participant Instructions

- The product is being tested, not you, it's okay if you don't know or don't understand something.
- The usage of the eye tracking glasses is safe and doesn't require any action from the user. Those can be worn as normal glasses.
- The glasses follow your eye movements. Further, there will be a small dot in the recordings which shows where you watched.

6.3.4. Task list

The task list is the same for all the methods and is attached in the appendix section.

6.3.5. Result analysis

The eye tracking produces a video file with gaze dot and sound of the participant. With this the analysis phase is more robust. Further, there is always a chance to go back in time and see what the participant actually looked at. There are quantitative metrics and qualitative results from the eye tracking:

- **Gaze Plots:** Gaze plots is a quantitative metric from the testing. You're able to identify specific plots for each participant where they were looking. E.g. identify the most frequently looked spots on the page and how long it took to find the right navigation bar for instance. This can be used to see if some specific area or text field is hard to locate.
- **Heatmaps:** One option is to create heatmaps for individuals (it's recommended to use averaged heatmaps with 39 or more participants). However, it's also recommended to utilize all three methods in analysis.
- **Gaze Replays**: With this method, in the analysis phase, one watches the video from the glasses (⅓ of speed is recommended). This is a qualitative measurement and really supports the analysis phase in the sense of being the easiest way of getting "inside" of a participant's head.

6.4. UX Questionnaire

6.4.1. Requirements

- At least 20 participants for statistics, not only insight. For a reasonably tight confidence interval might need more. Sometimes participants would be divided into groups.
- Moderator/facilitator: a person who runs the test and assists participants, before they fill in the questionnaire.
- Evaluators to analyze the quantitative results form questionnaire
- Testing environment:
  An actual working instance of the portal with demo data in it. (including demo accounts so that participants don't need to use theirs).
- An electronic or paper task list
- An electrical questionnaire form in participants' first language.
- Testing space: usability lab or improvised lab where participants and facilitators can sit at a computer without being disturbed/distracted. It should be in a computer-equipped room with a computer dedicated to each of the participants.
- Instructions for moderator/facilitator and participants (provided below)
- Tasklist (provided in the appendix)
- Screening questionnaire (provided in the appendix)
- Consent forms (provided in the appendix)

6.4.2. Execution Steps

- Test run
- Recruiting of participants
- Grouping and guiding participants to the testing place
- The moderators introduced themselves and collect basic demographic data
- Introduction to the research topic and the intention of the test
- Consent signing
- The actual test session
- The participants are asked to explore and get familiar with the product
- The participant was then asked to solve the described tasks displayed on the screen or paper
- After the participant finished the last task, the screen was turned off
- The moderators introduce and answer the question from SUMI
- The participant fills out the SUMI Questionnaire
- Data analysis
- [optional] Can use two kinds of questionnaires in the same test

Time Estimate: The introduction session should be within 10 minutes, participants might use 20 to 40 minutes to finish their tasks(depending on the number of tasks). As we have around 20 to 30 people, reasonable grouping can save a lot of time. The questionnaire might need 10 to 20 minutes to fill in.

6.4.3. Participants

For the questionnaire, it is mainly for providing a quantitative view to combine with the data from other methods that focus on the qualitative view. This method needs enough participants to ensure reliability and accuracy. The recommended number of participants might be 20 to 30, tight confidence intervals require even more users. If the quality of the samples can be controlled well, 12 participants are also fine.
Based on our recommended target user group participants should be recruited from very new users of the system, users that haven't used it for quite a long time or people that are not users at all yet. That way the study avoids being biased by users that already remember all the wrong patterns that are potentially present in the navigation design.

6.4.4. Moderator Instructions

- This method is more relevant to quantitative research, so make sure you get involved with enough participants (often more than 30).
- Make sure there are no giant differences between each participant and their testing environments.
- Have a clear understanding about the standard questionnaire you choose. SUMI is used for user experience, while SUS, IBM and TAM are broadly used for usability.

- It is a well-defined, strictly controlled study condition. Make sure every task has a single well-defined answer.
- Get familiar with the questions and the tasks.
- Prepare a practice task to help participants get familiar with the study setup and with the product being evaluated to reduce variety in experts and novices.
- Meet and explain the purpose of the evaluation session and present the methodology of SUMI evaluation. In the explanatory session, the experimenters were present to assist with any difficulties with the questionnaire and to answer questions as they possibly arose.
- During the evolutionary session, we don't answer questions about questionnaires.
- Make sure the participant knows it's the product that is being tested not them.
- Make sure the participants understand the task in the same meaning.
- Usually there is no think-aloud in this testing
- Personal information will increase the variability of the study, because different people have different data.
- Almost all the standard questionnaires have visions in different languages, preparing what the participant is most familiar.

6.4.5. Participant Instructions

- The product is being tested, not you, it's okay if you don't know or don't understand something.
- Please be honest in your comments and feedback. Both good and bad are helpful.
- Start the formal testing whenever after you get familiar with the practice tasks.
- Choose the vision of questionnaire in your mother language
- During the testing sessions users were not allowed to ask the evaluator questions.

6.4.6. Task list

The task list is the same for all the methods and is attached in the appendix section.Participants also need to explore the website before they start doing the task lists.

6.4.7. Evaluation questions

Here, we choose SUMI (Software Usability Measurement Inventory) as the main method for user experience evaluation. We use this form as an example: http://sumi.uxp.ie/en/

6.4.8. Result analysis

The SUMI Questionnaire would provide information on efficiency, affect, helpfulness, control and learnability. There are already some automatic platforms that can be used for data analysis. The result can also be measured manually.
The results from SUMI evaluation are often presented by the terms of median, upper and lower confidence levels. These three levels come from the global usability scale. The median is the middle score when we set the answers into a numerical sequence. The upper and lower confidence limits reflect where 95% of true scores fall. The researcher also needs to calculate

the average score and standard deviation. It is also helpful to find which part of the platform getting the lowest or highest score.
Although we would involve a large number of participants in the questionnaire method, it is a self-reported data, which can be unreliable. They are subjective perception, not objective performance. The result would be more reliable when combined with performance metrics. Every participant might have different attitudes to the same rating scale. This result only tells you the satisfaction of the experience but not where the exact weakness or strength is.
Low quality or lack of enough participants would lead to unreliable results.

6.5. Final Recommendations

We presented 4 methods that should be used for evaluation of the navigation structure of the Maisa portal. Those methods serve are Moderated and unmoderated usability testing, eye tracking and UX questionnaires. Naturally from the description of those methods provided earlier, they serve as complementary tools to achieve the common objective of the study, which is to improve the navigation of the portal. For this very reason we have given a common task list.

The base for this study should be moderated usability testing coupled with UX questionnaires which will provide both the insight to how users understand the system, how they behave when interacting with the system and also what their feelings are from using it. Furthermore this combination provides clues about where problems with the design are and How is the mental model of users different from developers' mental model of the system.

We recommend starting with the combination of those two as it has relatively quick turnaround and combines nicely qualitative and quantitative data. Starting with moderated testing with few test participants also allows to remove biggest errors in the design before moving on to larger scale tests with more participants.

Once done with the initial phase of the test it will be beneficial to continue refining the system with more continuous testing using unmoderated tests, which makes the test be closer to real usage. In the end eye tracking can be used as an additional point of view during moderated testing to confirm observations of researchers or tell when the participants are subconsciously lying.

Since none of these methods are intended to be stand-alone methods our recommendation is that the data gathered from each method are analyzed together so that any results regarding the navigation are triangulated [18] and provide better evidence for the analysed results.

[18] https://interactions.acm.org/archive/view/november-december-2006/triangulation1

7. Conclusion and Future Work

Following the Heuristic Evaluation conducted, we developed this evaluation plan to test the usability issues identified in the Maisa portal. Another focus of this plan was the overall user experience of the portal. The plan focussed on examining issues with the information architecture of the system. Even though the focus may seem limited, we deemed it essential to remove the drawbacks of the current information architecture of the Maisa portal in order to address other usability concerns later on.

Even though the evaluation plan focuses on the most critical aspects of the portal, it has some drawbacks. The narrow focus of the plan may prevent researchers from identifying usability issues in other aspects of the portal. We considered this limitation in our planning, but it is overcome by the importance of building a robust information architecture as a foundation for the portal. Secondly, the tasks proposed in the task list might not address the objective of evaluation because we didn't have the chance to properly validate them against the portal. So when the actual evaluation is carried out, it is recommended to validate the task list. This will allow detecting and resolving issues such as incomplete requirements, missing information, ambiguous content, while also allowing an opportunity to add clarification. Thirdly, the evaluation plan assumes that the study is carried out with the final portal. Due to this, users may have a tendency to focus on visual design issues rather than other usability issues and the flow of organization within the system. In order to overcome this, we recommend using some low fidelity prototypes of the Maisa portal. This can be done in design tools like InVision or Figma, and will ensure that users do not get distracted by the visual language of the portal.

In order to proceed with this evaluation plan, we recommend going through the suggested methods to understand the further steps for each. Researchers must account for the limitations mentioned above and validate and refine the task list, if needed. Researchers must collaborate with the legal department to finalize the forms attached in the Appendix. If deemed necessary, researchers must also prepare some low fidelity prototypes of the portal to ensure they are collecting the right information from the users. Following this, recruitment of participants can be started in order to implement the rest of the evaluation plan.

8. Author's Contribution

This evaluation plan was written and compiled by Group A with a fairly equal amount of work from everyone. The main responsibilities were distributed as follows:
Jehan Khattak was responsible for the compilation of the document and gathering of the information needed for the appendix. She also contributed to the incorporation of the user study feedback for the final evaluation methods. Jehan also worked on the plan's objectives and general user groups.

Ondřej Brém was responsible for work related to Moderated Usability Testing and contributed to selecting of the methods, preparing tasklist and setting up the structure for describing the methods.
Shanshan Hou was responsible for work relevant to the UX Questionnaire. She also worked on making the prototype for our mini study. She also participated in method selecting and information gathering.
Katyayani Singh was responsible for the work related to unmoderated remote usability testing, while also contributing to method selection and preparing the tasklist.
Tomas Villikka was responsible for the work related to eye tracking as well some smaller parts of mini report etc.

9. Appendix

9.1. Consent Form

Consult the forms with whoever is responsible for legal matters to make sure you don't have any extra issues that need to be covered. Then print two copies, one to be signed and kept by the moderator and one copy for the user to keep.

Maisa Portal Evaluation
Participant Consent Form

Please read this page carefully before you sign it.

You have agreed to participate in a usability study that will evaluate the Maisa Portal. By participating in this study, you will help Maisa and Apotti to improve the Maisa Portal and its navigational structure.

Research staff will observe you and record information about how you work with the system. We will also ask you to fill out questionnaires about your experience and answer follow-up questions. We may record your comments and actions using written notes and video cameras. We will use the data from your study session, including videotapes, solely for the purposes of evaluating Maisa Portal and sharing the results of these evaluations with Apotti.

Your full name will not be used during any presentation of the results of this study. We will follow the standard GDPR principles.

By signing this form, you give your permission for Apotti to use any recordings that take place during the evaluation.

If you need a break at any time, please inform the study facilitator immediately. If you have questions about how the session will proceed, you may ask them at any time. You may withdraw from your study session at any time.

If you agree with these terms, please indicate your agreement by signing below.

Signature:
Print name:
Date:

9.2. Screening Questionnaire

Maisa Portal Evaluation
Screening Questionnaire

1: What is your name? *[Participants names should be anonymized when analysing results]*

2: Which age category do you fall under?

- 20-30
- 31-40
- 41-50
- 50-

3: Please select your gender:

- Male
- Female
- Do not wish to disclose

4: What is your occupation, if any?

5: Are you familiar with online booking systems? Such as booking appointments or scheduling?

6: Have you used an online healthcare system before? If yes, could you please tell us more.

7: How comfortable on a scale on 1-5 (with 1 being not comfortable at all, and 5 being an advanced user) are you with basic technology, like navigating web pages?

8: Have you ever heard of the Maisa Portal? If yes, have you used it?

Thank you!

9.3. Task List

Usability Evaluation of the Maisa Portal

For each of the following tasks read the description of the task to the participant. The expected outcome is just for your understanding and is not to be shared with the participant during the test.

1. **Appointment reservation**

    a. **Description:**
    Imagine you are about to travel to a country that requires a compulsory vaccination for Yellow Fever. Get an appointment with a practitioner to get one. [You are leaving in July and the shots need to be done at least two weeks prior to your departure]

    b. **Expected outcome:**
    The participant has booked an appointment with their general practitioner/vaccination center to get the shots before mid June. The date appears in the dashboard of upcoming events for the patient.

2. **Appointment Checking:**

    a. **Description:**

    Imagine you have booked several appointments for the upcoming month but do not remember the exact dates and time. Login to your homepage of the Maisa portal and find where your latest appointment details are. If there are any actions that need to be done before the appointments check those too, and complete the pre-appointment tasks.

    b. **Expected Outcome:**

    The user should use the top navigation bar and/or quick links to access the upcoming actions. This will allow the user to view all of their upcoming appointment dates and times. Additionally the user should be able to find the tasks that need to be completed before the appointment. They should follow the step by step navigation to finish the task.

3. **Checking your Test Results:**

    a. **Description:**

You have performed a blood test and have just been notified that the results of the test are now available. Login to your Maisa portal homepage and view the test results named “Blood Test Result- 1”.

b. Expected Outcome:

The user will be able to successfully login and use the top menu bar to correctly find where the “Results” will be located. Once located the user should navigate the page to find the exact results named “Blood Test Result -1”. Alternatively the user may use a quick link on the side menu bar to find the same results page.

4. Making and Editing an Appointment:

a. Description:

You need an appointment with your primary care physician next Monday 4pm, for an ache in your back. Use the quickest way you can find to book an appointment and then return to your homepage. After you booked your appointment you have realized that you can't make it on Monday after all and that you need to change the appointment time. Starting from the homepage, find the appointment and edit the time so that you now have it on Tuesday 4pm.

b. Expected Outcome:

The user should use the correct navigational menu to quickly find the option to make an appointment. They should be able to successfully follow the steps for making an appointment and by clicking the main logo they should be able to go back to their homepage. They should then be able to find the same appointment again and edit the time accordingly. These changes should be submitted.

5. Filling in Self-Assessments and Questionnaires:

a. Description:

Before going for your appointment you need to fill in a Self-Assessment and answer some Questionnaires. Find out where these forms are located and fill your details into them. You need to fill the Type-2 Diabetes Risk Assessment Form and Alcohol Risks (AUDIT) form.

b. Expected Outcome:

The user should use the top navigation bar, select My Activities, and then select the option ‘Self-Assessments and Questionnaires’. From the list shown, they should select the option Alcohol Risks (AUDIT), fill in the details, and Submit the questionnaire. The same procedure needs to be followed with ‘Type-2 Diabetes Risk Assessment Form’.