MASTER

Implementation of a user experience evaluation tool in the innovation process of MRI systems

de Groot, B.

Award date:
2021

Implementation of a user experience evaluation tool in the innovation process of MRI systems

Eindhoven University of Technology

Innovation Management

Date:	13-08-2021
Name:	Bas de Groot (1012333)
Mail:	(b.d.groot.1@student.tue.nl)
First assessor TU/e:	dr. J.C.C.M. (Boukje) Huijben
2nd assessor TU/e:	dr. Vikrant Sihag
3rd assessor TU/e:	dr. Jan A. Millemann
Company supervisor:	Koert Bloemers
2nd Company supervisor:	Rens Kruisbrink

Management Summary

Medical systems are complex systems to operate, and this complexity results in decreased safety and comfort of patients and operators of medical systems (Bitkina, Kim & Park, 2020). Developers need to better understand how operators experience medical systems innovations to be able to improve their medical systems (Maia & Furtado, 2016; Roto, Obrist & Väänänen-Vainio-Mattila, 2009). This research aimed to provide guidelines for developers on the implementation of tools to evaluate the experience of medical systems. The main research question that this research addressed was: *How can a standardized user experience evaluation tool be implemented into the innovation process of a company that develops Magnetic Resonance Imaging (MRI) systems?*

This research adopted a Design Science methodology, where knowledge derived from theory by conducting a systematic literature review was evaluated in the empirical context of a case company following three broad phases: (1) Exploration and synthesis, (2) Creation, and (3) Evaluation (Keskin & Romme, 2020). This research was conducted at a case company and focused specifically on the evaluation of MRI systems. Design principles were created based on the theoretical analysis that together with design requirements for the case company resulted in the creation of a solution design (Keskin & Romme, 2020).

The main result of this research is a solution design that provides guidelines for the implementation of a user experience evaluation tool in the innovation process of MRI systems. The solution design consists of two main phases: Preparation and Evaluation. The Preparation phase is needed to customize and select a tool for the specific needs of the company. The evaluation phase shows how to conduct the actual evaluation.

This research showed that the focus of evaluating medical systems in practice and academic literature is currently on evaluating pragmatic parameters (usability and utility) and that the focus should be increased to the whole user experience which includes hedonic parameters (looks and feel) next to the pragmatic parameters (Bitkina et al., 2020). The results of this research show that developers should use standardized user experience evaluation tools that allow to include the relative importance of pragmatic and hedonic parameters for operators as these tools are reliable and developers can easily adopt these tools (Díaz-Oreiro et al., 2019; Laugwitz et al., 2008).

This research emphasizes the importance of benchmarking user experience scores to give meaning to the scores. Instead of either benchmarking scores to previous internal evaluations or an online database of previous evaluations, this research showed that the two can be combined to be comprehensive and facilitate the interpretation.

Furthermore, developers should repeat the evaluations over time and specify the specific context of the evaluation by adding labels to the evaluation, as this facilitates the understanding of the temporal and contextual factors of the user experience.

The steps of the preparation of the solution design have been conducted for MRI systems and the results of this research can therefore be used for the evaluation of MRI systems for the case company and other companies that want to evaluate MRI systems. For developers of other medical systems, the preparation steps of the solution design can be followed to be able to implement a user experience evaluation tool in the innovation process.

Table of Contents

List of Figures

List of Tables and Key abbreviations

Key abbreviations:

MRI – Magnetic Resonance Imaging

KPI – Key Performance Indicator

UEQ – User Experience Questionnaire

FOK – First of a Kind

UX – User Experience

AD – Advanced Development

PDLM – Planning Development Launch and Maintenance

1. Introduction

Medical systems are based on continually evolving technologies and the full risks associated with the working environment of medical systems are not fully understood (Pickup, Nugent & Bowie, 2019). Difficulties with operating medical systems directly affect the lives of people and therefore pose a major threat (Chandran, Al-Sa'di & Ahmad, 2020). The design of medical systems is an important process because of its direct influence on the safety and comfort of patients and medical staff (Bitkina, Kim & Park, 2020).

User-centered design is an increasing trend for general products, but recently also for medical systems, that places the needs of the user central to the innovation process to increase the usability and user experience of products and systems (Stola, 2018). Usability and user experience are both focal points of the higher-level concept of user-centered design. According to a literature review by Bitkina, Kim, and Park (2020), most researchers regard usability as a subset of the user experience of a product. For this reason, this research will also regard usability as a subset of the user experience.

The definition for user experience (UX) according to the International Organization for Standardization (ISO), consists of the user's perception and responses that result from the use or anticipated use of a system, product, or service (ISO, 2019). The definition for usability according to the ISO, consists of the effectiveness and efficiency of a system and the satisfaction that a user experiences with the system (ISO, 2019). These definitions will be used in this research when referring to the user experience and usability.

Developers of medical systems aim to prevent misuse and thus often focus on improving the usability and safety of their products (Bitkina et al., 2020). Although medical systems were previously merely tested on usability, there is a need to broaden the scope to the whole user experience of medical systems (Bitkina et al., 2020). It is necessary to understand how a user feels to improve the experience of users with a product (Maia & Furtado, 2016; Roto, Obrist & Väänänen-Vainio-Mattila, 2009).

As innovation processes for designing medical systems become increasingly focused on the users, developers subsequently want to measure the value of the user experience of their products to understand how well they fulfill the needs of their users (Bitkina et al., 2020; Zarour & Alharbi, 2017). Measures that can be tracked throughout the innovation process to see whether objectives will be met, can be integrated into the process, referred to as Key Performance Indicators (KPI) (Pan & Wei 2012). Many innovation processes include objective measures, but there is less focus on the subjective measures of innovations (Velimirović, Velimirović & Stanković, 2011). If solely relying on the objective measures, the perspective of the user is missing in the innovation process. Quantification of the user experience will allow developers to have a performance indicator from the user perspective on their innovations instead of merely technological or financial indicators.

User experience is a fuzzy term in academic literature and there is no clear consensus on what parameters determine the user experience and how to measure it in different contexts (Lallemand, Gronier & Koenig, 2015). Although the user experience of general products has previously been researched, there is a gap in the academic literature on the implementation of quantitative user experience tools in the innovation process and a gap in the literature on the important parameters of the user experience in the context of medical systems (Bitkina et al., 2020). To understand how user experience tools can be implemented in the context of medical systems, a case study was conducted. The context of this research and the problem statement will be discussed in more detail in Section 1.1.

The case company is working on innovating its medical imaging systems and making them as user-friendly as possible. For this purpose, the case company is increasingly focusing on implementing user-centered methods in its innovation process. There is a need for the case company to quantify the user experience of their medical imaging systems so the perspective of the operators of medical systems can be better incorporated in the innovation process. Quantifying the user experience will allow the case company to quickly assess the impact of innovations and track the impact of innovations on the user experience of clinical system operators over time. This research focused on developing guidelines for implementing a quantitative tool for measuring the user experience of operators of Magnetic Resonance Imaging (MRI) systems in the innovation process of the case company. This research focused on the Dutch context.

This research attempts to fill the academic research gap on guidelines for implementing a quantitative user experience evaluation tool in the context of MRI systems. Process guidelines were developed for implementing a user experience evaluation tool for the case company to evaluate the perspective of the operators in their MRI innovation process. The developed solution design explains the steps for the preparation and evaluation of the user experience of MRI systems. This research adopted a Design Science methodology, as it allows the combination of theoretical and empirical knowledge and can result in a practical solution, which will be described in detail in the methodology section (Keskin & Romme, 2020). The next section will discuss the case company context and the problem statement for the case company.

1.1 Case company context and problem statement

The context of the case company and the problem statement for the case company will be shortly discussed here. The case company has two innovation processes for MRI systems named Advanced Development (AD) and Planning, Development, Launch, and Maintenance (PDLM). This research will specifically focus on the First of a Kind period of the PDLM process. Both innovation processes and the focus on the FOK period will be explained in detail in Section 4.1.

The case company is involved in developing different types of medical systems, each based on different technologies and with different medical applications. They are working on developing high-quality systems to help improve the healthcare of many countries around the world. The medical systems that they are developing include medical imaging systems, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) systems. Although the aim for the case company was to measure the user experience for all their medical imaging systems, this research focused on MRI systems specifically. MRI is a medical application of nuclear magnetic resonance which is used for medical diagnosis. MRI systems are complex systems, which makes the systems difficult to use and can therefore impact the safety of patients and operators of MRI systems (Pickup, Nugent & Bowie, 2019).

There are several different types of clinical operators of MRI systems. The most important primary operators of MRI systems in Dutch hospitals are radiologists, technologists, and lead-technologists (Shah & Robinson, 2008). The technologists operate the MRI system and guide and prepare the patient. The lead-technologists are responsible for the scanning of patients on a higher level in terms of preparation. The radiologists interpret the scan results in terms of a medical diagnosis. The case company is aware that these clinical operators experience a lot of time pressure and stress due to the complexity of the systems which leads to low satisfaction with the systems. The case company wants to improve the experience of the clinical operators with their medical systems by determining the value of the innovations for the clinical operators throughout the innovation process.

The case company has clinical application specialists that work full-time within the case company as well as clinical application specialists that work full-time at hospitals by helping users operate MRI systems. The clinical application specialists that work full-time within the case company are focusing on integrating the user perspective in innovations but also are continuously in contact with clinical system operators in hospitals to help them if there are any issues with the MRI systems. The clinical application specialists are representatives or proxies of the previously described clinical system operators within the innovation process of the case company. Most clinical application specialists have previously worked in hospitals as clinical operators of MRI systems. Additionally, the patients are very important stakeholders as they are the final beneficiary of the MRI systems (Shah, Robinson & AlShawi, 2009). As the focus of this research was on the experience of operating MRI systems, patients were not involved in this research.

This research focused on quantitatively measuring the user experience of clinical system operators of MRI systems in Dutch hospitals. As the focus was on guidelines for the implementation of a user experience evaluation tool in the innovation process of the case company, the empirical analysis was done with employees that are involved in the innovation process of the case company. This research focused on MRI system operators, which will be referred to as operators from now on in this research.

The case company wants to know what the important parameters of the user experience of their MRI systems are and how they should be measured. The case company wants to increase their knowledge of the concept of user experience and implement new measures in its existing innovation processes. A requirement for the case company is to use quantitative measures, as these allow to quickly assess the impact of innovations on the user experience of their MRI systems and can be tracked throughout the innovation process. This means that tools that focus on qualitative detailed information were excluded from this research. The focus of this research was on quantitative tools, which are often questionnaires, that transform the subjective experience of operators into a score often with the use of a Likert scale (Díaz-Oreiro, López, Quesada & Guerrero, 2019).

As the case company is a global company, they do not only have different types of users but also users with individual and cultural differences. The case company wants to understand the differences between users in terms of their needs, so they can take these differences into account when developing their MRI systems. As the case company eventually wants to incorporate different types of users and wants to use a user experience evaluation tool for different medical imaging systems, it is desired to have a flexible user experience evaluation tool. This means that the tool can be used throughout different stages of the innovation process, for different medical imaging systems and it can be used for operators with individual and cultural differences. This research focused on the medical imaging technology MRI and operators working in the Netherlands as a pilot test for the further development of the implementation guidelines. Future research can focus on other contexts based on the results of this research.

This research had access to online databases from the case company about current innovation processes. Furthermore, several employees from different departments of the case company were interviewed, which will be discussed in detail in the methodology section. This research conducted a systematic academic literature review in preparation for the next steps of this research. The research gaps that were identified from the results of this literature review will be described in the next section below.

1.2 Academic research gaps

This research started with a systematic literature review. Two academic research gaps that were identified from the literature review will be discussed here, as they guided the further steps of this research. The methodology of conducting the literature review will be described in the methodology section below. The full results of the literature review can be seen in the theoretical results section of Chapter 3.

Research gap 1. *The important parameters of the user experience of MRI systems are not defined*

As discussed before, developers of medical imaging systems currently often focus on measuring the usability of their systems, which consists of merely evaluating the pragmatic (utility and usability) parameters of systems and not the hedonic (looks and feel) parameters of the systems (Borsci, Buckle & Walne, 2020; Hassenzahl, 2003). Developers are likely to benefit from focusing on the whole user experience, but it is not clear yet what the user experience entails in the medical context and thus what specific parameters should be measured (Bitkina et al., 2020). Previous research has shown that it is difficult to define the parameters of the user experience of all types of products as there are different interpretations of the user experience both in practice and in academic literature (Lallemand, Gronier & Koenig, 2015). Some researchers have aimed to define a comprehensive list of important parameters for the user experience of general products (Hassenzahl, 2003; Mahlke, 2008; Zarour & Alharbi, 2017). Different lists of parameters are defined in academic literature, which indicates that a single list of parameters is probably not possible for all types of products (Mahlke, 2008; Zarour & Alharbi, 2017). The important parameters of the user experience seem to be dependent on the specific context of the system (Hassenzahl, 2003; Mourouzis et al., 2006). The user experience of medical imaging systems, including MRI systems, has not been considered in previous academic research. If developers of MRI systems want to understand and measure the user experience of their systems, they should understand what the important parameters are in the context of MRI systems.

Research gap 2. *Guidelines for the implementation of a standardized user experience evaluation tool in the innovation process of MRI systems are missing*

As discussed above, the important parameters for the user experience of MRI systems are not defined yet. If the parameters are defined, developers still need tools and guidelines to be able to measure the defined parameters (Zarour & Alharbi, 2017).

Previously, developers in the medical context mostly used usability tools to evaluate their products (Sousa & Lopez, 2017). There is a need in the medical context to broaden the scope to evaluate the whole user experience (Bitkina et al., 2020). Three standardized tools for measuring the whole user experience are currently available in academic literature, which are the Modular Evaluation of Key

Components of User Experience (meCUE), Attrakdiff, and the User Experience Questionnaire (UEQ) (Díaz-Oreiro et al., 2019).

The three user experience evaluation tools have each been used in the medical context, but more research is needed to identify the most appropriate tool for MRI systems and guidelines need to be developed on how the tools should be implemented in the innovation process (Zarour and Alharbi, 2017). The UEQ seems to be the most appropriate tool for the context of MRI systems, due to the modularity and flexibility of the tool (Hinderks et al., 2019). This will be discussed in detail in Chapter 3.

Guidelines for the selection of the tool for the user experience evaluation as well as guidelines for the full process of preparation and evaluation of the user experience, are currently missing in academic literature. Furthermore, more research is needed on the context of the user experience evaluations (Zarour & Alharbi, 2017).

This research addressed both these research gaps by combining theoretical knowledge from previous academic research with the empirical context of the case company (Keskin & Romme, 2020). The research questions that followed from these research gaps and the problem statement will be described in the next section.

1.3 Research questions

The identified research gaps in academic literature and informal interviews with employees from the case company in preparation for this research resulted in one main research question and 5 sub-questions. The sub-questions are subsets of the main research question and together answered the main research question, which is visualized in Figure 1.

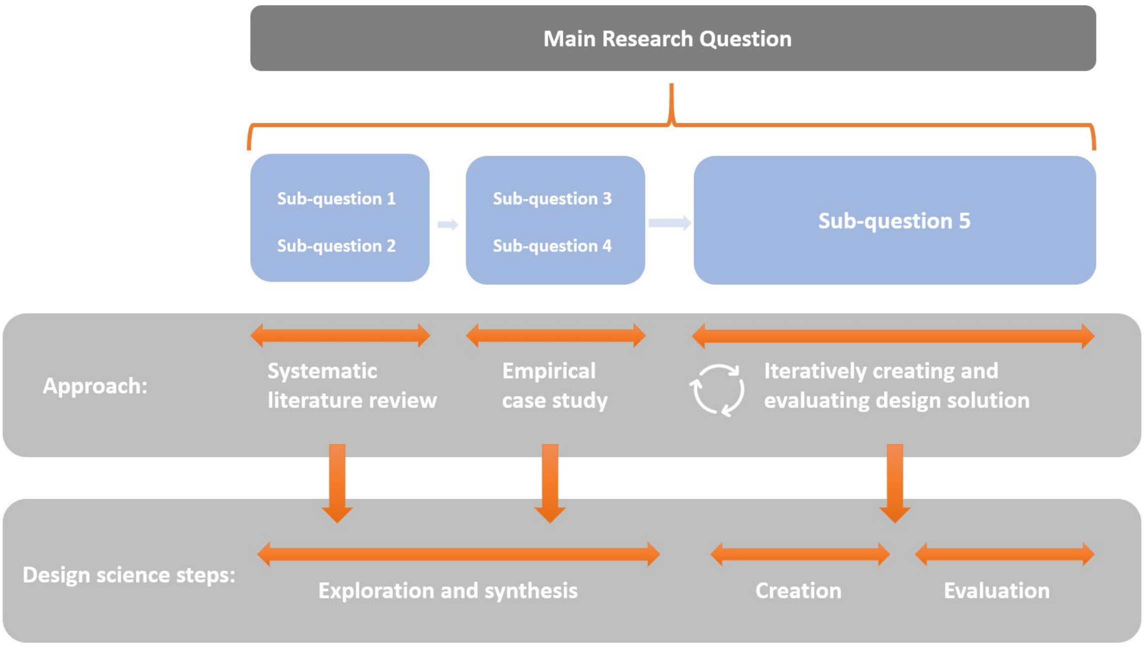


Figure 1. Research Design

The research questions together addressed the academic research gaps 1 and 2 described in the previous section. To answer the main research question, this research adopted the Design Science methodology, which is divided into phases of Exploration and synthesis, Creation, and Evaluation, which will be described in detail in the methodology section (Keskin & Romme, 2020). The first two sub-questions were answered by conducting a systematic academic literature review. The second two sub-questions were answered by analyzing a database of current innovation processes of the case company and conducting semi-structured interviews and informal interviews with employees from the case company. For answering the last sub-question, the answers from the first four sub-questions were used to iteratively create and evaluate a solution design with the use of focus groups. The process of each step will be described in the methodology section.

Main Research Question: How can a standardized user experience evaluation tool be implemented into the innovation process of a company that develops MRI systems?

To answer the main research question, each sub-question needed to be answered. The first sub-question was answered by identifying the important parameters of the user experience but also the previously used tools to quantitatively evaluate the user experience in academic literature. The second sub-question was answered by analyzing the academic literature on the implementation of quantitative user experience tools in the innovation process of companies. The first two sub-questions were addressed by a systematic literature review which was conducted in the exploration and synthesis phase of the Design Science methodology (Keskin & Romme, 2020). These two sub-questions address research gap 1 described before.

Sub-question 1: What are important parameters of the user experience and what standardized tools can developers use that can quantitatively evaluate these parameters according to academic literature?

Sub-question 2: How can a standardized user experience evaluation tool be implemented in the innovation process according to academic literature?

After addressing the first two sub-questions based on academic literature, the research shifted to the empirical analysis with employees from the case company. The insights from the theoretical analysis of the first two sub-questions were used in the empirical analysis to identify the design requirements for the case company and to determine the appropriateness of the identified user experience tools in academic literature for the context of MRI systems (Hevner, 2007). Sub-question 3 was answered by analyzing the current innovation process of the case company to understand how a standardized user experience evaluation tool can be implemented into its innovation process. Sub-question 4 was answered by identifying the design requirements for implementing a standardized user experience evaluation tool in the innovation process of the case company. The third and fourth sub-questions were addressed by an empirical analysis with the case company in the phase of exploration and synthesis of this research (Keskin & Romme, 2020).

Sub-question 3: How does the case company currently innovate their MRI systems and how do they integrate the perspective of operators in the innovation process?

Sub-question 4: What are the design requirements for the implementation of a standardized user experience evaluation tool in the innovation process of MRI systems for the case company?

The answers to the first two sub-questions were combined with the answers to the second two sub-questions to answer the fifth sub-question. To answer the fifth sub-question, a solution design for the implementation of a standardized user experience evaluation tool for the case company was iteratively created and evaluated. This means that the insights from the exploration and synthesis of the first four questions were used in the form of design requirements and design principles for the creation and evaluation of a design solution for the case company (Keskin & Romme, 2020). The fifth sub-question addressed research gap 2 on the lack of practical guidelines for the implementation of a standardized user experience evaluation tool in the innovation process, which was described before (Bitkina et al., 2020). This sub-question focused specifically on the First of a Kind (FOK) period of the innovation process of the case company, which will be discussed in detail in Section 4.1.

Sub-question 5: How to implement a standardized user experience evaluation tool to measure the experience of operators of Magnetic Resonance Imaging (MRI) systems in the First of a Kind (FOK) period of the MRI innovation process of the case company?

This research aimed to create a design solution for the evaluation of the user experience of MRI systems within the FOK period of the MRI innovation process of the case company, by combining the theoretical and empirical analysis of the Design Science methodology (Keskin & Romme, 2020). The next chapter will discuss the methodology of this research in detail.

2. Methodology

This chapter outlines the research methodology of this research. This research adopted the Design Science methodology, which will be discussed in broad terms according to the generic Design Science cycle introduced by Keskin and Romme (2020). Following the discussion of the research design, the methodology of the systematic literature review and the empirical analysis for the phase of Exploration and synthesis will be described. Afterward, the methodology of the iterative creation and evaluation of a design solution for the case company will be described. This chapter concludes by discussing the validity and reliability of this research.

2.1 Research design

This research adopted a theory-driven Design Science methodology to answer the previously defined research questions in Section 1.3 (Keskin & Romme, 2020). As previous research about the evaluation of the user experience has mostly remained theoretical, more practical knowledge is needed to be able to implement a user experience evaluation tool into innovation processes (Zarour & Alharbi, 2017). Adopting the Design Science research methodology allowed on the one hand to bring a pragmatic solution for the case company and on the other hand, addressed the gap of academic research on guidelines for the implementation of a user experience evaluation tool in the innovation process. Additionally, the Design Science methodology allowed switching between theory and practice which aided in answering the research questions, as new theory could iteratively be added and empirically validated to determine the appropriateness to the context of MRI systems (Keskin & Romme, 2020).

The generic Design Science cycle is an iterative process that consists of the phases of Exploration, Synthesis, Creation, and Evaluation, as can be seen in Figure 2. This research iteratively switched between the different phases but followed three main phases which will be explained below: *(1) Exploration and synthesis, (2) Creation, and (3) Evaluation* (Keskin & Romme, 2020). A general explanation of the research design of this research will be given first and afterward, each phase will be described in detail.

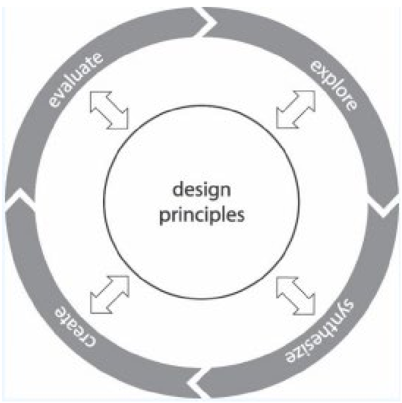


Figure 2. The Generic Design Science cycle. From “Mixing oil with water: How to effectively teach design science in management education?” by D. Keskin and G. Romme, 2020, *BAR-Brazilian Administration Review, 17,* p. 11.

The research design of this research is visualized in Figure 3. The first main phase referred to as Exploration and synthesis, consisted of a theoretical and empirical analysis. For the theoretical part, a systematic literature review was conducted on the topics of theoretical frameworks of the user experience, practical tools for evaluating the user experience, and considerations for the implementation of a user experience evaluation tool in an innovation process. The literature review addressed the sub-questions 1 and 2 that were previously described and additionally resulted in constructing design principles that guided the solution design for the case company. The empirical part of the Exploration and synthesis phase consisted of conducting formal semi-structured interviews, examining current innovation processes available in the case company database, and conducting informal interviews. The empirical analysis in this phase resulted in a description of the current MRI innovation process of the case company and the design requirements for a solution design for the case company. The empirical results of the first phase addressed sub-questions 3 and 4. The insights from the theoretical and empirical part were synthesized into design principles and design requirements that guided the development of the solution design creation for the case company, which will be explained below.

The second main phase referred to as Creation, consisted of using the developed design principles and design requirements from the first phase of Exploration and synthesis to develop a solution design for the case company. The focus of the solution design was on developing guidelines for the implementation of a user experience evaluation tool in the MRI innovation process of the case company.

The last main phase referred to as Evaluation, consisted of empirically evaluating the developed solution design both in terms of pragmatic value and theoretical value by reflecting on the constructed design principles. Together the phases of Creation and Evaluation answered the fifth sub-question. All three main phases of this research will be described in detail below.

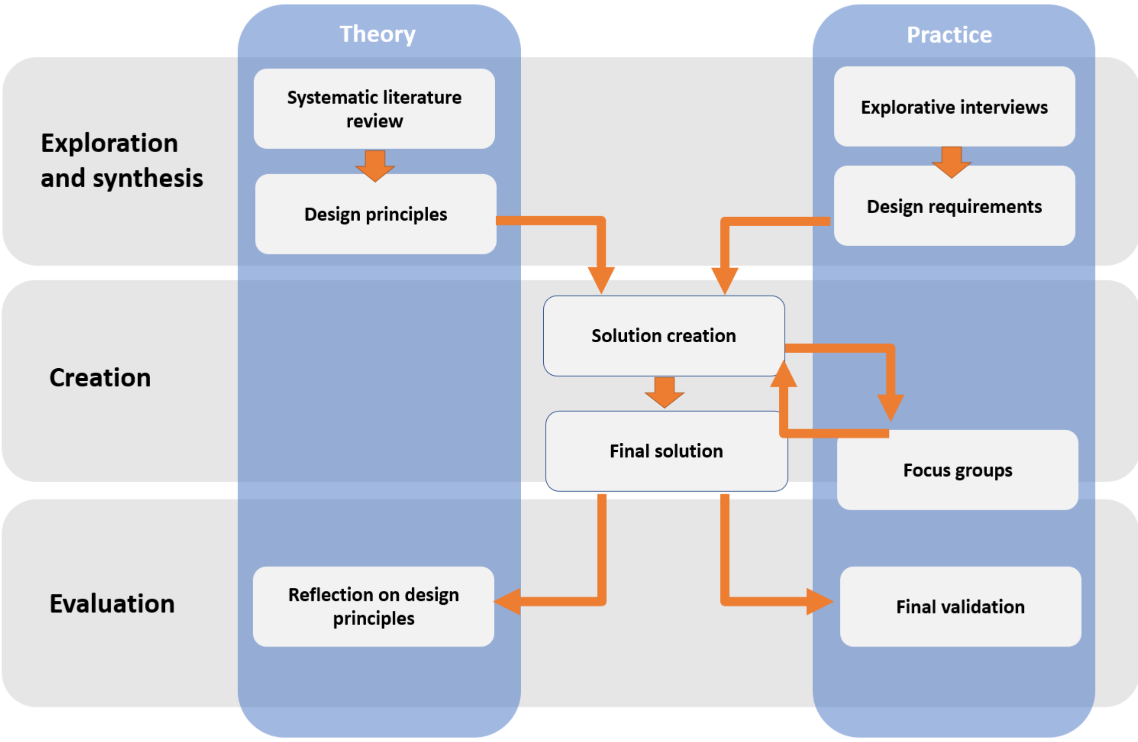


Figure 3. Research Design steps

2.2 Exploration and synthesis

The first main phase of this research examined the problem from a theoretical and a practical perspective, which will be described here. First, the methodology for the theoretical part will be discussed, which consisted of conducting a systematic literature review. Then the methodology for the empirical analysis will be described, which consisted of identifying the current MRI innovation process of the case company, identifying the design requirements for a solution design for the case company, and empirically evaluating insights from the academic literature review (Hevner, 2007). This chapter will conclude by describing the methodology of synthesizing the theoretical and empirical parts. The synthesis consisted of developing design principles and design requirements for the development of a solution design for the case company. As suggested by Keskin and Romme (2020), there was switched between the theoretical and empirical analysis during the research.

2.2.1 Theoretical analysis

The first step of this research was to examine the problem from a theoretical perspective, by conducting a systematic academic literature review. The subject of this literature review was based on the problem statement of the case company, which was identified via informal interviews. A literature review can have different aims, which then guide the specific steps of conducting the literature review to fulfill these aims (Randolph, 2009). The academic literature review had two main aims in this research. The first main aim was to delimit the research problem by identifying gaps in current academic literature (Gall, Borg & Gall, 1996). The second main aim was to relate the ideas and theory from previous studies to practice. The methodology of this literature review will be outlined in this chapter by discussing the taxonomy, the theoretical lenses, and the search strategy (Randolph, 2009). The literature review provided an answer to sub-questions 1 and 2.

2.2.1.1 Taxonomy

Defining the characteristics of a literature review helps to ensure an effective process of conducting a literature review (Randolph, 2009). Cooper (1988) identified six characteristics to define a literature review, which was elaborated on by Randolph (2009). The characteristics of a literature review defined by Cooper are (1) Focus, (2) Goal, (3) Perspective, (4) Coverage, (5) Organization, and (6) Audience. The characteristics and their respective category for this research are shown in Table 1 and will each be explained below.

Table 1. Characteristics of the literature review using the taxonomy of Cooper

Characteristic	Category
Focus	Research outcomes and practices or applications
Goal	Integration, Criticism, and Identification of central issues
Perspective	Espousal of position
Coverage	Central or pivotal
Organization	Conceptual
Audience	Supervisors, practitioners, and researchers

The focus of the literature review was on research outcomes and practices or applications (Randolph, 2009). The focus on research outcomes was to delimit the research problem by identifying research gaps in academic literature. The focus on practices or applications was to identify relevant insights for the implementation of a standardized user experience evaluation tool in the context of innovating MRI systems to help guide the solution creation of this research.

The goals for this literature review, using the taxonomy created by Cooper (1988), were to integrate available literature, to critically look at previous literature, and to identify the central issues in literature. The main goal was to determine the appropriateness of previous theories and tools for evaluating the user experience in the context of MRI systems (Randolph, 2009).

Although the aim was to be neutral in the representation of the current academic literature, the perspective might be more linked to the espousal of position, which means that the researcher took an active role in the accumulation and selection of relevant academic literature (Cooper, 1988; Randolph, 2009). Randolph (2009) suggests that is best to explain the preexisting biases when conducting qualitative research. The main bias in this research is that some articles were selected to be useful for designing a solution design for the case company. This also means that some articles were left out of the literature review as they did not provide useful information for the case company.

In terms of coverage, it is difficult to be exhaustive in this field, due to the different interpretations of user experience and the many different tools related to evaluating the user experience in academic literature (Lallemand, Gronier & Koenig, 2015). For this reason, the coverage of this literature review was central and pivotal (Randolph, 2009). This means that articles were selected that had an important contribution to the field of user experience evaluation by for example proposing new tools or research directions (Cooper, 1988).

The organization of the literature review was conceptual (Randolph, 2009). The review was organized conceptually around three main areas of academic literature, which will be discussed in the next section.

The audience of the literature review is the supervisors of the Eindhoven University of Technology, the supervisors of the case company, and practitioners and researchers interested in the implementation of a user experience evaluation tool in the innovation process of MRI systems (Randolph, 2009).

2.2.1.2 Theoretical lenses

As there are many different definitions of the user experience in academic literature, the literature review will start by explaining the definition of the user experience that was adopted in this research (Lallemand, Gronier & Koenig, 2015). After discussing the definition adopted in research, the literature review focused on three main areas of academic literature, which are visualized in Figure 4. The three lenses depict the scope of the literature review, which each had separate search strings to find relevant articles within each lens.

The first lens focused on previously developed key theoretical frameworks that explain the parameters of the user experience that can be used to evaluate the user experience of MRI systems. This lens aimed to identify whether any theoretical frameworks show what the parameters of the user experience of MRI systems are and could therefore be used in the next steps of this research.

The second lens focused on the current standardized tools in academic literature for evaluating the user experience of all types of products and systems. These tools were assessed on the appropriateness for the context of MRI systems.

The third lens focused on the current academic literature on considerations for the implementation of a user experience evaluation tool in the innovation process of companies. This lens focused on considerations for the implementation of the user experience in the innovation process of general products as the research on implementation in the medical context, and specifically on MRI systems, is limited (Bitkina et al., 2020). Together the three lenses resulted in identifying the two academic research gaps described before. The main insights for the development of a solution design for the case company from the literature review were synthesized in the form of seven design principles, which will be discussed in the synthesis section below.

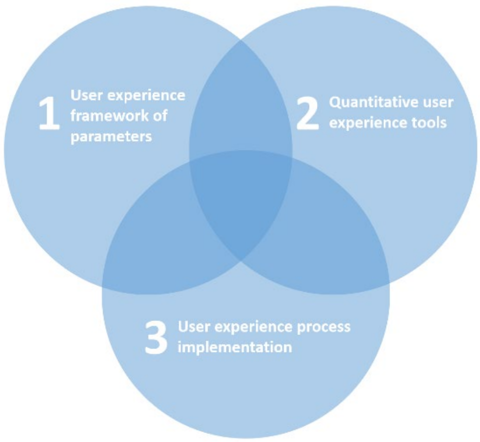


Figure 4. Theoretical lenses of the literature review

2.2.1.3 Search strategy

The search strategy for finding relevant articles during the academic literature review consisted of several steps. Web of Science was used as the primary database for the literature review ("Web of Science", 2021). Following an initial search to get acquainted with the relevant terminology, search strings were made for each of the lenses described above. An iterative approach was adopted to derive the most accurate search strings of which the final search strings are included in Appendix 1. Only journal articles and conference papers were included and only if they were available in English. Articles that focused specifically on the medical context or MRI systems were included. To assess which articles should be included, the derived articles from the search strings were first analyzed based on the relevancy of the title, then on the relevancy of the abstract.

As the user experience is a fragmented field in academic literature, some additional methods for finding relevant academic literature were used. Three secondary methods for searching relevant literature were adopted: expert recommendation, Google Scholar, and a snowballing technique. Google Scholar was used for finding relevant articles on specific topics found in the articles from the primary database. Especially looking forward and backward in references, often referred to as forward and backward snowballing (van Aken, Berends, & van der Bij, 2012) resulted in many relevant articles. The final list of included articles with the method of finding each article is added to Appendix 2.

2.2.2 Empirical analysis of Exploration and synthesis

The insights derived from the literature review guided the second part of the exploration. The second part of the exploration consisted of an empirical analysis by conducting semi-structured interviews, informal interviews, and examining an online database of the case company. These different methods for the empirical analysis functioned as method triangulation which means that different methods were used for the analysis of the same subject, which increases construct validity (Carter, Bryant-Lukosius, DiCenso, Blythe & Neville, 2014). The informal interviews are documented in field notes which can be seen in Appendix 3. The formal interviews were semi-structured as it does not restrict both the interviewer and the interviewee in an open discussion, but it does give some structure to the conversations, which makes it a useful flexible method for qualitative research (Longhurst, 2003). Two different guidelines were developed for the semi-structured interviews in this phase, focusing on the MRI innovation process of the case company and the requirements and needs for the case company. The interview guides can be seen in Appendix 4, together with PowerPoint slides that were created to facilitate the discussion.

The empirical analysis in the phase of Exploration and synthesis had multiple purposes. The first purpose was to identify the design requirements for a solution design for the case company, which provided an answer to Sub-question 3. The second purpose was to identify the current innovation process of the case company, to understand how a user experience evaluation tool can be implemented in the innovation process, which provided an answer to Sub-question 4. The last main purpose was to discuss insights from theory to determine the appropriateness of the insights to the context of MRI systems.

As the focus of this research was to develop guidelines for the implementation of a standardized user experience tool into the MRI innovation process of the case company, the semi-structured interviews were conducted with relevant employees of the case company, which are shown in Table 2.

Table 2. Participants involved in the empirical analysis

Department	Number of participants	Language
Clinical Applications	5	All Dutch (Quotes translated)
Design	2	Both English
Quality and Regulatory	1	Dutch (Quotes translated)
Marketing	1	English

To assess the problem from different perspectives, data source triangulation was used, which means that employees from different departments of the case company were involved in the empirical analysis (Carter et al., 2014). Five different employees from the department of Clinical Applications were interviewed during the Exploration and synthesis phase, as they are representatives of the operators within the case company. One employee from the department of Quality & Regulatory, one from the department of Marketing, and two employees from the department of Design were involved in the empirical analysis as their departments are all involved to some degree in the evaluation of the experience of operators with MRI systems.

One interview with Clinical Application specialist 4 and one interview with Designer 1 focused specifically on the current MRI innovation process of the case company and requirements and needs were only shortly discussed during these two interviews. The interview guide A for these two interviews can be seen in Appendix 4.

The other seven interviews focused specifically on the requirements and needs for the case company for the implementation of a user experience evaluation tool in the MRI innovation process. The interview guide B for these interviews can be seen in Appendix 4.

To ensure that ethical and privacy requirements were met, the participants were asked to give informed consent via a consent form and the process for conducting this research was subjected to the Ethical Review Board (ERB) from the Eindhoven University of technology, of which the approval can be seen in Appendix 5. The consent form that was used during this research can also be seen in Appendix 5. Next to the empirical analysis in the exploration phase, the solution design was iteratively evaluated empirically with the use of focus groups, which will be discussed in Section 2.4. The next section will discuss the data analysis of the phase of Exploration and synthesis.

2.2.2.1 Data analysis of Exploration and synthesis

After the completion of the semi-structured interviews, the results of these interviews were analyzed in combination with the informal interviews and main insights from the database of the case company. As discussed before, the empirical analysis in the exploration phase focused on understanding the current innovation process and user involvement within the case company and the design requirements for the creation of a solution design for the case company.

The semi-structured interviews from the empirical analysis were transcribed, coded, and analyzed to develop insights that can be used for the development of a design solution. The coding was done systematically to categorize the insights and aid the analysis. The qualitative data analysis software NVivo was used to code and analyze the semi-structured interviews from the empirical analysis (NVivo, 2018). The provisional coding method was adopted as it allows to be flexible in the coding as knowledge on the subject develops (Saldaña, 2009). Provisional coding is an exploratory method that uses pre-specified provisional codes based on a literature review. The codes can be revised, modified, deleted, and expanded based on the qualitative data throughout the analysis (Saldaña, 2009). The start codes and the final resulting codes can all be seen in Appendix 6.

The transcriptions of all the interviews are added to Appendix 7. The results from the theoretical and empirical analysis were synthesized into design principles and design requirements, which will be explained below.

2.2.3 Synthesis

The first main phase of Exploration and synthesis involved synthesizing the insights from the theoretical and empirical analysis into design principles and design requirements.

The theoretical analysis described above resulted in seven design principles on four main topics, which were created to guide the creation of a solution design for the case company (Keskin & Romme, 2020). Design principles can be described as *"involving a coherent set of normative ideas and propositions, grounded in research, which serve to design and construct detailed solutions"* (Van Burg et al., 2008, p. 116). In this paper, CAMO logic (Denyer, Tranfield, Van Aken, 2008; Van Burg & Romme, 2014) was used to formulate the design principles. The CAMO logic makes it possible to formulate the design principles targeted at a specific context (C), where a certain action (A), triggers a particular Mechanism (M), to get to the desired outcome (O) (Keskin & Romme, 2020). The seven design principles and the four main topics that resulted from the theoretical analysis will be described in Chapter 3. The design principles were used in the final phase of this research to reflect on how useful the design principles were for the creation of a solution design for the case company. Furthermore, two additional design principles were created based on the final validation of this research.

As discussed before, design requirements were synthesized from the semi-structured interviews conducted at the exploration phase (Hevner, 2007). The insights from the semi-structured interviews and informal meetings with different employees were synthesized to a final list of requirements, which will be shown in Chapter 4. The development of the requirements guided the identification of the design space for creating a solution design for the case company (Dresch, Lacerda & Antunes, 2015). The design requirements were separated into functional requirements, user requirements, boundary conditions, and design restrictions (van Aken et al., 2012). The functional requirements were focused on what the tool must be able to do and therefore form the basis of the creation of a solution design. The user requirements were focused on specific requirements that were relevant for users in the use of a user experience evaluation tool. The boundary conditions showed what the external boundaries for the case company were in which a solution design could be created. Design restrictions were focused on requirements that show the solution space for the case company to make the design successful.

2.3 Creation

After the first main phase of Exploration and synthesis was concluded, the second main phase of Creation started. The synthesized design principles and design requirements from the Exploration and synthesis phase were used as guidelines to develop a solution design for the case company. The process of developing the solution design was an iterative approach, consistent with the suggestion by Keskin and Romme (2020). This means iteratively switching between the creation of the design, evaluation via interviews, and searching new relevant literature.

The main focus was to develop guidelines for the implementation of a standardized quantitative user experience evaluation tool in the FOK period of the MRI innovation process of the case company. The solution design consists of two main phases: Preparation and Evaluation, which will be discussed in detail in Chapter 4. The first phase of the solution design consisting of the preparation has been completed as a pilot study of the solution design and furthermore provides the case company with the results of the completed steps.

The completed steps of the pilot study included the identification of the main aim, the selection of a user experience evaluation tool, the identification of relevant parameters, the customization of a specific user experience evaluation tool, and the identification of contextual factors that are important for the evaluation for the case company, which are added in the form of labels. The results of these steps will be described in combination with the solution design in Chapter 4.

A user experience evaluation tool was developed and customized with the use of the software Electronic Feedback Management (EFM) Verint, which is a tool that allows to develop and deploy questionnaires and facilitates the collection and analysis of feedback online that is currently used for sending questionnaires within the case company (EFM Verint, 2018). The process of the creation of the solution design was iteratively evaluated, which will be described below.

2.4 Evaluation

To evaluate whether the developed solution design worked, it needed to be subjected to empirical evaluation (Holmström, Ketokivi & Hameri, 2009). Two focus groups were organized and one final validation session in the form of a focus group was organized. As a focus group will give different results than a semi-structured interview it will also have added value for the method triangulation (Longhurst, 2003). The final evaluation can best be categorized as a summative evaluation (Haynes et al., 2004). Together the creation and evaluation phase were used to answer the fifth sub-question, which was previously discussed. The main benefit of a focus group is that it allows for discussion between interviewees that are knowledgeable in their respective areas and gives the interviewer more time to observe the participants (Longhurst, 2003). Therefore, the focus groups were less structured than the formal interviews in the phase of Exploration and synthesis, but they did follow a guide to be more structured, which can be seen in Appendix 10.

The first focus groups focused mostly on the implementation and use of a standardized user experience evaluation tool in the FOK period of the innovation process of the case company. This focus group consisted of two Clinical Application specialists, one Designer, and one employee from Marketing.

The second focus group focused mostly on customizing the content of the user experience evaluation tool to specifically evaluate MRI systems. This focus group consisted of three Clinical Application specialists.

The final validation session had a focus on evaluating the final solution design and therefore implicitly and explicitly evaluating the design principles that guided the creation of the solution design. This final validation session consists of one Clinical Application specialist and the manager of the department of Clinical Application specialists. The next section will describe the data analysis of the main phase of Evaluation.

2.4.1 Data analysis of evaluation

Consistent with the data analysis of the Exploration and synthesis phase, the focus groups were transcribed and coded with the use of the software NVivo (NVivo, 2018). The transcriptions of the focus groups can be seen in Appendix 7. The codes were added to the previously coded results from the main phase of Exploration and synthesis, which resulted in a final list of codes, which can be seen in Appendix 6.

The results of the first two focus groups were mainly used to improve the solution design. The final validation session focused on evaluating the final solution design and reflecting on the previously constructed design principles from theory to understand how appropriate they were to the context of innovating MRI systems within the case company. The reactions of the participants will be exemplified with the use of quotes in this research. The next section will discuss the methodology for ensuring the validity and reliability of this research.

2.5 Validity and reliability

According to Yin (2017), there are four criteria for judging the quality of a research design, each provided with recommended tactics for dealing with these criteria. The four criteria identified by Yin are construct validity, internal validity, external validity, and reliability. Internal validity is only applicable to an explanatory study according to Yin and was therefore excluded from this research. Several tactics were adopted from different researchers to ensure validity and reliability which will be described below (Carter et al., 2014; Golafshani, 2003; Yin, 2013; Yin, 2017).

Yin defined construct validity as ensuring that the correct operational measures are used for the concept being studied. To increase construct validity, method and data source triangulation were used for this study, as described before (Carter et al., 2014; Golafshani, 2003; Yin, 2013). For the method triangulation, semi-structured interviews, focus groups, a literature review, a database of current innovation processes, and informal interviews documented in field notes were used to assess the problem from different points of view. For the data source triangulation, different participants that are working in different departments of the case company were involved to increase the construct validity. Yin suggests using key informants to increase construct validity, which was done in this study via weekly informal interviews with two employees from the case company, which are documented in field notes.

External validity means that the findings of the case company can be generalized to other areas (Yin, 2017). For a case study, analytic generalization is needed to link the results of this research to the initial concepts identified in the literature that guided the research design (Yin, 2017)**.** Analytic generalization means that an abstract level of ideas will be extracted from the results of the case study (Halkier, 2011; Yin, 2013). This research reflected on the design principles from theory and developed new principles to discuss the external validity. Furthermore, this research reflected on the appropriateness of the solution design to other areas within the case company.

Reliability means ensuring that the operations of this study can be repeated with the same results. To ensure reliability, several steps were taken suggested by Yin. The first step is that the background information of the case company was described with the aim of the research for the case company. The second step is that a rationale was given for using the Design Science methodology and developing the research questions. The third step is that the interviews and the focus groups were conducted following an interview guide to give more structure and to be replicable. Additionally, a diverse set of participants was selected to ensure an objective view of the subject.

This chapter described the three main phases of the Design Science methodology that this research followed as well as the validity and reliability of this research. The next section will describe the theoretical results of this research and the design principles that were synthesized for the development of a solution design for the case company in the phase of Exploration and synthesis.

3. Theoretical results

This chapter will discuss the results from the systematic literature review, which was conducted during the Exploration and synthesis phase of this research. With the use of the systematic literature review, the first and second sub-questions were addressed, which were:

"What are important parameters of the user experience and what standardized tools can developers use that can quantitatively evaluate these parameters according to academic literature?"

And:

"How can a standardized user experience evaluation tool be implemented in the innovation process according to academic literature?"

First, different definitions used in academic literature will be explained and the definition that will be adopted in this research will be explained. Second, several theoretical frameworks that define the parameters of general products and medical products specifically will be discussed. Third, available tools for quantitatively measuring the user experience or a subset of the user experience will be discussed. The User Experience Questionnaire (UEQ) will be discussed in detail as this tool was used for the further creation of a design solution for the case company. The theory section concludes with discussing the academic literature on the considerations for the implementation of a user experience evaluation tool in the innovation process.

From the systematic literature review, seven design principles were synthesized, which guided the development of a solution design for the case company (Keskin & Romme, 2020). The seven design principles were categorized based on four main topics from academic literature: (1) Standardized user experience tools, (2) Experience over time, (3) the Context of use, and (4) Benchmarking and communication. Each topic and the corresponding design principles will be discussed below in this order, and they will be summarized at the end of this chapter. Chapter 4 reflects on the design principles based on the empirical evaluation during this research.

3.1 User experience (UX)

There are different definitions for the user experience in current academic literature (Lallemand et al, 2015; Zarour & Alharbi, 2017). This section will discuss the definition of the user experience adopted in this research. The user experience is closely linked to the concept of usability and they are sometimes treated as a synonym (Hassenzahl, 2003). For this reason, the definition of usability adopted in this research will also be explained

User experience and usability are both parts of the higher-level concept of user-centered design. User-centered design places the focus on the user perspective in the design of new products and services instead of the product itself (Lallemand et al, 2015). It is necessary to increase the user-centered approaches in medical-system design to create medical systems that are more user-friendly (Bitkina et al., 2020).

A survey among practitioners and researchers revealed that there is no single accepted definition for the user experience (Lallemand et al, 2015). However, the key definitions in academic literature for the user experience and usability are summarized in Table 3.

Table 3. Key definitions of usability and user experience in academic literature

Authors	Usability	User experience
International Organization for Standardization (ISO) (2019)	"the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use."	"the user's perceptions and responses that result from the use and/or anticipated use of a system, product or service."
Roto, Obrist & Väänänen-Vainio-Mattila (2009)	Usability emphasizes effectiveness and efficiency, which can be measured objectively.	The user experience includes hedonic and pragmatic parameters and is thus subjective. We need to understand how a user feels, which cannot be measured with objective measures.
Brade, Lorenz, Busch, Hammer, Tscheligi & Klimant (2017)	Usability describes the fitness of use of a product and summarizes pragmatic parameters of the product or system. Contrary to user experience, usability does not respect hedonic quality parameters.	User experience includes the holistic assessment of the user because it extends common usability factors with esthetics, joy-of-use and attractiveness
Brooks & Hestnes (2010)	No definition.	The user experience extends beyond traditional measures of user perception, user satisfaction, and usability. User experience includes both objective and subjective measures.
Mahlke (2008)	Usability and utility are distinct concepts part of the instrumental qualities. The usability consists of the efficiency, controllability, helpfulness and learnability.	The user experience consists of the perception of instrumental qualities, emotional user reactions and the perception of non-instrumental qualities.
Hassenzahl (2003)	Usability is a pragmatic attribute. Pragmatic attributes involve the manipulation of the environment which requires functionality (or utility) and ways to access the functionality (usability).	The user experience encompasses all parameters of interacting with a product, which includes pragmatic and hedonic attributes. Hedonic attributes do not focus on functionality but involve stimulation, identification and evocation, which is strongly related to pleasure or psychological well-being .

As can be seen in Table 3 above, there is no clear consensus among researchers regarding the relationship and the differences between usability and user experience (Bitkina et al., 2020). Although there are differences between the definitions, many researchers regard usability and user experience from the viewpoint of the International Organization for Standardization (ISO) (Bitkina et al., 2020). Furthermore, most definitions in academic literature do agree that the user experience is broader than usability and therefore consider usability as a subset of the user experience (Bitkina et al., 2020; Väänänen-Vainio-Mattila, Roto & Hassenzahl, 2008b). For this reason, this research will adopt the definitions of the ISO for usability and user experience and regard usability as a subset of the user experience.

On the contrary to the definition by Roto et al., (2009), the user experience will not be regarded as solely subjective, as objective measures such as physiological and observational tools in the form of video analysis or biometrics sensors can be used to measure (a part of) the user experience (Brooks & Hestnes, 2010; Hussain et al., 2018). Still, this research will only include subjective measures in the form of standardized questionnaires, as they are easy to use, valid, and reliable for evaluating the user experience (Díaz-Oreiro, López, Quesada & Guerrero, 2019).

Hassenzahl (2003) introduced the notion that the user experience consists of pragmatic and hedonic parameters, which other researchers then adopted (Brade et al., 2017; Roto, Obrist & Väänänen-Vainio-Mattila, 2009). Usability is mostly related to pragmatic parameters and the user experience includes both pragmatic and hedonic parameters, which are described in more detail in Table 4**.** This research will regard pragmatic parameters as parameters related to the usability and utility of a system, whereas the hedonic parameters are regarded as parameters related to the looks and feel of a system and the pleasure that it might result in.

Previously the focus was mostly on usability and thus pragmatic parameters for evaluating medical systems (Bitkina et al., 2020). There is a consensus among most researchers that measuring the pragmatic usability on itself is not enough for developers to effectively improve their products and systems (Hassenzahl, 2003). For this reason, this research focused on evaluating the whole user experience of MRI systems. Including measures for the whole user experience will inform the design and lead to better, more satisfying, and more pleasurable products (Hassenzahl, 2003).

To summarize, the definition of usability and user experience will be consistent with the definition of the International Organization for Standardization (ISO) and the usability will be regarded as a subset of the whole user experience. The focus of this research will be on the full subjective user experience of MRI systems, which includes both pragmatic and hedonic parameters. What parameters the full user experience consists of in different contexts and what specific parameters should be measured has gotten insufficient attention in previous research (Zarour & Alharbi, 2017). The next section will discuss key theoretical frameworks that aimed to capture the parameters of the user experience.

3.2 Frameworks of the parameters of the user experience

As discussed in the previous section, there are different definitions of the user experience. Although some definitions already include some parameters that are important for the user experience, they do not provide a comprehensive theoretical framework of all the important parameters. This section will discuss several theoretical frameworks that aimed to identify the important parameters of the user experience. Table 4 below shows the included parameters in key frameworks of the user experience.

Table 4. Key theoretical frameworks in academic literature

Authors	Focus	Included parameters
Zarour & Alharbi (2017)	Framework that covers UX dimensions, aspect categories, aspects and measurement methods	**Brand Experience (BX):** branding, everyday operations, marketing, business communications, context of use, spatio-temporal and user journey. **User experience (UX):** cultural, emotional, hedonic, trustworthiness, esthetics, fun, privacy, sensual, usability, functionality and usefulness. **Technology experience (TX):** platform technology, infrastructure, service response time and visual attractiveness.
Mahlke (2008)	Components of User Experience (CUE) model	**The perception of instrumental qualities:** utility, usability, efficiency, controllability, helpfulness and learnability. **Emotional user reactions:** subjective feelings, motor expressions, physiological reactions, cognitive appraisals and behavioral tendencies.

		Perception of non-instrumental qualities: aesthetic aspects (visual, haptic quality and acoustic quality), symbolic aspects (communicative symbolics and associative symbolics) and motivational aspects.
Hassenzahl (2003)	Key elements of the user experience	**Pragmatic attributes:** Manipulation **Hedonic attributes:** Stimulation, identification and evocation.
Aldoihi & Hammami (2020)	Usability attributes for CT	Context of use, easiness, effectiveness, efficiency, efficient to use, error prevention, functionally correct, helpfulness, image quality, information communicativeness, learnability, operability, productivity, safety, speed of performance, standardization, system functions, system performance, training, useableness, usefulness and user satisfaction
Handayani et al., (2018)	Framework for user acceptance of medical technologies	**Human characteristics:** Individual attitude toward using technology, self-efficacy, perceived usefulness, perceived ease of use, compatibility with the work process, information security expectancy and social influence. **Technology characteristics:** Information quality and system quality. **Organizational characteristics:** management leadership, facilitating conditions, training, participation of end-users in the communication phase, participation of end-users in the design phase and participation of end-users in the implementation phase
Mourouzis, Antona, Boutsakis and Stephanidis (2006)	User experience lifecycle	**Visibility:** The product is made visible to non-users **Perceived usefulness & ease of use:** Non-users are motivated to gain a personal experience of the system **Availability/approachability:** Actual users find it easy and acceptable to reach the product **Quality of interaction experience:** Actual users find it useful, easy and acceptable to interact with the product **Relationship maintainability and subjective usefulness & ease of use:** Previous users are motivated to become long term users **Competitiveness:** Product users are not offered more promising and satisfying alternatives

What becomes clear from Table 4 is that different frameworks include different parameters. For this reason, it remains unclear what the most important parameters are for evaluating the user experience. The common factor in almost all frameworks is that the specific context of use and user characteristics are mentioned as important considerations for evaluating the user experience (Hassenzahl, 2003; Mahlke, 2008; Mourouzis et al., 2006; Zarour & Alharbi, 2017). This suggests that the specific context and user characteristics of the evaluation need to be considered to understand the important parameters in that specific context.

Aldoihi and Hammami (2020) have listed usability parameters of a CT scanner, which is closely linked to the context of MRI systems. However, only pragmatic and no hedonic parameters are included in this list and no guidelines are provided on how to measure these parameters. Therefore, this framework does not provide a comprehensive list of important parameters.

Bitkina, Kim and, Park (2020) depict the state of the art regarding the literature on the context of user experience for medical systems in Figure 5. Medical systems can be regarded as a subset of expert systems, which differ from consumer goods (Bitkina et al., 2020). To what extent consumer goods and expert systems differ in terms of parameters that determine the user experience has not yet been defined, but it seems that the user experience cannot be treated the same due to the differences in the context (Bitkina et al., 2020).

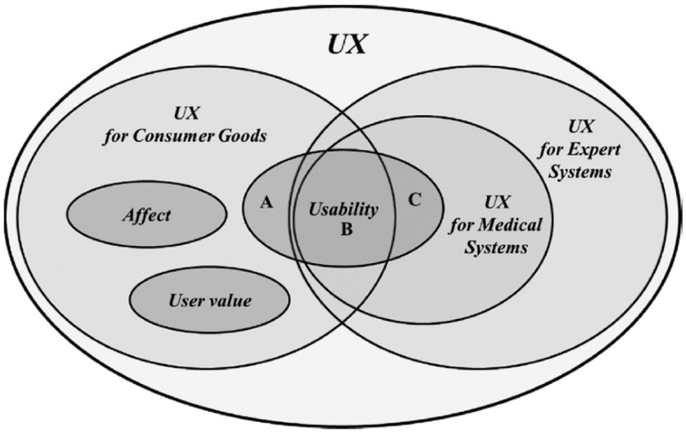


Figure 5. Framework of UX and usability concepts. From: "Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges." By, O. V. Bitkina, H. K. Kim & J. Park, 2020. *International Journal of Industrial Ergonomics, 76,* p.8

The frameworks in Table 4 were not developed with the aim of evaluating the user experience of MRI systems. More research is needed on the specific context of the user experience of medical systems and MRI systems specifically (Bitkina et al., 2020; Zarour & Alharbi, 2017). The user characteristics and specific context of use will be discussed in more detail in Section 3.4.

Furthermore, most parameters that are mentioned in the frameworks are on an abstract level and tools and guidelines still need to be developed for developers to actually evaluate the theoretical parameters of the user experience (Zarour & Alharbi, 2017).

To summarize, different theoretical frameworks of the user experience include different parameters. It seems that a single comprehensive theoretical framework is not possible, as the parameters of the user experience depend on the specific context of use and the user characteristics (Bitkina et al., 2020; Zarour & Alharbi, 2017). The specific context of MRI systems has not been considered yet in relation to a comprehensive theoretical framework of the user experience. The next section will discuss practical evaluation tools that either aim to measure the full user experience or a subset of parameters of the user experience.

3.3 Tools for evaluating the user experience

The previous section discussed theoretical frameworks of the user experience to identify what the important parameters are for evaluating the user experience of MRI systems. The frameworks and parameters are mostly on an abstract level and do not guide developers in measuring the user experience in practice (Zarour & Alharbi, 2017). This section discusses practical tools and questionnaires that can quantitatively evaluate the subjective user experience.

This section focuses on the most used standardized tools discussed in academic literature, as they are easy to use, valid and reliable for evaluating the user experience (Díaz-Oreiro, López, Quesada & Guerrero, 2019). The most recognized standardized tools for quantitatively measuring the user experience are the self-reported measures, usually in the form of a questionnaire with a Likert scale or semantic differentials (Díaz-Oreiro et al., 2019). This section will therefore discuss these self-reported questionnaires.

This research focused on tools that can evaluate the whole user experience, because of the added value compared to measuring merely usability described before (Hassenzahl, 2003). If developers still only want to evaluate the usability of their systems, they can choose a standardized usability tool, which is described in Appendix 9. As discussed before, the user experience of MRI systems is likely to differ from the user experience of other products such as consumer goods. For this reason, the tools from academic literature will be discussed in terms of the appropriateness for evaluating the user experience of MRI systems.

3.3.1 User experience questionnaires

The three most common user experience questionnaires are the Attrakdiff, the User Experience Questionnaire (UEQ), and the Modular evaluation of key components of user experience (meCUE) (Díaz-Oreiro et al., 2019). Only the UEQ will be described here in detail, as the UEQ was used for the further steps in this research. The UEQ was selected for multiple reasons: it allows the calculation of a Key Performance Indicator, it allows for customization of the included parameters, it allows for internal and external benchmarking and it includes the relative importance of each evaluated parameter (Laugwitz, Held & Schrepp, 2008; Schrepp, 2018; Schrepp, Hinderks & Thomaschewski, 2017a; Schrepp & Thomaschewski, 2019). Each reason will be discussed in more detail below.

The Attrakdiff is a user experience evaluation questionnaire developed by Hassenzahl, Burmester, and Koller (2003) that consists of 28 fixed items that focus on the *(1) pragmatic quality, (2) hedonic quality – identity, (3) hedonic quality – stimulation, and (4) attractiveness* of a product or system (Walsh et al., 2014). The modular evaluation of key components of user experience (meCUE) is a questionnaire based on the Components model of User Experience (CUE) developed by Thüring and Mahlke (2007) which was shortly discussed before (Minge, Thüring, Wagner & Kuhr, 2017). The meCUE includes four separate modules: *(1) instrumental and non-instrumental product perceptions, (2) user emotions, (3) consequences of usage, and (4) an overall judgment of the attractiveness* (Minge et al., 2017).

A more detailed description of the Attrakdiff and the meCUE, as well as other related tools, can be seen in Appendix 9. The UEQ will be described in detail first and afterward, the three most standardized tools will be discussed in terms of their appropriateness for evaluating the user experience of MRI systems.

User experience questionnaire (UEQ)

The UEQ was developed as the developers recognized that other user experience evaluation tools put a greater emphasis on the hedonic parameters compared to the pragmatic parameters (Laugwitz et al., 2008). The developers of the UEQ wanted to provide a better balance between the pragmatic and hedonic parameters to get a better overall score of the user experience (Laugwitz et al., 2008).

The tool aimed to evaluate the user experience in a simple and immediate way while covering a comprehensive impression of the product or system (Laugwitz et al., 2008). Three different versions of the UEQ have been created: a traditional version, a short version, and a modular version (Laugwitz et al., 2008; Schrepp et al., 2017a; Schrepp & Thomaschewski, 2019). Each version will be described in this section.

The traditional version of the UEQ is a questionnaire that measures the parameters *Attractiveness Perspicuity, Efficiency, Dependability, Stimulation, and Novelty* (Laugwitz et al., 2008). In total, the questionnaire consists of 26 items which are measured on a 7-step scale whose poles are opposite adjective pairs. The full list of items and the relative parameter they belong to is included in Figure 6 and the full questionnaire is included in Appendix 9. This traditional version of the UEQ can be benchmarked to an online available dataset of product evaluations, which will be discussed in more detail in Section 3.5.

The short version of the UEQ, which is named the UEQ-S, only includes 8 out of the 26 items of the traditional UEQ but aims to measure the same parameters as the traditional version of the UEQ (Schrepp et al., 2017a). As this version does not give the detailed information the traditional UEQ gives, the authors mention that it should only be used when there is not enough time to conduct the full traditional UEQ (Schrepp et al., 2017a). The short version is added to Appendix 9.

The modular version of the UEQ, which is named the UEQ+, allows developers to select the important parameters for their specific product or system (Schrepp & Thomaschewski, 2019). This is consistent with the previous discussion of the importance of the context of use for identifying the important parameters of the user experience (Hassenzahl, 2003; Mahlke, 2008; Mourouzis et al., 2006; Zarour & Alharbi, 2017). The modular UEQ provides developers of systems with the possibility to include other parameters than the six parameters of the traditional UEQ, based on their specific needs (Schrepp & Thomaschewski, 2019).

Currently, for the modular version, the developers of the UEQ have created the parameters *esthetics, Adaptability, Usefulness, Intuitive use, Value, Trustworthiness of Content, Quality of Content, Trust, Haptics, Acoustics, Clarity, Response behavior, Response quality, and Comprehensibility,* next to the traditional six that were described before (Schrepp & Thomaschewski, 2019). Schrepp (2018) has made suggestions for including specific parameters for different product categories. However, no suggestions were made yet for MRI systems or medical systems in general.

UEQ Adjective Pairs	UEQ Factor
annoying / enjoyable	Attractiveness
attractive / unattractive	Attractiveness
friendly / unfriendly	Attractiveness
good / bad	Attractiveness
unlikable / pleasing	Attractiveness
unpleasant / pleasant	Attractiveness
meets expectations / does not meet expectations	Dependability
obstructive / supportive	Dependability
secure / not secure	Dependability
unpredictable / predictable	Dependability
fast / slow	Efficiency
impractical / practical	Efficiency
inefficient / efficient	Efficiency
organized / cluttered	Efficiency
conservative / innovative	Novelty
creative / dull	Novelty
inventive / conventional	Novelty
usual / leading edge	Novelty
clear / confusing	Perspicuity
complicated / easy	Perspicuity
easy to learn / difficult to learn	Perspicuity
not understandable / understandable	Perspicuity
boring / exciting	Stimulation
motivating / demotivating	Stimulation
not interesting / interesting	Stimulation
valuable / inferior	Stimulation

Figure 6. UEQ adjectives and UEQ factors. From: "Generalized User Experience Questionnaire (UEQ-G): A Holistic Tool for Measuring Multimodal User Experiences." By C. S. Boothe, 2020. Mississippi State University p.26

An example for the parameter Efficiency is given in Figure 7. After the evaluation of each parameter of the user experience, the users are asked how important they believe that parameter is on a 7-point scale. The four initial items (e.g., slow / fast) will be referred to as items from now on. The last item to assess the relative importance will be referred to as the importance question from now on.

To achieve my goals, I consider the product as

slow	○	○	○	○	○	○	○	fast
inefficient	○	○	○	○	○	○	○	efficient
impractical	○	○	○	○	○	○	○	practical
cluttered	○	○	○	○	○	○	○	organized

I consider the product property described by these terms as

Completely irrelevant	○	○	○	○	○	○	○	Very important

Figure 7. Parameter Efficiency of the UEQ. From: "Handbook for the modular extension of the User Experience Questionnaire." By M. Schrepp and J. Thomaschewski, 2019, *Mensch & Computer,* p.17

In addition to analyzing each parameter, the UEQ allows calculating an overall UEQ Key Performance indicator (KPI), which combines the scores for each parameter to a single overall score for simple interpretation and communication (Hinderks, Schrepp, Mayo, Escalona & Thomaschewski, 2019). The process of the calculation of the UEQ KPI based on the traditional UEQ will be explained. Based on the importance question of each parameter, a value between -3 and +3 is obtained for each of the parameters: *attractiveness* a_i *, perspicuity* p_i *, efficiency* e_i *, dependability* d_i *, stimulation* s_i *, and novelty* n_i (Hinderks, Schrepp, Mayo, Cuaresma & Thomaschewski, 2018).

The overall UEQ KPI is calculated with the use of the relative importance for each parameter per participant with the following formula:

$$UEQ\ KPI = \frac{1}{n}\sum_{i=1}^{n}(Aw_i * A_i + Pw_i * P_i + Ew_i * E_i + Dw_i * D_i + Sw_i * S_i + Nw_i * N_i)$$

Figure 8. Formula to calculate the relative importance of attractiveness. From: "UEQ KPI Value Range based on the UEQ Benchmark" By A. Hinderks, M. Schrepp, F.J.D. Mayo, M.J.E. Cuaresma and J. Thomaschewski, 2018, *University of Seville, Tech. Rep.,* p.3

The process described here is for the traditional UEQ but the KPI can also be calculated for the modular UEQ+ (Schrepp & Thomaschewski, 2019). The UEQ KPI has been validated in a more pragmatic and a more hedonic context and both contexts seemed to be appropriate for the UEQ KPI (Blunck, 2020; Hinderks et al., 2019). The developers of the tool mention that further research is needed to assess whether the UEQ KPI can be implemented in an organization as a KPI and that the emphasis should be on interpretability and acceptance (Hinderks et al., 2019).

The UEQ was identified as the most appropriate user experience tool for evaluating MRI systems. The theoretical part of identifying the most appropriate user experience tool, which involves a discussion of the meCUE and the Attrakdiff for MRI systems, will be shown in the next section. All three standardized user experience evaluation tools have been used in the medical context before, but not for the evaluation of MRI systems specifically. The previous use of the three standardized user experience tools in the medical context is discussed in detail in Appendix 9.

3.3.2 Appropriateness of user experience tools for MRI systems

As mentioned before, the UEQ was chosen to be the most appropriate for the context of MRI systems, both based on the theoretical and empirical results of this research. This section will explain why the UEQ seems to be the most appropriate out of the three main standardized user experience evaluation tools for the context of MRI systems based on the theoretical results.

The meCUE is a modular questionnaire, which allows developers to only select the relevant parameters for their evaluation (Minge et al., 2017). However, when looking at the items in the meCUE questionnaire they do not seem to fit the context of MRI systems (Lallemand & Koenig, 2017). Items such as *"I could not live without this product"* and *"The product is like a friend to me"* can confuse participants and make them feel that they are wasting their time (Lallemand & Koenig, 2017). The meCUE can still be used in the context of MRI systems, but developers might choose to adopt a questionnaire that better fits the context of MRI systems.

The Attrakdiff has successfully been applied to many contexts, including the medical context (Shih & Zheng, 2020). Again, the focus of this questionnaire does not seem to be completely appropriate for MRI systems, as the questionnaire includes predominantly hedonic parameters, and pragmatic parameters are limited (Hassenzahl, 2004).

In response to the Attrakdiff questionnaire, the UEQ was developed to have a better balance between the hedonic and pragmatic parameters, which seems to better fit the context of MRI systems (Laugwitz et al., 2008). Additionally, the modular UEQ version allows developers to choose the relevant parameters specifically for the context of MRI systems (Schrepp & Thomaschewski, 2019). As there is a lack of research on the important parameters of the user experience of MRI systems the modularity will have added value. Furthermore, the UEQ tool allows to analyze the chosen parameters on the importance with the importance questions mentioned before (Bitkina et al., 2020). Based on the importance questions, developers can over time exclude irrelevant parameters for operators from their user experience evaluation. This importance question also allows to calculate an overall KPI score, which facilitates the interpretation and communication of the results (Hinderks et al., 2019). The importance question in the UEQ tool will ensure that parameters that are not important for operators, will not weigh much in the overall UEQ KPI score (Hinderks et al., 2018).

In the meCUE and Attrakdiff, there is no such importance question to include the subjective importance of each parameter for operators. For example, in the Attrakdiff questionnaire, the hedonic parameters weigh more than the pragmatic parameters in an overall score as more items focus on hedonic parameters (Hassenzahl et al., 2003).

On the contrary to the meCUE and the Attrakdiff, the UEQ has the benefit of externally benchmarking the parameters from the traditional UEQ with an online available database **(**Schrepp et al., 2017a)**.** This external benchmarking can facilitate the interpretation and communication which will be discussed in more detail in Section 3.5.1.

Another option for developers of MRI systems is to develop their own questionnaire that fits the specific context. This would not only require more effort but the tool will also not be statistically validated and there will be no benchmark data other than the data gathered by the developers.

Concluding, the UEQ seems to be the best option for measuring the user experience in the context of MRI systems as it allows for the calculation of a Key Performance Indicator, allows for customization of the included parameters, allows for external benchmarking and it includes the relative importance of each parameter (Laugwitz et al., 2008; Schrepp, 2018; Schrepp et al., 2017a; Schrepp & Thomaschewski, 2019). This research will assess the appropriateness and usefulness of the UEQ in the MRI innovation process of the case company.

From the theoretical analysis currently, the UEQ tool seemed the most appropriate for MRI systems. As new user experience tools could be developed in the future, the design principle that was synthesized will be on a more general level. It is best to use a standardized tool for measuring the user experience as using a standardized questionnaire does not require any effort and time for the construction and has previously been shown to be reliable (Díaz-Oreiro, López, Quesada & Guerrero, 2019). This resulted in the following design principle for the creation of a solution design:

Design Principle 01. *If an organization wants to integrate user experience evaluations within their innovation process of medical systems (C), it should use a standardized user experience evaluation tool (A) to quickly evaluate the medical systems with a reliable tool (M) which leads to an accurate quantitative evaluation of the user experience of medical systems that can be tracked over time (O).*

Furthermore, balancing the hedonic and pragmatic parameters based on the importance of the included parameters for operators is important in the context of MRI systems, which resulted in the following design principle:

Design Principle 02. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should include the relative importance of pragmatic and hedonic* parameters *for the operators (A) to balance the importance of hedonic and pragmatic parameters ensure that hedonic* parameters *do not disproportionally represent the evaluation of a product (M) which leads to a more accurate user experience evaluation of medical systems (O).*

As discussed before, the context of the user experience evaluation is important. The next section will discuss the context of evaluating the user experience in more detail.

3.4 Context of evaluating the user experience

In the previous section, practical tools for quantitatively evaluating the user experience of MRI systems were discussed. As mentioned before, the specific context of the evaluation is important for determining the user experience. This section will discuss important considerations for the context of user experience evaluations discussed in academic literature. This section will focus on the user characteristics and the specific context of the use of user experience evaluations. Section 3.5 will discuss important considerations for implementing user experience evaluation tools in the innovation process of companies, with a focus on the interpretation and communication of the results.

The user experience of a user with a system is not static as it depends on the context and evolves over time (Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos & Sinnelä, 2011). Two important considerations are the user characteristics and the specific context of use (Mahlke, 2008). This section will discuss both these considerations.

3.4.1 User characteristics

User characteristics have been shown to influence the outcome of a user experience evaluation (Mahlke, 2008). There are several different user characteristics, such as age, gender, and the internal state of the user. The internal state of the user consists of predispositions, expectations, needs, motivation, mood, and other factors (Hassenzahl & Tractinsky, 2006). The influence of the prior experience of users will also be discussed in this section.

There are differences in quality perceptions of users, as users rate products differently on the same parameters but also rate the importance of the same parameters differently (Mahlke, 2008). For some users the visual esthetics of a system are important and for others, it not important (Bloch, Brunel & Arnold, 2003). Cultural differences can also impact the ratings for both esthetics and usability of a system and therefore also impact the user experience (Kortum & Oswald, 2018; Tractinsky, 1997).

The UEQ was evaluated with Indonesian and German students, which rated the same products differently and also mentioned different parameters to be more important as can be seen in Figure 9 (Santoso, Schrepp, Hinderks & Thomaschewski, 2017). This effect could however be caused by the language of the interface of the evaluated system (Santoso et al., 2017). The advantage of the UEQ is that users with different characteristics can be analyzed separately from each other of which an example can be seen in Figure 9.

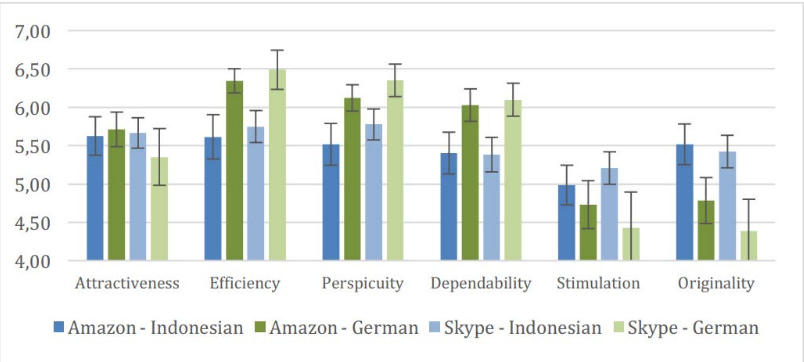


Figure 9. Judgments concerning the importance of the UEQ aspects. From: "Cultural differences in the perception of user experience." By H. Santoso, M. Schrepp, A. Hinderks and J. Thomaschewski, 2017, *Mensch und Computer,* p.269

Von Wilamowitz-Moellendorff, Hassenzahl and Platz (2006) found that the more pragmatic qualities (utility and usability) seem to be stable in their importance or become slightly more important over time and the hedonic qualities (beauty, identity, stimulation) seem to become less important over time, which can be seen in Figure 10. Especially the usability parameters increase but also are more important when a user has more experience with a product or system (Von Wilamowitz-Moellendorff et al., 2006). The hedonic or parameters become less important after the increased experience with a product or system (Borsci et al., 2015; Von Wilamowitz-Moellendorff et al., 2006).

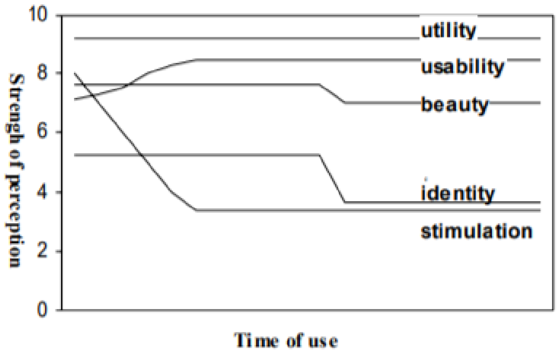


Figure 10. UX qualities over time. From: "Dynamics of user experience: How the perceived quality of mobile phones changes over time" By M. von Wilamowitz-Moellendorff, M. Hassenzahl and A. Platz, 2006. *User Experience-Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer interaction,* p.77

Several researchers have tested this effect of prior experience of users on the outcomes of different user experience questionnaires and have shown that the prior experience should be considered in the evaluation as it can have a great impact on the scores (Berkman & Karahoca, 2016; Borsci et al., 2015; McLellan, Muddimer and Peres, 2012; Walsh et a., 2014). Developers could benefit from measuring the user experience several times to see how the user experience develops over time independent of product iterations (Kujala et al., 2011). The research on the temporal dimension of the user experience resulted in the following design principle:

Design principle 03. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should repeat user experience evaluations over time (A) to measure the dynamic changes in the experience of users with a product (M) which leads to a better understanding of how the experience of operators with medical products develops over time (O).*

As described before, next to the user characteristics and their experience with a product, the specific context of use during the evaluation is also important to consider for developers. This will be discussed in the next section.

3.4.2 Context of use

Next to the user characteristics, the context and the environment in which the product or system is used also influence the user experience evaluation and should therefore be considered in the user experience evaluation (Hassenzahl & Tractinsky, 2006). The context of use is a broad concept, but this section will discuss the context or reason of using a system next to the physical and social context of a system (Wigelius & Väätäjä, 2009).

Marc Hassenzahl (2003) mentioned that the usage mode can be important when measuring the user experience. Hassenzahl argues that usage of a product always consists of behavioral goals and actions to fulfill these goals. When in goal mode, the goal determines all the actions and when in action mode, the user is exploring the product, and the goals are determined "on the fly" (Hassenzahl, 2003). An example Hassenzahl gives is that when in goal mode effectiveness and efficiency become more important than when in action mode.

The spatial context which consists of conditions such as temperature, noise, and light affect the user experience of users with a system (Wigelius & Väätäjä, 2009). The social context also influences the user experience, which can be important for the user experience of medical systems as the interaction with a patient might have a big impact on the user experience (Wigelius & Väätäjä, 2009).

If developers want to evaluate the user experience of their products, they should specify the context of use to understand the user experience for that specific context. When evaluating a product, it seems that giving the user a goal to fulfill will put the user in goal mode and will give different results than allowing the user to explore the product (Hassenzahl, 2003). This resulted in constructing the following design principle:

Design Principle 04. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should specify the specific context of use for user experience evaluations (A) to analyze the effects of influencing factors on the user experience outcomes (M) which leads to a more accurate understanding of the experience of operators with medical systems (O).*

Next to specifying the specific context of use of user experience evaluations, there are other important considerations for implementing a user experience tool in the innovation process. This will be discussed in the next section.

3.5 Implementation of a user experience tool in the innovation process

Where the previous section focused on factors that influence the user experience, this section will focus on the interpretation and communication of the results of the user experience tools in the innovation process of developers of MRI systems. Benchmarking will be discussed, followed by visualization of the results.

3.5.1 Benchmarking

Developers can interpret each of the parameters of the user experience independently to see how well they are doing on that specific parameter. However, a score can be difficult to interpret, and benchmarking is usually needed to be able to draw meaningful conclusions from the user experience evaluation (Sauro & Lewis, 2016). Benchmarking in this research means comparing a user experience evaluation score to a previous user experience evaluation score to see whether it improves or not. There are two different methods for benchmarking, which will be referred to here as external and internal benchmarking. Developers can use external benchmark data available from standardized user experience questionnaires (Sauro & Lewis, 2016). Alternatively, developers can build their own internal benchmark data by evaluating the user experience multiple times (Schrepp & Thomaschewski, 2019).

Two UEQ was partially chosen as a user experience evaluation tool in this research to evaluate MRI systems as it has an online database of previous evaluations (Schrepp et al., 2017a).

One additional advantage of the online benchmark dataset is that it does not only show how well you do in comparison to other products, but it can also show what the real value range is in practice (Schrepp, 2015). In Figure 11, classifications are made for the UEQ scores that show how well you do relative to the benchmark database in terms of the UEQ score (Schrepp, 2015). The UEQ has a possible range from +3 to -3, but the results are not evenly spread across the range. It is likely that you get a score above 0 but this does not necessarily mean that you are doing well compared to other products (Schrepp, 2015). The products are classified into 5 different categories for the UEQ to allow for easier interpretation:

- Excellent: In the range of the 10% best results.
- Good: 10% of the results in the benchmark data set are better and 75% of the results are worse.
- Above average: 25% of the results in the benchmark are better than the result for the evaluated product, 50% of the results are worse.
- Below average: 50% of the results in the benchmark are better than the result for the evaluated product, 25% of the results are worse.
- Bad: In the range of the 25% worst results.

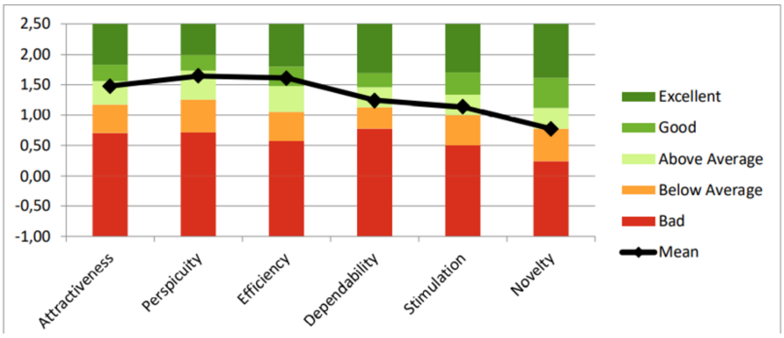


Figure 11. Benchmark for the parameters of the traditional UEQ. From: "User experience questionnaire handbook". By M. Schrepp, 2015, P.6

The UEQ KPI that can be calculated from the UEQ scores also has a possible range of -3 to +3 (Hinderks et al., 2019). In practice, the real value range for the UEQ KPI is from -0.286 to 2.143 according to the current benchmark data, which shows again that very low numbers do not occur (Hinderks, Schrepp, Mayo, Cuaresma & Thomaschewski, 2018). The scores without benchmarking can easily be misinterpreted which shows the importance of benchmarking. Ideally, developers would compare the user experience scores to similar products in the database, but this is not possible for MRI systems currently (Schrepp et al., 2014).

A disadvantage when using external benchmarking data to interpret the user experience scores is that developers are limited to the parameters included in the standardized questionnaires and cannot change them to fit their own purposes (Schrepp & Thomaschewski, 2019). So, when a standardized questionnaire does not fit the context and developers want to incorporate relevant parameters for their purpose, they can choose to build their own internal benchmarking data.

Developers can for example score their current product and a new product iteration to assess whether there is a statistically significant improvement of the user experience in the new product iteration (Schrepp, 2015; Schrepp et al., 2017a). A two-sample t-test can be used to assess whether the average difference between two groups is significant or if it is due to random chance (Schrepp, 2015). Alternatively, developers can compare the user experience of different product designs with each other to assess which of the design options will result in the highest user experience of which an example can be seen in Figure 12 (Schrepp et al., 2014). Another option is to track the user experience over time to observe how it develops either independent of changes or iterations or with changes to the system or product (Schrepp & Thomaschewski, 2019).

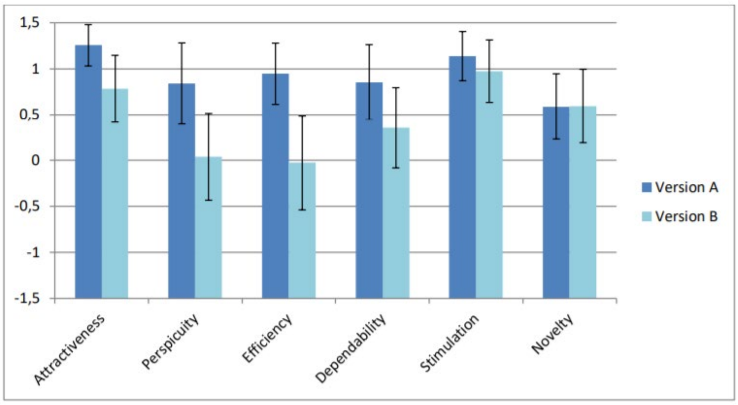


Figure 12. Comparison of two products with the traditional UEQ. From: "Applying the user experience questionnaire (UEQ) in different evaluation scenarios." By M. Schrepp, A. Hinderks, J. Thomaschewski, 2014, *International Conference of Design, User Experience, and Usability,* p.386

As both internal and external benchmarking have different advantages and limitations, two design principles were constructed:

Design principle 05. *If an organization wants to integrate user experience evaluations within their innovation process of medical systems (C), it should use internal benchmarking (A) to facilitate the evaluation and interpretation of the user experience of their systems with the use of a customized tool (M) which leads to an accurate understanding of the experience of operators with medical products that can be tracked over time and product iterations (O).*

Design principle 06. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should use external benchmarking (A) to facilitate the interpretation of the user experience evaluation of their systems through comparing with an online database (M) to understand how the user experience of the medical systems is relative to other products (O).*

Next to benchmarking to facilitate the interpretation and communication of the user experience scores, visualizations can be used which will be described in the next section.

3.5.2 Visualizations

As the previous section showed, user experience scores such as the UEQ can easily be misinterpreted (Sauro & Lewis, 2016; Schrepp, 2015). Developers should consider the interpretation and communication of their user experience evaluations preferably before starting the actual evaluations. One option identified in previous research to facilitate the interpretation and communication of a user experience evaluation is through visualization (Lachner, Naegelein, Kowalski, Spann & Butz, 2016). Figures 11 and 12 in the previous section, showed two options to visualize the results of the UEQ.

One example of a tool to visualize the scores from a user experience evaluation is the quantified UX evaluation tool (QUX) developed by Lachner et al. (2016). They describe that the user experience is a fuzzy concept and visualizations could help to simplify the interpretation of a user experience score outcome. The identified benefits of the QUX visualization are that it provides an overview, helps to prioritize, allows for benchmarking, and facilitates communication (Lachner et al., 2016).

As visualizations can facilitate the interpretation and communication, the following design principle was constructed:

Design principle 07. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should use visualizations (A) to simplify the interpretation and communication of the user experience of their systems (M) which leads to a better understanding and communication of the experience of operators with medical products (O).*

The systematic literature review resulted in identifying two research gaps and seven design principles for the creation of a design solution for this research. Below the two research gaps and the seven design principles described throughout the theoretical results, will be summarized.

3.6 Summary theoretical results

This section will summarize the main insights from the systematic literature review. The two research gaps that were identified and that this research focused on will be repeated and the design principles will be summarized that were used for the creation of a solution design for the case company. Furthermore, the systematic literature review provided an answer to the first two sub-questions of this research, which were:

"What are important parameters of the user experience and what standardized tools can developers use that can quantitatively evaluate these parameters according to academic literature?"

"How can a standardized user experience evaluation tool be implemented in the innovation process according to academic literature?"

The answers to these sub-questions will be summarized in Chapter 5.

The two research gaps that were identified that this research will address are:

Research gap 1. *The important parameters of the user experience of MRI systems are not defined*

Research gap 2. *Guidelines for the implementation of a standardized user experience evaluation tool in the innovation process of MRI systems are missing*

The design principles that were identified during the systematic literature review are summarized in Table 5 below. The design principles were constructed around four main topics of previous academic research: (1) Standardized user experience tools *(design principle 01 and 02)*, (2) Experience over time *(design principle 03)*, (3) the Context of use *(design principle 04)* and (4) Benchmarking and communication *(design principle 05, 06 and 07)*. The next section will discuss the empirical results of this research.

Table 5. Design principles from theory

Context	Action	Mechanism	Outcome
Design Principle 01. *If an organization wants to integrate user experience evaluations within their innovation process of medical systems (C),*	***it should use a standardized user experience evaluation tool (A)***	*to quickly evaluate the medical systems with a reliable tool (M)*	*which leads to an accurate quantitative evaluation of the user experience of medical systems that can be tracked over time (O).*
Design Principle 02. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C),*	***it should include the relative importance of pragmatic and hedonic parameters for the operators (A)***	*to balance the importance of hedonic and pragmatic parameters ensure that hedonic parameters do not disproportionally represent the evaluation of a product (M)*	*which leads to a more accurate user experience evaluation of medical systems (O).*
Design principle 03. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C),*	***it should repeat user experience evaluations over time (A)***	*to measure the dynamic changes in the experience of users with a product (M)*	*which leads to a better understanding of how the experience of operators with medical products develops over time (O).*
Design Principle 04. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C),*	***it should specify the specific context of use for user experience evaluations (A)***	*to analyze the effects of influencing factors on the user experience outcomes (M)*	*which leads to a more accurate understanding of the experience of operators with medical systems (O).*
Design principle 05. *If an organization wants to integrate user experience evaluations within their innovation process of medical systems (C),*	***it should use internal benchmarking (A)***	*to facilitate the evaluation and interpretation of the user experience of their systems with the use of a customized tool (M)*	*which leads to an accurate understanding of the experience of operators with medical products that can be tracked over time and product iterations (O).*
Design principle 06. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C),*	***it should use external benchmarking (A)***	*to facilitate the interpretation of the user experience evaluation of their systems through comparing with an online database (M)*	*to understand how the user experience of the medical systems is relative to other products (O).*
Design principle 07. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C),*	***it should use visualizations (A)***	*to simplify the interpretation and communication of the user experience of their systems (M)*	*which leads to a better understanding and communication of the experience of operators with medical products (O).*

4. Empirical results

The previous chapter discussed the results of the systematic literature review. The theoretical insights in the form of design principles were combined with the empirical analysis to create a design solution for the case company. This chapter will discuss the empirical results of this research of the Exploration and synthesis phase, the Creation phase, and the Evaluation phase. First, the current innovation process of the case company will be described which was identified during the phase of Exploration and synthesis and answered Sub-question 3:

"How does the case company currently innovate their MRI systems and how do they integrate the perspective of operators in the innovation process?

Second, the design requirements for the case company in a user experience evaluation tool will be discussed which were identified during the phase of Exploration and synthesis and answered Sub-question 4:

"What are the design requirements for the implementation of a standardized user experience evaluation tool in the innovation process of MRI systems for the case company?".

Third, the steps for the creation of a solution design and the final solution design will be described, which took place during the second main phase of Creation. Last, the final validation of the solution design will be described which took place during the last main phase of Evaluation, which together with the final solution provided an answer to sub-question 5:

"How to implement a standardized user experience evaluation tool to measure the experience of operators of Magnetic Resonance Imaging (MRI) systems in the First of a Kind (FOK) period of the MRI innovation process of the case company? ".

4.1 Current innovation process

This section will discuss the current innovation process of the case company, which answered Sub-question 3. The current innovation process was analyzed to understand how a solution can be implemented in the innovation process of the case company. The case company has two innovation processes for MRI systems which will be discussed here, with a focus on the operator (representative) involvement.

The two processes were identified by conducting semi-structured interviews and informal meetings with the case company employees and analyzing documents in an online database. First, the Advanced Development (AD) process will be discussed, followed by the Planning, Development, Launch, and Maintenance (PDLM) process. Both processes were customized for the innovation of MRI systems within the case company. The steps of the processes are the same for software and physical MRI system innovations. This section concludes with a discussion of the focus of this research on the First of a Kind (FOK) period of the innovation process of the case company.

4.1.1 Advanced development (AD)

The first step in terms of innovation is the Advanced Development (AD) process. The case company has a regular and a so-called 'lite' AD process. The lite version is used for small projects and the regular version for larger projects. The regular version will be discussed here, as it includes all the steps that are also in the lite version.

The AD process is mostly concerned with market research and identifying opportunities and solutions. The AD process is a very iterative process that starts with a project Charter that explains all the details and budgets of the project. The AD starts with high uncertainty, and concepts and designs are increasingly formalized throughout the AD steps and the PDLM process that follows. AD consists of five main steps, which can be seen in Figure 13.

The first step *"Propose AD project"* consists of the input collection and the decision to start or not start the AD process. This step includes filling in the project Charter that explains all the details and requirements of the project.

The second step *"Define AD project"* consists of defining the project objectives and scope, understanding the customer relevancy and business opportunity, generating concept options, and building a project plan for evaluating these options.

The third step *"Select technology or concepts"* evaluates the previously generated concept options and reduces the number of concept options. This is the most iterative step, where many options are evaluated, and only promising options go on to the next step.

The fourth step *"Prove concepts"* proves the technical feasibility of the concept options in the appropriate environment. This can be a simulated environment, laboratory environment, or a real application environment.

The final step *"Develop function"* refines the concept and proves that it works in the relevant environment and that it satisfies the agreed user and business requirements. This is mostly concerned with eliminating major risks. Compared to the previous step, this step is more focused on the operators than on technical feasibility.

In terms of the involvement of operators in the AD process, concepts and concept directions are evaluated with actual operators and clinical application specialists, with either prototypes or demonstrations. The main aim of the AD process is to learn fast or fail cheap in the innovation process. Although the process does not seem to be very iterative, it is cheap for the case company if the AD steps are not successfully concluded, as no large investments have been made. For this reason, they can easily initiate a new project. When the AD process is successfully concluded it moves on to a PDLM process, which is described below.

<Removed for confidentiality>

Figure 13. Advanced development (AD) process

4.1.2 Planning Development Launch and Maintenance (PDLM)

If the AD process is successfully concluded, the project is initiated, and it becomes a ‘real’ project. This means that resources are committed, and the PDLM process steps will be conducted until the product is finished and commercially launched on the market.

The process follows a stage-gate process, which involves 10 gates where higher management is involved and decides on a “Go”, “Recycle”, “Hold” or “Kill” of the project. The PDLM process is mostly like a waterfall process, where each step is completed before moving on to the next step, which makes it less iterative than the AD process. The PDLM process can be seen in Figure 14. The process will be separated into four main phases: *(1) Developing needs and finalizing requirements, (2) Creating designs and integrating product, (3) Verifying and validating designs, and (4) Launching and monitoring product.*

<Removed for confidentiality>

Figure 14. Planning Development Launch and Maintenance (PDLM)

4.1.2.1 Developing needs and finalizing requirements

The first step *“Develop user and business needs and concepts”* consists of understanding the business and user needs of all relevant stakeholders, including hospitals and suppliers. The concepts are created to meet these needs and are checked whether these are in line with the value proposition, which has some overlap with the AD process, but this is done to further de-risk investments in the next steps of the PDLM process.

The second step *“Finalize high level requirements, designs and plans”* consists of translating the user and business needs into requirements for the product but also for the service, manufacturing, service operations, supply chain, and marketing that all depend on the specific product that will be created. The high-level designs are finalized, and a plan and activities are defined to develop and launch the new product.

In these early steps of the PDLM, the main involvement of operators is that their current practices in hospitals are being analyzed and observed. The needs for improvements in practice are also identified by analyzing complaints and enhancement requests. The clinical application specialists have the main role of identifying and translating these complaints and enhancement requests into user needs.

Mostly internal clinical application specialists are included in the evaluation of product concepts. This is consistent with the suggestion by Roto et al., (2009) to involve expert users in the early stages of the innovation process as they are better at recognizing future needs.

In these steps, no final product concept has been chosen yet and decisions still have to be made for selecting the best product concept or idea. Low-fidelity prototypes are sometimes made which can be evaluated. In terms of the experience of operators, these early steps are the most difficult phase to evaluate, as there are often no functional prototypes yet. Currently, the concepts are evaluated by application specialists in these early steps by expressing their feelings, but there is a need to quantify these feelings of the expected user experience for actual operators.

Also, a usability engineering plan is developed. In this plan, it is assessed whether the product needs to be evaluated with operators in the upcoming steps of the PDLM. It can be that many so-called formative and summative usability tests are needed to evaluate the product, or it can be that operators do not have to be involved at all, and anything in between. Formative usability tests have a focus on improving the usability and summative usability tests have a focus on evaluating what has been accomplished in terms of usability. In these usability tests, usability engineers conduct evaluations of the product with actual operators. Usually, operators are asked to perform a specific task with a product, which will be observed, and the operator will evaluate the product. For the next steps of the PDLM, there will be assumed that the actual operator involvement is needed according to the usability engineering plan, as this research focuses on operator involvement.

4.1.2.2 Creating designs and integrating product

The third step *"Detail designs"* consists of creating detailed designs for the product, service, manufacturing, operations, and the supply chain. Also, the launch dates of the product are committed.

The fourth step *"Integrate product"* consists of realizing manufacturing equipment and service delivery means and performing engineering runs. The parts and elements of the innovation are verified, and the elements are integrated into a full MRI system.

In these steps, the main concept is chosen, and designs are made and eventually integrated in a full product. These are the first steps where a functional product can be evaluated with operators.

First, separate components are evaluated and finally, a full product is integrated and evaluated. Often integrating the elements in a full product, results in identifying areas of improvement that were not identified before. This is an important step where the experience of the interrelating items in a full product can be evaluated. These steps consist of iteratively improving the product, which means that evaluations are repeated to track the impact on the usability and experience of operators over the iterations.

The focus of these steps starts very broad with a focus on the whole experience. With the use of the previously described formative usability testing, the components and products are improved. Usability engineers, which have the main responsibility of improving the usability, prefer to do the formative usability testing with actual operators, but they are sometimes done with internal application specialists if only minor improvements are tested. When the end of the integrate product step is reached, the focus is mostly on safety and complying with regulations, which require several mandatory formative and summative usability testing, depending on the country.

Recently, "Testathons" are included in the integration step, which involves performing workflows freely and identifying how the systems respond. The "Testathons" are conducted with internal application specialists, who also fill in the System Usability Scale (SUS) after using the system to evaluate the subjective usability.

4.1.2.3 Verifying and validating designs

The fifth step *"Verify designs"* consists of performing a manufacturing pilot run and service verification run. Verification is concerned with checking whether the integrated product is developed in the right manner and meets the requirements.

The sixth step *"Validate and release designs"* consists of validating the product, obtaining final regulatory approval, and releasing the supply chain, which are prepared in the previous steps. In this step, the evaluations are mostly concerned with safety and risks. Validation is concerned with evaluating whether the product meets the needs that it should address.

Verify designs are more technical evaluations and validate designs are more focused on usability and the experience of operators. These steps are mostly internal evaluations with internal application specialists, but it sometimes still includes usability testing with actual operators.

These steps are more summative in terms of assessing whether it satisfies prespecified goals and needs. Ideally, the product does not need any improvements and the evaluations, and the improvements that follow are more safety related.

4.1.2.4 Launching and monitoring product

The seventh step *"Prepare launch"* consists of completing the market preparation for delivery and starting up the production of the product.

The eighth step *"Monitor launch"* consists of monitoring the product quality, volume, manufacturing, and delivery. It starts with a First of a Kind (FOK) period, where the product is released at some sites to easily monitor the product and ensure a successful introduction of the product release into the markets. The distribution of the FOK customers is discussed with all involved parties and is based on requirements such as division across markets, type of customers, and types of configuration of the current product. A product support engineer will support the FOK installation on-site or remote support will be given. Reported issues are registered and communicated to the relevant teams for investigation.

For MRI innovations, approximately 20 hospital sites receive an initial version of the product and evaluate the product in the actual context. At each site, approximately 3 operators will evaluate the product. This is the first phase where the actual context of the product is considered, which is important for accurately evaluating the user experience (Väänänen-Vainio-Mattila et al., 2008a).

The feedback from the FOK sites is used as input for the full release of the product. Currently, the System Usability Scale (SUS) is used in the FOK period after 1 week, 4 weeks, and 12 weeks. The score is repeated to assess how the subjective usability develops over time.

After the FOK period, the products are continuously monitored, and feedback is often integrated into future projects. As the product is out already, no major changes will be made to the product.

Marketing is involved in the FOK period to retrieve proof points of the product in terms of quantitative data or quotes. Marketing wants to have information from the FOK sites to be able to communicate this with the decision-making units for eventually selling and marketing the final product. For this reason, marketing is mostly regarded with retrieving positive results.

The focus of this research will be on the FOK period as a pilot test, to determine how a user experience evaluation tool can be integrated in the innovation process. This phase was identified as the most important phase for the case company to do evaluations of the experience of operators, as this will help the case company in identifying the most important areas for improvement. Furthermore, if the experience evaluations in hospitals are conducted, they can function as a baseline for benchmarking future product improvements. There is also a need to evaluate the user experience in other steps of the innovation process, but they are currently not the main focus for the case company.

These sections discussed what the innovation process of MRI systems for the case company looks like, with a focus on how the operator (representatives) are involved. This provided an answer to Sub-question 3:

"How does the case company currently innovate and how do they integrate the perspective of operators in the innovation process?".

The insights on the current innovation process were used to create a solution design for the case company medical systems, which will be discussed in the next sections.

4.2 Design requirements for the case company

Next to the identification of the current innovation process of the case company, the design requirements were identified which together with the design principles guided the creation of a solution design. The design requirements were based on the semi-structured interviews and informal meetings with employees of the case company from several departments conducted during the main phase of Exploration and synthesis during this research. Five employees from the department Clinical Applications, two from the department Design, one from the department Quality & Regulatory and one from Marketing were interviewed about their needs and requirements in a tool for evaluating the user experience of operators of MRI systems. The design requirements were divided into four functional requirements, two user requirements, two boundary conditions, and one design restriction, which were explained in the methodology section (Van Aken, Berends & van der Bij, 2012). The full list of requirements can be seen in Table 6.

Table 6. Design requirements for the case company

Type of requirement	Requirement
Functional requirement 01:	The tool should be able to measure the user experience of operators of MRI systems.
Functional requirement 02:	The tool should allow for internal and external benchmarking.
Functional requirement 03:	The tool should help to increase the focus on the experience of operators of MRI systems in the innovation process.
Functional requirement 04:	The tool should allow the evaluation of the experience of both software and physical MRI system innovations.
User requirement 01:	The tool should be easy to understand by operators.
User requirement 02:	The results of the tool should be easy to analyze and communicate within the company.
Boundary condition 01:	The tool should be able to be sent online.
Boundary condition 02:	The tool should be compatible with the current innovation process.
Design restriction 01:	The tool should not take longer than 5 to 7 minutes to complete for operators.

Most requirements are still on a high level and are self-explanatory. Insights from theory were discussed during the informal meetings and semi-structured interviews, which resulted in some overlap between the design principles and the requirements. This overlap already indicates that some design principles are useful in practice.

The functional requirements were used as the starting point of this research. For functional requirement 02 there were some conflicting opinions. While two clinical application specialists mentioned that external benchmarking is needed to communicate the results, one usability designer mentioned that he did not want to compare the user experience of MRI systems to completely different products. Eventually, it was decided to still include this requirement as other participants explicitly mentioned that this external comparison is needed to successfully communicate the results of the UEQ. However, in terms of marketing the communication of the internal benchmark to customers can give some issues, as one employee from Marketing mentioned:

"You don't want to cannibalize an existing product of your own. You also want to be careful how you word that, so that you don't cannibalize some other product saying it is 70% less good." (Interview Marketing, 2021).

For the design restriction in Table 6 there were some conflicting opinions. Three clinical application specialists argued that operators will both have enough time and are willing to evaluate the experience of a product for longer than 10 minutes. Two employees from the design department argued that they want to have enough time for other types of evaluations and that the tool might need to be shorter than 5 minutes. The decision to restrict the time to a maximum of 5 to 7 minutes is based on a restriction from the clinical application manager, who argued:

"Anything not longer than 5 to 7 minutes will be fine for operators (in the field) " (Informal meeting 5, 2021).

For the other requirements, there were no conflicting opinions. This section provided an answer to Sub-question 4:

"What are the design requirements for the implementation of a standardized user experience evaluation tool in the innovation process of MRI systems for the case company?".

All the requirements that were described here, were combined with the design principles from the theoretical analysis to create a solution design for the case company. The next section will describe the process and results of the creation of a solution design for the case company.

4.3 Solution design

The theoretical analysis resulted in design principles and the empirical analysis resulted in design requirements, which were both explained before. Based on the design principles and design requirements a solution design was created for the case company. As discussed in the methodology, the focus of the solution design was on guidelines to implement a user experience evaluation tool in the innovation process of MRI systems. The creation of the solution design included focus group sessions, semi-structured interviews, and informal meetings. The focus groups involved the use of PowerPoint slides and a MIRO board where visualizations could be made, to facilitate the discussions. The slides and MIRO board screenshots are included in Appendix 10. The process of creating the solution design was described in more detail in the methodology section.

The final solution design in terms of process steps can be seen in Figure 15. This section will explain each of the process steps of the final solution design. The solution design consists of two main phases: Preparation and Evaluation. First, the Preparation phase will be explained, which needs to be done before the evaluation of the user experience can start. The steps of the preparation for user experience evaluations have been conducted during this research for MRI systems for the case company, for which the main results will be summarized at each step below.

The process steps for the phase of Preparation are: *Identifying the main aim, Selecting an appropriate tool, Identifying relevant parameters, Customization, and Identifying the contextual labels.* Each step of the preparation and the results for the case company will be discussed below.

After discussing the Preparation phase, the Evaluation phase will be explained, which consists of the process steps needed for the evaluation of the user experience in practice. The steps of evaluation have not been conducted during this research, but the usefulness of this process was validated during a final validation session. The steps of evaluation consist of *Labelling, Conducting the evaluation, Analyzing the results, and Communicating the results.*

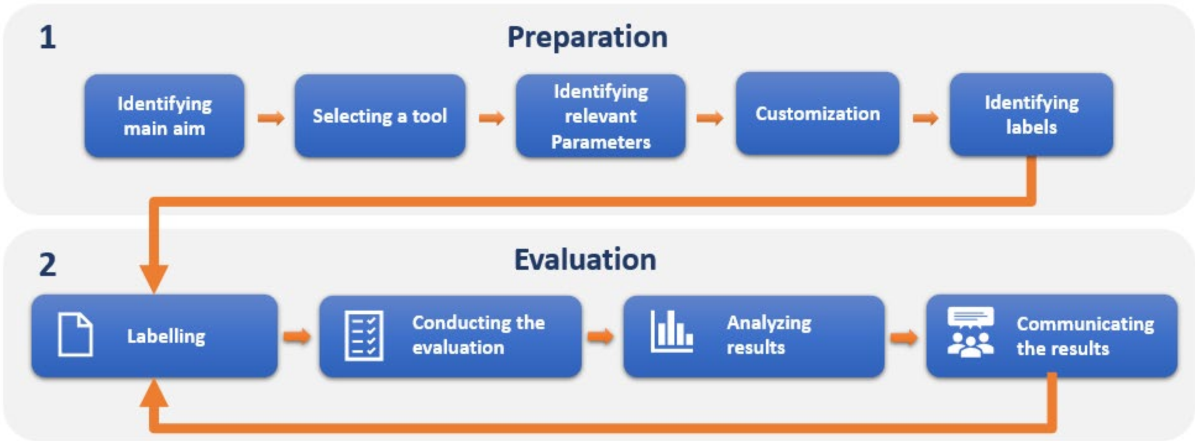


Figure 15. Final solution design

4.3.1 Preparation

The first phase of the solution design consists of the preparation phase. The preparation phase consists of 5 steps: *identifying the main aim, selection of tool, identifying relevant parameters, customization, and identifying labels.* Each step will be described with a focus on the FOK period of the innovation process of MRI systems of the case company. The preparation phase needs to be completed once to prepare for the evaluation of the user experience of MRI systems. After completing the steps of the preparation phase, the evaluation phase can be repeatedly used for different user experience evaluations without repeating the steps for the phase of Preparation.

Identifying main aim

The first step of implementing a user experience evaluation tool in the innovation process is the identification of the main aim of the evaluations. This is an important step that has to be done first. as the following steps depend on the decisions that are taken in this step. This section will discuss some important considerations for identifying the main aim and the main aim that was identified for the case company. Three important considerations are summarized in Table 7.

The first important consideration is whether developers want to do a quick evaluation to get a general understanding or do a comprehensive evaluation to be as accurate as possible. When only a general understanding is required, developers can choose to use a standardized tool from the literature that has been validated, without spending too much time on customizing it. If developers want to have a more comprehensive understanding, it is better to include the most relevant parameters for measuring the user experience and to customize the questionnaire.

The second important consideration is whether the aim is to derive a single overall user experience score or that specific parameters of the user experience need to be evaluated (e.g., efficiency). If an overall score is the main aim, developers should select the parameters that have (or are expected to have) the most impact on the overall experience. If specific parameters need to be evaluated, they can choose to only select the parameters that they want to evaluate.

The last important consideration is whether a product is evaluated only once or that evaluations are repeated over time. If a product is evaluated only once, developers can best choose a tool that allows for external benchmarking to an available database of evaluations without customizing it, as a score without benchmarking has not much meaning (Sauro & Lewis, 2016). If more evaluations will be done, the scores can be benchmarked to previous internal evaluations, and developers can decide to also compare scores to the external benchmark dataset (Schrepp & Thomaschewski, 2019).

The main aim for the case company will be described here, of which the results will be explained in the next steps. The case company wanted to focus on developing a comprehensive understanding of the user experience and wanted to have an overall score to track the impact of MRI innovations on the experience of operators over time. They want to create a benchmark dataset, which can be shown on a user experience dashboard to easily see which areas of the MRI system need to be improved. The case company wanted to use both internal and external benchmarking, which requires that both a standard version of a tool with an available benchmark dataset is included, as well as to include specific parameters that are important in the context of MRI systems. This means that the evaluation will be relatively long for operators and this can only be done if operators are willing to spend time and effort to evaluate many different parameters.

It is important that in this first step different departments are involved in a discussion of the main aim of the user experience evaluation. Different departments can have different aims for measuring the user experience of their products, which can result in making different decisions in the following steps. Some tools, such as the UEQ described before, do have the advantage that they can fulfill different aims, without too much extra effort.

Table 7. Important considerations

Option A:	Option B:
Quick evaluation: use a standardized tool without spending too much time on customization	**Comprehensive:** spend time on selecting the most important parameters and customizing the questionnaire
Overall score: include the parameters that have (or are expected to have) the most impact on the overall experience of operators	**Specific parameters:** include the specific parameters you want to measure (the importance question can be left out of the questionnaire)
Evaluate once: use a standardized tool without customization as this allows for external benchmarking. The score on itself will not have much meaning.	**Repeated evaluations:** customize the questionnaire to the specific needs as the scores can be internally benchmarked to previous evaluation. This can be combined with an external benchmark.

Selecting a user experience evaluation tool

The second step of the implementation of a user experience tool in the innovation process is the selection of an appropriate user experience tool. Three different standardized user experience evaluation tools were identified with the systematic literature review in current academic literature, which are the meCUE, the Attrakdiff, and the UEQ (Díaz-Oreiro et al., 2019). Selecting a standardized user experience tool has the advantage that the tool has previously been validated, which ensures that the tool is reliable and easy to use (Díaz-Oreiro et al., 2019). In this second step, it is important that developers consider the context of their product or system and select a tool that is appropriate for that context.

As discussed in the theoretical results section of this research, the UEQ tool seems to be the most appropriate tool for the context of MRI systems. From the semi-structured interviews that were conducted in the main phase of Exploration and synthesis during this research, the appropriateness of the UEQ tool for measuring MRI systems for the case company was evaluated. The empirical analysis showed indeed that the UEQ tool would be appropriate for evaluating MRI systems.

The main reasons for selecting the UEQ were that it allows the calculation of a Key Performance Indicator, it allows for customization of the included parameters, it allows for internal and external benchmarking and it includes the relative importance of each evaluated parameter, which were described in detail in the theoretical results of this research (Laugwitz et al., 2008; Schrepp, 2018; Schrepp et al., 2017a; Schrepp & Thomaschewski, 2019).

One clinical application specialist mentioned:

"I think to measure the overall impression, this (the UEQ) is the best way. This can also easily be repeated over time, as you are always asking the same questions." (Clinical Application specialist 4, 2021).

Identifying relevant parameters

The third step of the implementation of a user experience evaluation tool consists of identifying the relevant parameters of the user experience in the specific context of the evaluations. This step is only needed if in the previous two steps it was identified that selection of the most important parameters is needed. Furthermore, the UEQ is the only user experience evaluation tool that allows to select relevant parameters, which makes this step redundant if a different user experience evaluation tool is selected.

If a product is only evaluated once or if a quick evaluation of the user experience is needed, the traditional UEQ can be used as it allows for external benchmarking to an online database and does not have to be customized. In these cases, this step of identifying the important parameters, will not be needed.

If the aim is to be comprehensive in the evaluation, the most important parameters of the user experience in that specific context still need to be identified. The modular UEQ offers a list of 20 different parameters that developers can choose from, which was listed in the theoretical results section. However, it is probably better to first have an open discussion with experts to make sure that relevant parameters in the specific context are identified.

If there are any relevant parameters identified during this discussion that are not included in the modular UEQ, they can either be constructed, which requires a lot of effort or they can be excluded from the analysis. If developers want to include a new parameter and want to construct this, they can repeat the steps taken in the research of Klein, Hinderks, Schrepp, and Thomaschewski (2020).

After deciding whether there are any missing parameters, the 20 parameters from the modular UEQ can be analyzed with experts to identify what the most important parameters are in that specific context.

For The case company, two parameters were identified that are not explicitly available in the modular UEQ tool. which were “Performance” and “Errorless use”. The case company employees were interested in these parameters as they are often referred to for internal communication within the case company. One clinical application specialist mentioned:

“We have a KPI from development how stable the system is, and also a performance measurement … So, it would be interesting if we could connect a subjective score from customers to this, but we could probably use efficiency for this purpose.” (Interview Clinical application specialist 3, 2021).

One designer also mentioned that “Efficiency”, which is available in the modular UEQ, could be a good replacement for the internally communicated parameter “Performance”. The clinical application manager mentioned:

“It is fine if we measure this implicitly, as we are mostly interested in the overall score, so the main purpose is not to say something specifically about the performance” (Interview Clinical application manager, 2021).

Errorless use is also not available in the modular UEQ. However, this was identified to not be problematic as there are enough objective measures within the innovation process of the case company to observe whether operators can use the system without major errors. Also, it is expected that the parameters “Trust”, “Efficiency” and “Dependability” will score lower if the system is not errorless, as they show some overlap. The case company is mainly interested in the overall score of the user experience and this missing parameter will have an impact on several other parameters, which suggests that it will not have a significant impact on the overall score if it is not measured explicitly.

The manager of the department of Clinical applications mentioned:

“I think we should use these terms from literature and change the definitions we use within our company.” (Interview Clinical application manager, 2021)

Based on the input from the semi-structured interviews, the relevant parameters of the UEQ for evaluating the user experience of MRI systems were selected during two focus groups.

The parameters from the modular UEQ that were excluded are “Value”, “Quality of Content”, “Acoustics”, “Response behavior”, “Response quality” and “Comprehensibility”. The parameters “Response behavior”, “Response quality” and “Comprehensibility” were specifically made for voice assistants (Klein, Hinderks, Schrepp & Thomaschewski, 2020). Although MRI systems do have a voice system, it was identified that not many innovations will include a voice system, which makes including the parameter redundant for most innovations. For this reason, these parameters were excluded from the questionnaire. The parameter “Acoustics” was also excluded from the questionnaire as most innovations do not involve changes to the acoustics, which again makes the parameter redundant in most cases as the scores for this parameter will be the same between different evaluations.

The parameter “Value” was excluded as it is referred to as personal value and being proud of possessing the product, which does not apply to MRI systems as they are not personally used and are unlikely to have personal value (Schrepp, & Thomaschewski, 2019). The parameter “Quality of content” was excluded as it showed similarity to “Trustworthiness of content” and the items that the parameter consists of were less appropriate than the items of “Trustworthiness of content” for MRI systems.

In total 14 different parameters of the modular UEQ were identified as interesting parameters for the evaluation of the user experience of operators of MRI systems. There was no consensus yet on the number of parameters to include in the questionnaires. One participant from the department Design mentioned:

“It is long so I don’t know whether we will be able to get that much feedback.” (Interview Designer 1, 2021).

The clinical application specialists that were interviewed did have a completely different view as they wanted to include all the relevant parameters. Different employees mentioned that it would be best to start the evaluations with all relevant parameters and remove parameters over time if they are not important or take too much time for participants based on analyzing the importance question of the UEQ, which was described in the theoretical results section.

The case company wanted to be able to evaluate software innovations as well as physical product innovations of MRI systems, which was one of the design requirements listed before. For MRI systems, different parameters are important for innovations that are software or that are physical products. For this reason, the case company wanted to have different questionnaires for software and physical products. Innovations that are a combination of these two can probably best use the software version, as the “Trustworthiness of the content” that is included in the software version is likely to be more important than the parameter “Haptics” that is included in the questionnaire for physical products. The two lists of included parameters in the questionnaires for the physical products and software products, as can be seen in Table 8. So, when evaluating an innovation, developers should either use the software or physical product questionnaire, depending on the type of innovation. This will be described in more detail in the step of identifying labels.

As mentioned before, in the first step the main aim needs to be aligned between different departments. If the main aim is not clear, different departments might want to select different parameters for their own purpose. For the case company, Design focuses on gathering the most negative points of the product, to further improve the product. Marketing on the contrary wants to retrieve positive results of very specific parameters, to market the product. The UEQ has the added value of fulfilling different purposes, as the final list of parameters includes relevant parameters for Design, Marketing, and for higher management.

Table 8. Specific parameters included for physical innovations and software innovations for MRI systems

Physical MRI system innovations:	Software MRI system innovations:
Pragmatic parameters	
Adaptability	Adaptability
Efficiency	Efficiency
Learnability	Learnability
Dependability	Dependability
Usefulness	Usefulness
Clarity	Clarity
Intuitive use	Intuitive use
Trust	*Trust (privacy)*
	Trustworthiness of content
Hedonic parameters	
Attractiveness	Attractiveness
Visual esthetics	Visual esthetics
Fun-of-use	Fun-of-use
Novelty	Novelty
Haptics	

Customization

The fourth step of the implementation of a user experience evaluation tool in the innovation process consists of the customization of the tool. As mentioned before, if only external benchmarking is needed, it is best to not change anything or limit the customizations of the tool as the comparison with the online database will be less reliable. However, if only internal benchmarking will be used, it is best to customize the parameters of the UEQ to make sure that the questionnaire is easily understood by users.

The names of the parameters or the way that the questions are asked can be customized. It is best to not change the four items that each parameter consists of in the UEQ, as changing these items will decrease the reliability of the questionnaire. The items should only be changed if they are really problematic and are not understood by users.

The names that are given to the parameters of the UEQ can be adjusted to fit the specific context to make it easier to interpret and communicate the results (Schrepp & Thomaschewski, 2019). Previously it was shown that being asked to evaluate parameters that do not fit the context can result in frustration, therefore it is important that wording is understood by users of the tool (Lallemand & Koenig, 2017). All customization changes for MRI systems for the case company can be seen in Table 9. The full customized questionnaires for both physical and software MRI system innovations can be seen in Appendix 11.

On the contrary to the traditional UEQ, the parameter names of the tool are shown to operators in the customized questionnaire to facilitate the understanding of what they are evaluating. This does mean that the names of the parameters should be clear to operators. The parameter names were discussed with clinical applications specialists, who are representatives of the operators and understand how operators work in practice.

Especially the parameter name “Perspicuity” was not understood by almost all of the participants. As this parameter is closely linked to learnability, the parameter name was changed to “Learnability”.

The parameter “Stimulation” was also not clear to all participants, the alternative name suggested by Schrepp and Thomaschewski (2019) was used, which is “Fun-of-use”. The parameter “Personalization” was changed to “Adaptability” as an MRI system is not personalized to the operator but adapted to the characteristics of a patient. The term “Adaptability” seems to be more appropriate in this context. “Adaptability” was also one of the alternative names suggested by Schrepp and Thomaschewski (2019).

As discussed before, the conclusion was drawn that different parameters should be included in a questionnaire for physical and software MRI innovations. For physical products, the parameter “Trust” focuses on the trust in the whole system. For software products, the parameter “Trust” focuses on the trust in the privacy and data handling of the system.

Each parameter has an introductory sentence to show participants what they are evaluating. The introductory sentences were customized to fit the context of MRI systems**,** which can be seen in Table 9. In the traditional UEQ, the importance question for asking the relative importance of the specific parameter is: *“I consider the product property described by these terms as:”,* which operators have to score from completely relevant to very important. As this was not clear enough and it was expected that it would confuse participants, the statement was changed to: *“For me ‘parameter name’ of the product is:”*

In the tool, some specific items raised some doubts about whether they would fit the context of MRI systems. For the parameter “Usefulness”, one of the items that operators have to evaluate is whether the product is rewarding or not rewarding, which two clinical application specialists thought would be confusing. These specific items of each parameter were not changed as it is likely that would have a significant impact on the reliability of the results if specific items are changed. Participants are therefore asked to answer the items with their first impression although the words might not completely fit the context.

Table 9. Customization changes to the questionnaire

Category	Previous	New
Parameter name:	Perspicuity	Learnability
Parameter name:	Stimulation	Fun-of-use
Parameter name:	Personalization	Adaptability
Parameter name (only physical innovations):	Trustworthiness of content	Trust
All importance questions:	“I consider the product property described by these terms as:”	“For me *‘Parameter name’* of the product is:”
Introductory sentence for Adaptability:	“Regarding my personal requirements and preferences, the product is:”	“Regarding patient characteristics and preferences, the product is:”
Introductory sentence for Usefulness:	“I consider the possibility of using the product as:”	“I consider the product as:”
Introductory sentence for Clarity:	“In my opinion the user interface of the product looks:”	“In my opinion, the product looks:”
Introductory sentence for Trust (only physical innovations):	“In my opinion the information and data provided by the product are:”	“In my opinion, the product is:”
Introductory sentence for Trust (privacy) (only software innovations):	“Regarding the use of my personal information and data, the product is: “	“Regarding the use of patient information and data, the product is:”

Identifying labels

The fifth step of the implementation of a user experience evaluation tool in the innovation process consists of identifying the contextual factors for the evaluations. From the systematic literature review, it became evident that it is important that the specific context of user experience evaluations is clearly defined (Hassenzahl & Tractinsky, 2006; Mahlke, 2008). This step is based on design principle 4 described before. The empirical analysis also showed that defining the specific context of the evaluation facilitates the analysis of the user experience scores.

Although the context was important for identifying the relevant parameters of the user experience and the customization of the UEQ tool, developers cannot create a new tool for every specific context. During the empirical analysis, a participant suggested that labelling can be used to specify the user experience evaluations for easier analysis.

Labels provide information on what the specific context of the evaluation was and what contextual factors could have influenced the user experience scores. As an example, if the prior professional work experience of operators is labelled, developers can analyze the differences in scores between operators with different levels of experience. This step consists of identifying all the possible labels that are needed for future user experience evaluations to facilitate the analysis.

The full set of labels that were identified for the context of MRI systems for the case company can be seen in Table 10. Each of the labels will be described here shortly. The labels were partially derived from academic literature and partially from the empirical analysis. These labels will not be shown to participants but will be constructed to facilitate the further analysis of the results. Most labels can be added by the developers that conduct the evaluation. Some specifics about the context need to be asked to the participants, such as the prior professional work experience of operators and the region they live in.

The first label is the type of product, which makes a distinction between software and physical MRI system innovations. Each type has a separate questionnaire for MRI systems as described before.

The second label is product fidelity. A distinction will be made between different types of fidelities. It will probably not make much sense to compare the user experience score of an idea to a fully integrated product, as they cannot be used in the same way. There will be distinguished between, ideas, concepts, prototypes, integrated products, and installed base. An integrated product means that the new software or physical system innovation is integrated into a fully functional MRI system. Installed base, means that the system is currently being used in practice.

The third label consists of the specific workflow step(s). For this label, the workflow steps of operators that will be evaluated can be labelled or developers can fill in that they evaluate the full experience of the product. Workflow steps are all the steps of scanning a patient from turning on the MRI system to interpreting the resulting scans of the MRI scan, which are shown in more detail in Appendix 12.

The fourth label is the innovation phase. The categories for the innovation phase are consistent with the phases described in Section 4.1. This allows to compare user experience scores over the different innovation steps and therefore over time and product iterations. Although the focus of this research is on FOK, which is in the monitor launch phase, this label was added as the aim is eventually to also evaluate the user experience in other phases of the innovation process.

The fifth label is the region. The operators are asked about the region they live in. The regions in Table 10 are the regions in which the case company distinguishes operators.

The sixth label is the prior professional work experience. Operators are asked about the years of professional experience they have with MRI systems. The categories are based on discussions with clinical application specialists, as these categories are expected to show the largest differences in user experience scores.

The seventh label is the user type, which can be a technologist, lead technologist, or radiologist. These different users were described in Section 1.1.

The eighth label is the user proxy. A distinction will be made between actual operators and the clinical application specialists, who are proxies of the operators. This is a different category than the seventh label as the clinical application specialists can take the role of each of the three user types.

The ninth label is the specific innovation. This label shows the specific name of the innovation that was evaluated. This allows to easily distinguish between evaluations of different innovations.

The tenth label is the usage mode, which is consistent with the suggestion by Hassenzahl (2003) to distinguish between usage modes as it can impact the results of a user experience evaluation. The evaluation can be guided, where users get a specific task to fulfill. Alternatively, the evaluation can be non-guided, where users are free to use the system as they like.

There are many more factors that can influence the user experience evaluations, which could be labelled. Factors such as the actual environment in terms of noise, temperature, but also social factors such as the characteristics of the patient or colleagues can all influence the user experience scores. No other labels were included as it would result in too many labels. The included labels were the labels that were deemed important and relatively easy to measure in the context of MRI systems. In other contexts, it is likely that other labels will be constructed.

Table 10. Included contextual labels for MRI systems

Label	Categories
Type of product	Software / Physical
Product fidelity	Idea / Concept / Prototype / Integrated product / Installed base
Innovation phase	AD / Develop user and business needs and concepts / Finalize high level requirements, designs and plans / Detail designs / Integrate product / Verify designs / Validate and release designs / Prepare launch / Monitor launch
Region	Africa / ASEAN + other pacific / BENELUX / Central + East Europe / DACH (Germany, Austria, Switzerland) / France / Greater China / Iberia / Indian subcontinent / Italy, Israel & Greece / Japan / Latin America / Middle east & Turkey / Nordics / North America / Russia, Ukraine & Central Asia / UK & Ireland
Prior professional experience	Less than 1 year / 1-3 years / 3-8 years / more than 8 years
User type	Radiologist / Lead-technologist / Technologist
User proxy	User / Application specialist
Solution name	Solution name
Usage mode	Guided / Non-guided use

4.3.2 Evaluation

When the phase of Preparation is concluded, the phase of Evaluation can start. The evaluation phase describes the process from starting the user experience evaluation until concluding the evaluation. The steps of the evaluation phase can be repeated if an evaluation round is concluded. The evaluation phase consists of *Labelling, Conducting the evaluation, Analyzing the results, and Communicating the results*, which will each be described below. These steps for the evaluation were identified through semi-structured interviews, informal meetings, and focus groups. On the contrary to the previous phase of Preparation, the evaluation steps have not yet been conducted for MRI systems within the case company.

Labelling

In the last step of the preparation phase, all the labels were identified that are needed for the analysis of the user experience scores. In the first step of the phase of Evaluation, the specific categories of the labels need to be selected and added to the evaluation. This labelling needs to be done before actually sending the questionnaire. This will help with understanding the specific characteristics of the evaluation. The specific category of the labels needs to be selected by the person that is responsible for the evaluations.

For the previously described labels for the case company, two labels can only be identified by asking the operators. These are the labels for the region and the prior professional work experience. The questions that operators need to answer to identify the categories of these labels are added to the UEQ questionnaire, which can be seen in Appendix 11. As discussed before, different questionnaires were made for software and physical MRI system innovations. Based on the labelling of the type of product, either the questionnaire for software or physical MRI systems needs to be sent to operators. Sending the questionnaires will be discussed in the next step.

Conducting the evaluation

In the second step of the evaluation, the customized questionnaires will be used to evaluate the user experience of operators of MRI systems. Within the previous step, it was established whether a software or physical MRI system innovation will be evaluated. The questionnaire for either the software products or physical MRI system innovations needs to be sent to users. Two important considerations for sending the questionnaires are the use of online tools for sending the questionnaires and repeating the user experience evaluations over time.

One of the requirements for the case company was that the tool can be sent online to easily evaluate the user experience of MRI systems in different geographical locations. The case company eventually wants to send the questionnaires to hospital sites at different geographical locations, both nationally and internationally.

Two versions of the UEQ were created in the Electronic Feedback Management (EFM) Verint tool, which is an online tool for sending and collecting feedback, which is currently being used for sending questionnaires within the case company (EFM Verint, 2018). This online tool allows to include a specific set of operators and set specific dates for sending the questionnaires. The tool can also automatically send reminders when operators do not fill in the questionnaire. Using such a tool can be useful as companies do not have to spend too much time conducting the evaluation.

Furthermore, the results from the questionnaire can easily be exported for further analysis and communication. An example is given of one parameter of the online questionnaire in Figure 16 and the full questionnaires are included in Appendix 11.

Currently, only an English version is created, but the website of the modular UEQ offers 26 other languages (“UEQ+”, 2021). The case company has a translation agency that is able to translate the questionnaire to other languages if needed. A validation round might be needed to verify whether the translated questionnaires are clearly understood by operators.

Figure 16. Example parameter Efficiency of online version of the tool

As discussed before, user experience evaluations must be repeated over tim, as the user experience is not a static phenomenon (Kujala et al., 2011). Within the FOK period of the innovation process of the case company, the questionnaire should be repeated after 1 week, 4 weeks, and 12 weeks of using the product. This will allow to see how the user experience develops over time. The reason for selecting these time intervals is that the expectation of the case company employees is that operators are relatively positive in the first week as they are usually positive about innovations. After 4 weeks operators usually recognize more issues and are relatively negative. After 12 weeks usually, all issues are fixed, and the score is likely to be more stable. These expectations are based on the experiences of the case company employees with previous MRI system innovations. These expectations can be verified by using the UEQ with these time intervals.

During the user experience evaluations, it is best to keep the context as consistent as possible. This will limit the impact of influencing factors on the evaluation of the user experience scores. Furthermore, developers should specify to users what they are evaluating whether it is the full experience or only specific workflow steps, which should be added to the introductory text of the questionnaire.

The more data that is collected, the better and more stable the results will be. There is no minimum number of participants you need to get reliable results, as it also depends on the level of agreement of the participants (Schrepp, 2015). Previously, for the original UEQ around 20 participants already gave stable results (Schrepp, 2015). Usually, in the FOK period of the MRI innovation process of the case company a minimum of around 30 different operators are involved, so aggregating these results, is likely to give reliable results. The next section will discuss the analysis of the results.

Analyzing the results

An online questionnaire tool has the advantage that the results from the user experience evaluations can easily be exported for analysis. The results can be exported to a prepared Excel file, where the analysis does not require a lot of effort, which was also one of the requirements for the case company. Screenshots of the excel files can be seen in Appendix 11. The Modular UEQ already provides an Excel file for quick analysis, which was adjusted for the case company to allow for the analysis of all the parameters that were selected from the modular UEQ. If developers want to do other types of data analysis, they have to create their own Excel file for their purpose. Statistical analyses such as two-sample t-tests can be done to observe whether there is a statistically significant difference between the averages of user experience scores or whether it is due to random chance (Schrepp, 2015).

As mentioned before, the questionnaire might be relatively long for operators, so if from the analysis can be concluded that some parameters are not very relevant and the evaluation takes too long, some parameters can be removed from the questionnaire and the analysis in future evaluations.

Communicating the results

After analyzing the results, the results still need to be communicated. Communication is essential for the success of the tool within the case company. If results cannot be clearly communicated, the evaluations will not have much added value (Lachner et al., 2016). The theoretical analysis showed that two methods to communicate the results of the user experience evaluations are visualizations and benchmarking, both were considered to be useful by the case company employees (Lachner et al., 2016; Sauro & Lewis, 2016). This section will discuss different methods of communicating and visualizing the results. The visualizations that are made here are created with simulated data, so the results do not represent the actual evaluation of MRI systems. The visualizations that are shown here, were identified as useful for the communication by Clinical Application specialists of the case company.

The main aim for the case company was to derive a single overall user experience score that can easily be communicated, preferably in the form of a dashboard. As was described in the theoretical results section, an overall UEQ KPI can be calculated. To be able to communicate and interpret the UEQ KPI score, either internal or external benchmarking is needed (Sauro & Lewis, 2016).

For the external benchmark, only the original six parameters of the traditional UEQ can be used, which was explained in the theoretical results section. For the internal benchmark, the customized parameters can be used from the modular UEQ. As the case company wants to do both internal and external benchmarking, communication of both internal and external benchmark scores will be discussed here.

With the external benchmark, an adjective rating can be added to the scores, which facilitates the interpretation of the scores (Schrepp et al., 2017a). With the use of the real value ranges in a database of the UEQ KPI, which was described in the theoretical results, Figure 17 was created. This score is especially useful if no internal benchmark is available and this adjective rating is likely to be easier to communicate with higher management (Hinderks et al., 2018). For example, it is easier to communicate that the score is "Good" compared to other products, instead of communicating that the user experience score is 1.5.

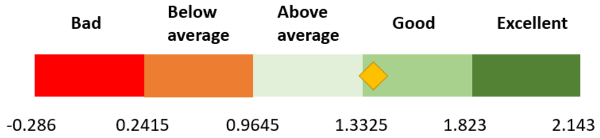


Figure 17. External benchmark range for the UEQ KPI

The six parameters of the traditional UEQ can also be externally benchmarked individually, which was shown in Figure 11 in the theoretical results section. The scores for each parameter can be categorized in the same way as the UEQ KPI in Figure 17.

If evaluations of the user experience are repeated, they can be visualized to understand how the user experience develops over time. In Figure 18 an example is shown how the user experience scores can be visualized for the three different evaluations within the FOK period of MRI innovation process for operators with different levels of prior professional work experience.

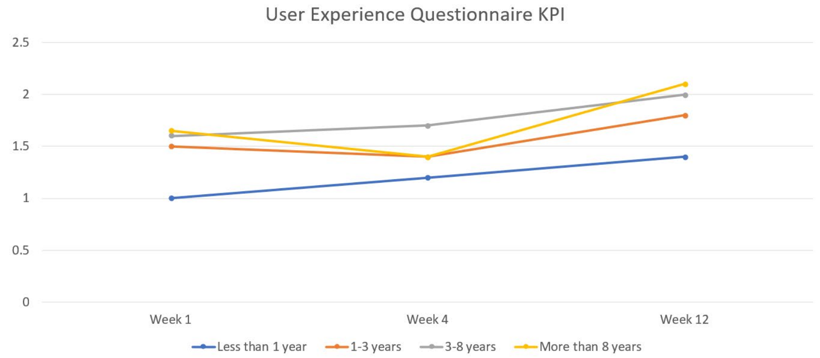


Figure 18. User experience scores over time

The importance question, discussed in the theoretical results section, can be analyzed separately to see whether there are any parameters that are either very important or not important at all. If there are parameters that are not important, they can be excluded from the questionnaire. If there are parameters that are very important, developers can decide to focus more on these parameters with their innovations. Figure 19 shows how the relative importance of each parameter can be visualized.

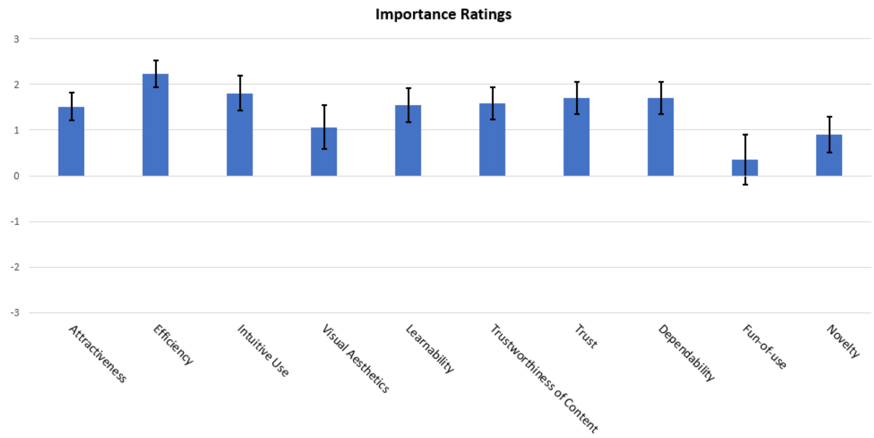


Figure 19. Importance scores for different parameters

Table 11 and figure 20 show how the user experience scores can be used on a dashboard. Eventually, the aim is to have both a baseline evaluation of the user experience, as well as an evaluation of the new MRI system innovation. With these two scores, the difference can be calculated. As mentioned before, developers can use statistical analyses such as two-sample t-tests, to understand whether there is a statistically significant difference between the scores.

Table 11. Dashboard for external and internal user experience KPI score

Score	Baseline	Innovation	Difference
External UEQ KPI	1.55 (Good)	1.62 (Good)	+ 0.07
Internal UEQ KPI	1.62	1.80	+ 0.18

Figure 20 shows a more detailed version of how the user experience scores can be used on a dashboard and at the top left of Figure 20, the labels are shown that were selected in the previous steps. In Figure 20 a user experience score is given to all to different workflow steps of scanning a

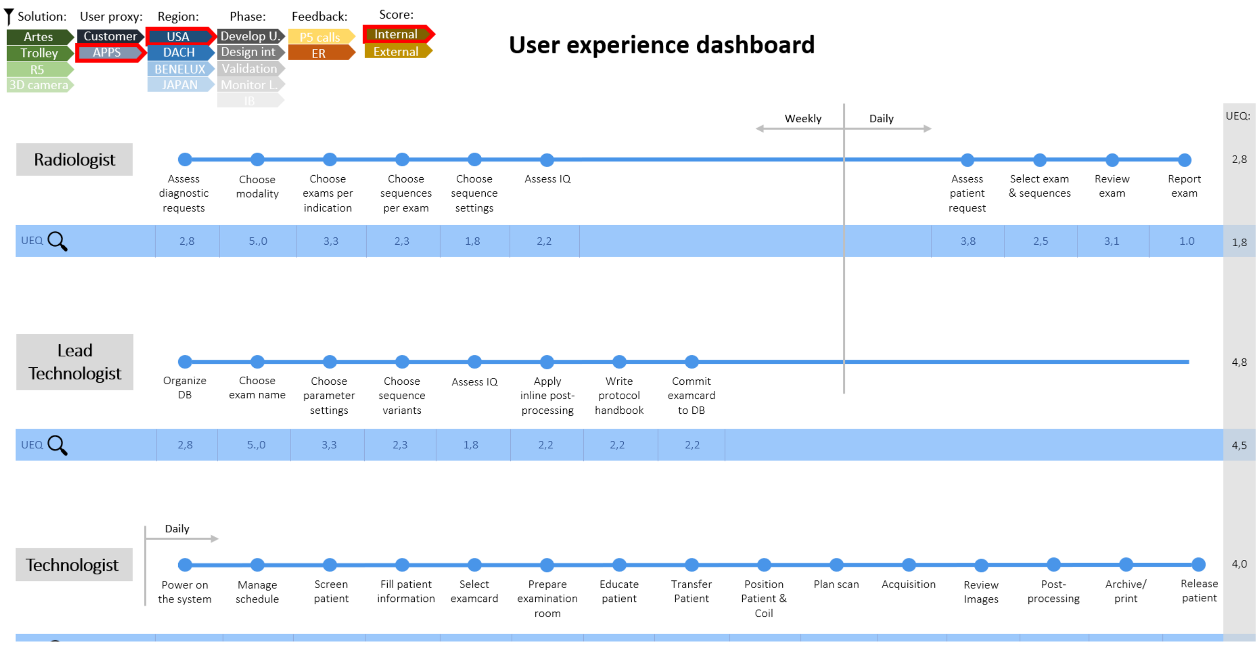


Figure 20. Example of user experience dashboard visualization

patient. This visualization was created by one of the application specialists. Next to the visualizations described here, many other visualizations can be made with the results of the questionnaire.

This section described the solution design that consists of two broad phases of Preparation and Evaluation, as well as the results of conducting the steps of preparation for the case company. This solution design answered Sub-question 5 which was:

"How to implement a standardized user experience evaluation tool to measure the experience of operators of Magnetic Resonance Imaging (MRI) systems in the First of a Kind (FOK) period of the MRI innovation process of the case company?"

The solution design was based on the design principles from theory. The next section will explain the reflection on the design principles, based on the empirical analysis of this research.

4.4 Reflection on design principles

The creation of a solution design was based on the design principles identified in the theoretical analysis. This section will reflect on these design principles based on the empirical analysis of this research. The reflection on the design principles will include results from the full empirical analysis. The design principles will be discussed around the four main themes: (1) user experience tools, (2) experience over time, (3) context of use, and (4) benchmarking and communication. At the end of this section, some additional insights will be discussed, which resulted in the construction of two new design principles.

4.4.1 User experience tools

Design Principle 01. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should use a standardized user experience evaluation tool (A) to quickly evaluate the medical systems with a reliable tool (M) which leads to an accurate quantitative evaluation of the user experience of medical systems that can be tracked over time (O).*

Design principle 01 was mainly validated with the explorative interviews of the empirical analysis during the main phase of Exploration and synthesis as the further steps of the solution design were based on the use of the standardized UEQ tool. It was evident that indeed a standardized user experience tool was desired in practice, due to the fact that it does not require much effort to set up as it has already been validated. The UEQ tool will fit the context of evaluating the user experience of MRI systems according to the empirical analysis. When showing the UEQ one employee from Design mentioned:

"It is really interesting, especially by the fact that it can be more focused on things that are more interesting to different departments." (Interview Designer 2, 2021).

The Clinical application manager mentioned:

"I think a standardized questionnaire that can be benchmarked is the best way to do this for now and the future as it provides with a tool to measure the user experience over time." (Interview Clinical application manager, 2021).

Overall, it seems that developers of MRI systems can benefit from using a standardized user experience tool, such as the UEQ. However, other standardized tools could be used if they are identified to be more appropriate in other contexts and new tools can be developed. Currently, only the UEQ allows to select the most relevant parameters for the specific context and includes the relative importance of each parameter, which will be discussed in the next design principle.

Design Principle 02. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should include the relative importance of pragmatic and hedonic parameters for the operators (A) to ensure that hedonic parameters do not disproportionally represent the evaluation of a product (M) which leads to a more accurate user experience evaluation of medical systems (O).*

The second design principle focused on the relative importance of all pragmatic and hedonic parameters for operators. There were some doubts about whether some hedonic parameters should be included in a user experience evaluation tool for MRI systems. However, multiple employees of the case company liked that the relative importance of each parameter ensures that the hedonic parameters are balanced with the pragmatic parameters, based on the importance question of the UEQ.

One employee from Marketing mentioned:

"I would not use it (importance of each parameter) for marketing material, but more just to know like what are the things that are really impactful for our customers?" (Interview Marketing, 2021).

One employee from Clinical applications mentioned:

"I think that it would be good to know what parameters bring the most value to users… and it would be good to say when something always scores low, we can exclude it from the evaluation." (Interview clinical application specialist 4, 2021).

Last, one employee from Design mentioned:

"I think that is needed right, otherwise it treats all the answers equally. And without this, actually, the score doesn't really make sense too." (Interview designer 1, 2021).

Developers that work on MRI systems that want to evaluate the user experience, should consider the relative importance of pragmatic and hedonic parameters. Some standardized tools treat pragmatic and hedonic parameters equally, which did not fit this context of MRI systems as it would probably overemphasize the importance of hedonic parameters. Other tools only use pragmatic parameters such as usability, which completely leave out hedonic parameters. It is still unclear how much impact the hedonic parameters have on the experience of operators of medical systems (Lallemand, Gronier & Koenig, 2015). An important additional insight from the second design principle is that analyzing the importance of each parameter allows removing parameters from the tool over time if they are deemed to be irrelevant by all operators.

4.4.2 Experience over time

Design principle 03. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should repeat user experience evaluations over time (A) to measure the dynamic changes in the experience of users with a product (M) which leads to a* *better understanding of how the experience of operators with medical products develops over time (O).*

As the user experience can develop over time, developers should repeat user experience evaluations over time to see how the user experience develops (Kujala et al., 2011). In practice, different employees from the case company have noticed with previous MRI system innovations that the opinion of operators can change over time.

However, they were unable to quantify this, which can be done with a user experience evaluation tool. As discussed before, it was decided to conduct the evaluations in the FOK period of the MRI innovation process after 1 week, 4 weeks, and 12 weeks. With these repetitions, trends can be observed in the adoption of the products. This already became clear from the interviews during the main phase of Exploration and synthesis:

"If we repeat the scores over time, we can see whether the scores increase or decrease over time … if we do this with the customer, we can see the adoption curve of the product ". (Interview Clinical application specialist 1, 2021).

Also, in the final validation session this was repeated:

"I think it will be good to evaluate the user experience multiple times, so we can see what the trend of the user experience is." (Final validation, 2021)

For MRI systems, the innovations are often quite complex, which takes time for operators to learn how to use them. When evaluating simple products, one evaluation might already give a stable result for the user experience evaluation.

4.4.3 Context of use

Design Principle 04. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should specify the specific context of use* ***with the use of labelling*** *for user experience evaluations (A) to analyze the effects of influencing factors on the user experience evaluation scores (M) which leads to a more accurate understanding of the experience of operators with medical systems (O).*

All participants agreed that the specific context of use is important for the evaluations. It was identified that it will be difficult to keep the context consistent for all the evaluations. For this reason, it was decided to add labels to the evaluations. The design principle 04 is adjusted as labelling was found to be a useful method for defining the context of the user experience evaluations. This way contextual factors can be analyzed separately, or different contexts can be compared to each other. The final list of labels was considered to be useful:

"I think these labels will help us with the analysis of the context. For some labels I doubt whether there will be any differences, but it won't harm to include these labels as it does not require any effort from the users" (Final validation, 2021).

One clinical application specialist did mention:

"You always need to do a deep dive on the experience. You cannot really fix something if you do not know the specific use cases. You really need to understand the context of the experience and you cannot fully attach that to a survey." (Interview Clinical application specialist 4, 2021).

Developers can decide to evaluate many different contexts, to limit the effect of external factors. However, this would require many more evaluations, which is usually not desired in practice. If very low scores occur on certain parameters of the user experience evaluation, developers can decide to do a qualitative follow-up interview with the operators to understand the issues.

4.4.4 Benchmarking and communication

Design principle 05. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should use internal benchmarking (A) to facilitate the evaluation and interpretation of the user experience of their systems with the use of a customized tool (M) which leads to an accurate understanding of the experience of operators with medical products that can be tracked over time and product iterations (O).*

Design principle 06. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should use external benchmarking (A) to facilitate the interpretation of the user experience evaluation of their systems through comparing with an online database (M) to understand how the user experience of the medical systems is relative to other products (O).*

All participants agreed that internally benchmarking the user experience evaluation will be useful. Relevant parameters from the UEQ were identified and customized for the evaluation of MRI systems for the case company, which can be benchmarked internally. The only downside for internal benchmarking is that it takes time to build up a database for the comparison of innovations to previous MRI systems.

One clinical application specialist thought that especially the internal benchmarking will add value:

"If you can benchmark compared to the previous version of the product … That will definitely have added value." (Interview Clinical application specialist 4, 2021).

Another clinical application specialist mentioned a different purpose:

"We can use these scores as an exit criterion, for example we want to reach this score for this product, otherwise we are not releasing." (Interview Clinical application specialist 3, 2021).

A participant from the department Quality & Regulatory mentioned:

"Personally, I think that it would be best to benchmark within our own scope. We could compare it to different medical systems within our company, but I think it will be difficult to compare to other products as you are implicitly or sometimes explicitly comparing to different contexts.". (Interview Quality & Regulatory, 2021)

As some participants argued that MRI systems should probably not be compared to completely different products, internal benchmarking seems to be the most appropriate option to evaluate the user experience of medical systems. Still, some participants explicitly wanted to be able to benchmark the results externally.

External benchmarking of the results of the UEQ can only be done with the six parameters from the traditional UEQ and does not allow to select the most relevant parameters of the user experience in the specific context. Although this will not result in the most comprehensive understanding of the user experience, it was still identified that external benchmarking will have added value. Especially the fact that the scores can be put in a category is expected to facilitate the communication within the case company.

The Clinical application manager emphasized the importance of the external benchmark:

"Having a framework for measuring the user experience that we can benchmark will be the most useful, we should not change too much to it as it is more important that the results can be compared with products in other domains or industries." (Clinical application manager, 2021).

Usually, external benchmarking is the best option when only evaluating a single product as it does not require to do multiple internal evaluations (Schrepp & Thomaschewski, 2019). With the selection of the relevant parameters for the evaluation of MRI systems, the results can be both internally and externally benchmarked. The advantages for developers of external benchmarking are that it can be done immediately without building up a database and that it can be categorized which can sometimes simplify the communication of the results.

Design principle 07. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should use visualizations (A) to simplify the interpretation and communication of the user experience of their systems (M) which leads to a better understanding and communication of the experience of operators with medical products (O).*

Visualizations are important for the communication of the results of the user experience evaluations (Lachner et al., 2016). However, it was identified that the overall score will be the most important to analyze for the case company. This means that not many different visualizations are necessary. However, it can depend on the audience of the results as one clinical application specialist mentioned:

"We can have quotes from different operators about their experience, but graphs tell a lot more and this is really what stakeholders want to see … It also depends on the audience, leadership is probably only interested in the high-level score, but others want to know what to improve" (Interview Clinical application specialist 4, 2021).

In the final validation, it was repeated that the overall score is the most important.

"*The overall score is the most important … with these visualizations we can do deeper analyses to understand what we need to improve"* (Final validation, 2021).

Visualizations are not always needed to communicate the results as sometimes only the overall user experience score is needed. However, visualization does have added value and the need for visualizations can depend on the goal and audience of the evaluation.

4.4.5 Additional insights

Next to evaluating the design principles, some new insights were developed from the empirical analysis. The first main insight was that there are differences in needs between departments. For example, the departments that are working on improving the MRI systems, want to retrieve negative points from the user experience evaluations to understand what to improve. On the contrary, Marketing wants to retrieve positive points from the user experience evaluations to communicate this with potential customers. An employee from the department of Marketing mentioned:

"Maybe on one scale it shows quite an average score, but there is one piece of that that's really high, I would like to be able to pull that out." (Interview Marketing, 2021).

One participant from the department Design mentioned:

"If you take the questionnaire to usability they will focus on efficiency and trust or something like that or confidence. And if you went by marketing they will focus on attractiveness and novelty for instance. But then how can we then compare with each other?" (Interview Designer 2, 2021).

It is important that the different departments do use the same questionnaires, as results can otherwise not be compared with each other.

Although the main aim for the case company was to derive a single overall score of the user experience, it was identified during the empirical analysis that many different methods have different advantages. The overall score, scores for specific parameters, the importance of the different parameters, and even results on single items were all identified to be useful for different departments.

The second main insight was that there are differences between software and physical MRI system innovations in terms of parameters that should be evaluated for the user experience. This resulted in the creation of a new design principle:

Design Principle 08. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should use different questionnaires for physical and software products (A) to allow to include different parameters in each of the questionnaires as different parameters are important for physical and software products (M) which leads to a more accurate understanding of the experience of operators with medical systems (O).*

Furthermore, the clinical application specialists are representatives or proxies of the actual users. They often have previously worked in hospitals and therefore understand the needs of users in practice. It was mentioned a few times that it is desired to identify whether there is a correlation between the clinical application specialists and actual users. If there is a high correlation, the proxies of the users can be more involved in testing innovation in early stages which leads to requiring less time and effort from actual users. This led to the development of the following design principle:

Design Principle 09. *If an organization want to integrate user experience evaluations within their innovation process of medical systems (C), it should measure the experience of operators and proxies or representatives of the operators (A) to measure whether there is a correlation between the actual operator and the representative (M) which leads to a better understanding of the accurateness of the opinion of operator representatives (O).*

All the initial design principles were discussed in this section and two new design principles were developed that can be used in future research and in practice. In the next, section the answers to the research questions will be explained and discussed.

5. Conclusion and Discussion

This research adopted the Design Science methodology, where insights from theory were evaluated in the empirical context of the case company and a solution design was created for this case company (Keskin & Romme, 2020). In this chapter, the answers to the five sub-questions and the main research question will be given. The Master Thesis research concludes with a discussion of the Theoretical implications, Managerial implications and Limitations, and future research.

5.1 Conclusion

This research consisted of a theoretical and empirical analysis to answer the main research question:

How can a standardized user experience evaluation tool be implemented into the innovation process of a company that develops MRI systems?

The answer to the main research question will follow from the collective answers to each of the five sub-questions. The answers to each sub-question and the main research question will be summarized below.

Sub-question 1: What are important parameters of the user experience and what standardized tools can developers use that can quantitatively evaluate these parameters according to academic literature?

The first and second sub-questions were answered by conducting a systematic academic literature review. Several researchers have previously tried to capture the important parameters of the user experience in a comprehensive theoretical framework (Hassenzahl, 2003; Mahlke, 2008; Mourouzis et al., 2006; Zarour & Alharbi, 2017).

From the theoretical analysis, it was concluded that it is difficult, if not impossible, to build a theoretical framework of all the important parameters of the user experience for all types of products (Zarour & Alharbi, 2017). Differences in the product influence what specific parameters are important for that specific context (Hassenzahl & Tractinsky, 2006; Wigelius & Väätäjä, 2009). Where previously the focus was mostly on evaluating the pragmatic (usability and utility) parameters of the medical systems, there is a need to broaden the focus to evaluating the whole user experience by also including hedonic (looks and feel) parameters (Bitkina et al., 2020; Brade et al., 2017; Hassenzahl, 2003; Roto et al., 2009). Although there is not enough evidence for the medical context, it seems that both pragmatic and hedonic parameters should be included in the evaluation of the user experience of medical systems (Hassenzahl, 2004).

There are three main standardized tools that focus on evaluating the whole user experience, which are the UEQ, the meCUE, and the Attrakdiff (Díaz-Oreiro et al., 2019). All three user experience evaluation tools include both pragmatic and hedonic parameters. The modular version of the UEQ allows to select the most relevant parameters and aims to balance the pragmatic and hedonic parameters by asking the participants how important they think that each parameter is (Hinderks et al., 2019). Therefore, the modular version of the UEQ seems to be the most appropriate user experience tool for the medical context (Hinderks et al., 2019).

Sub-question 2: How can a standardized user experience evaluation tool be implemented in the innovation process according to academic literature?

From the theoretical analysis, several important considerations for the implementation of a user experience tool in the innovation process were identified.

First of all, the specific context of the evaluation is important. User experience evaluations are likely to differ in focus and context for each evaluation, which requires that the context of the evaluations is specified (Hassenzahl & Tractinsky, 2006).

Second, it is important that the results can be clearly interpreted and communicated. Two main options for communicating the results are benchmarking and visualizations (Sauro & Lewis, 2016). Benchmarking is needed to give meaning to the user experience score, as a score on itself without benchmarking is difficult to interpret (Schrepp & Thomaschewski, 2019). Two types of benchmarking were identified, which will be referred to as internal and external benchmarking. External benchmarking requires to use of prespecified parameters but can be compared to an online database (Sauro & Lewis, 2016). Internal benchmarking allows the customization of the parameters included in the evaluation to be more comprehensive but if the tool is customized, it does not allow the comparison to an online database of evaluations (Schrepp et al., 2014). In addition, visualizations can also facilitate the interpretation and communication of the results (Lachner et al., 2016).

The last important consideration is to repeat the user experience evaluations over time, as previously it was shown that the user experience is not a static phenomenon and can develop over time (Kujala et al., 2011). Developers can therefore decide to do multiple user experience evaluations over time, without making changes to the product to see how the user experience develops over time.

Sub-question 3: How does the case company currently innovate their MRI systems and how do they integrate the perspective of operators in the innovation process?

The third and fourth sub-questions were answered by performing semi-structured interviews, informal meetings, and analyzing a database during the main phase of Exploration and synthesis of this research. The MRI innovation process of the case company can be divided into two main processes: Advanced Development (AD) and Planning Development Launch and Maintenance (PDLM). The AD process is a more iterative process where new ideas can be tested in a cheap and fast way. The PDLM process is a more structured process where commitment decisions are made for investing in innovations. The focus of this research was specifically on the FOK period of the Monitor launch phase of the PDLM process, which involves sending the product to a few sites to observe how the operators experience the product in the actual context. In the FOK period, the main focus is on evaluating the user experience of the actual operators, but clinical application specialists working within the case company are also involved as representatives of operators.

Sub-question 4: What are the design requirements for the implementation of a standardized user experience evaluation tool in the innovation process of MRI systems for the case company?

The design requirements for the case company were used as input for the creation of a solution design for the case company. The first functional requirement for the case company was that the tool should be able to evaluate the user experience in the context of operators of MRI systems. Furthermore, this tool was required to help to increase the focus on the experience of operators of

MRI systems in the innovation process. To help increase this focus on the experience of operators, it was required that the tool can be used for internal and external benchmarking, as internal benchmarking allows to be comprehensive by including the most relevant parameters and external benchmarking can facilitate the communication by comparing to an external database of evaluations (Schrepp et al., 2014). Furthermore, the tool was required to allow the evaluation of both physical and software MRI innovations.

In terms of requirements for the operators, it was required that the tool can be easily understood, as not understanding the evaluated parameter can result in frustration (Lallemand & Koenig, 2017). Furthermore, it was required that the results of the tool are easy to analyze and communicate within the company, otherwise, the tool would not bring much added value.

The boundary condition for the user experience tool was, that it should be able to be sent online and that it should be compatible with the current innovation process. Last, a design restriction was that the tool should not take longer than 5 to 7 minutes to complete for operators. These design requirements were used to create a solution design for implementing a standardized user experience evaluation tool for evaluating MRI systems, which will be described in the sub-question below.

Sub-question 5: How to implement a standardized user experience evaluation tool to measure the experience of operators of Magnetic Resonance Imaging (MRI) systems in the First of a Kind (FOK) period of the MRI innovation process of the case company?

Sub-question 5 was mainly answered by the creation of a design solution for the case company with the use of the answers to the previous four sub-questions. The design solution consists of process steps for the implementation of a standardized user experience evaluation tool in the FOK period of the T MRI innovation process. The solution consists of two main phases: Preparation and Evaluation. The preparation phase has to be done once and consists of identifying the main aim of the evaluation, the selection of an appropriate user experience tool, identifying the relevant parameters of the user experience, the customization of the tool, and the identification of the contextual labels. The labels are needed to specify and later analyze the specific context of the evaluation.

When the preparation phase is concluded, the phase of evaluation starts. This phase consists of labelling the evaluation, conducting the evaluation, analyzing the results, and communicating the results. The steps of the evaluation phase can be repeated for the following evaluation rounds.

It was identified during a final validation session that this process will help in evaluating the user experience for operators of MRI systems, as it provides a comprehensive understanding of the user experience of MRI innovations that can be analyzed with the use of benchmarking and labelling.

Main research question: How can a standardized user experience evaluation tool be implemented into the innovation process of a company that develops MRI systems?

The answer to the main research question followed from the answers to the five sub-questions mentioned above**.** The design solution, which was summarized for sub-question 5 shows specifically how the case company should implement a standardized user experience evaluation tool in their MRI innovation process. On a more general level, the main research question is answered by reflecting on the design principles synthesized from theory and the development of two new design principles. These principles and thus the main conclusions will be summarized here.

Developers should use standardized user experience evaluation tools that allow to include the relative importance of pragmatic and hedonic parameters for operators as developers can easily adopt these tools which have a focus on the whole user experience (Díaz-Oreiro et al., 2019; Laugwitz et al., 2008). The standardized tool allows developers to either use internal or external benchmarking or both, to give meaning to the user experience evaluation scores. Additionally, visualizing these results can have added value for the interpretation and communication of the results.

Developers should repeat the evaluations over time and specify the specific context of the evaluation by adding labels to the evaluation, as this facilitates the understanding of the temporal and contextual factors of the user experience. In this research, it was found that two different user experience evaluation questionnaires should be used for either software or physical MRI system innovations, as different user experience parameters are important for software and physical MRI innovations. Furthermore, this research found that developers can benefit from evaluating the user experience of their MRI system innovations by including actual operators of MRI systems as well as proxies or representatives of the operators. This allows to measure whether there is a strong correlation between the two and to identify whether evaluating with representatives is likely to give reliable user experience evaluation results. This research showed that the suggestions described here will help developers with the implementations of a user experience evaluation in their innovation process.

5.2 Discussion

This section will discuss the Theoretical implications, Managerial implications and Limitations, and future research.

5.2.1 Theoretical implications

This research followed a Design Science methodology to address the current gap in the academic literature on guidelines for the implementation of a user experience evaluation tool in the innovation process of companies (Keskin & Romme, 2020; Zarour & Alharbi, 2017).

Previous academic literature on evaluating medical systems mostly discussed the evaluation of pragmatic (usability and utility) parameters of medical systems (Bitkina et al., 2020; Hassenzahl, 2003; Sousa & Lopez, 2017). Bitkina et al., (2020) showed that developers of medical systems are likely to benefit from evaluating the whole user experience of medical systems which also includes evaluating hedonic (looks and feel) parameters. This research confirmed that focusing on the usability is not enough and that the whole user experience and thus both pragmatic and hedonic parameters should be considered for evaluating MRI systems and probably for evaluating other types of medical systems (Brade et al., 2017; Hassenzahl, 2003; Roto et al., 2009).

Although there are three main standardized user experience evaluation tools in the current academic literature that include both pragmatic and hedonic parameters, only the User Experience Questionnaire (UEQ) allows to balance these parameters by asking the operators how important they think the respective parameter is (Díaz-Oreiro et al., 2019; Hinderks et al., 2019)**.** For this reason, the flexibility of the UEQ is an added advantage for evaluating medical systems, as other tools have a fixed set of pragmatic and hedonic parameters, which might overemphasize the importance of the hedonic (looks and feel) parameters in an overall user experience score for evaluating MRI systems (Díaz-Oreiro et al., 2019). This research emphasized that including the relative importance of each parameter is an important consideration not only for MRI systems but also for other medical systems and products in general, as the user experience of different product types usually consists of different important parameters (Bitkina et al., 2020; Schrepp & Thomaschewski, 2019). The added value of the flexibility of tools for evaluating the user experience of products is not evident in the current academic literature, although developers of products and systems are likely to benefit from an increased focus on flexible and modular user experience evaluation tools.

The importance of each parameter of the user experience depends on the specific context and many researchers have suggested taking the context into account when evaluating the user experience of a product or system (Hassenzahl, 2003; Hassenzahl & Tractinsky, 2006; Mahlke,2008; Wigelius & Väätäjä, 2009). This research emphasizes the importance of considering the context of the user experience evaluations and suggests a new method to take the context of the evaluations into account, which is referred to as labelling. A label is a description of a contextual factor of the user experience evaluation that can be added to the evaluation. The main advantage of labelling is that over time, developers can analyze scores either with the same labels or with different labels in a database of evaluations, to understand what kind of impact these contextual factors might have on the user experience score.

For the interpretation and communication of the user experience evaluation results, two important considerations were identified in previous research, which were benchmarking and visualizations (Lachner et al., 2016; Sauro & Lewis, 2016; Schrepp, 2015; Schrepp & Thomaschewski, 2019). Benchmarking is an important consideration as s score on itself is difficult to interpret and previous research has suggested to use either internal or external benchmarking (Schrepp et al., 2014; Schrepp & Thomaschewski, 2019). However, this research showed that external and internal benchmarking can be combined to have the advantage of being comprehensive in the evaluation, but also facilitate the interpretation and communication of the results by comparing the scores to an external online database of evaluations. Although visualizations can facilitate the interpretation and communication of the results, visualizations are mostly needed for doing deeper analysis instead of using the user experience scores as an overall score or Key Performance Indicator, which was the main focus for the case company in this research (Lachner et al., 2016).

There are indications in previous research that expert users are important to involve in the innovation process of products and systems, as they are better at recognizing future needs (Roto et al., 2009). This research underlines these indications, as there was a strong need within the case company to evaluate the user experience of representatives of MRI systems operators. The main reason for this need is that the user experience scores can be correlated between actual users and user representatives to understand whether using user representatives in the innovation process is likely to give reliable results.

5.2.2 Managerial implications

This section will discuss the managerial implications of this research for the case company and for other companies that are developing MRI systems or other types of medical systems. This section will discuss how companies that develop medical systems should implement a user experience evaluation tool in their innovation process.

Currently, the main focus within the case company for evaluating MRI systems is on evaluating the pragmatic parameters (utility and usability) of the MRI systems. The results of this research showed that developers should focus on evaluating the whole user experience of their MRI systems and therefore also include hedonic (looks and feel) parameters next to the pragmatic parameters. This broader focus leads when evaluating innovations of MRI systems leads to a better understanding of how MRI system innovations are experienced by operators of these systems (Bitkina et al., 2020; Hassenzahl, 2003; Sousa & Lopez, 2017). A better understanding of the performance of the MRI system innovations allows to effectively improve the design of MRI systems, which has previously been shown to directly influence the safety and comfort of both patients and operators of MRI systems (Bitkina et al., 2020).

To increase the focus of the evaluation of MRI systems to the whole user experience, developers can best use a standardized user experience evaluation tool, as it is easy to use, reliable, and does not require much effort to construct (Díaz-Oreiro et al., 2019). Currently, the User Experience Questionnaire (UEQ) is the most appropriate tool in the context of MRI systems for evaluating the user experience as it balances the importance of the pragmatic and hedonic parameters by asking the operators how important they think each parameter is (Díaz-Oreiro et al., 2019; Hinderks et al., 2019)**.**

Furthermore, the UEQ has the advantage that an overall user experience Key Performance Indicator (KPI) can be calculated, which can be helpful for developers as it allows to easily keep track of the user experience and simplifies the analysis (Hinderks et al., 2019). Developers currently often use technical and financial Key Performance Indicators, to keep track of their performance and react if the scores decrease (Pan & Wei, 2012; Velimirović et al. , 2011). With the KPI of the UEQ the user experience of operators with MRI systems can be tracked next to keeping track of the technical and financial performance indicators to have an increased focus on the experience of operators in the innovation process (Hinderks et al., 2019)**.**

This research provides a customized UEQ questionnaire and guidelines for evaluating the user experience of MRI systems, that can be adopted by developers of MRI systems in the short-term. In the long-term for the implementation of a standardized user experience evaluation tool in other contexts, developers first need to consider what the main aim is of their evaluation, select the most appropriate tool, identify the most important parameters that they want to include in their evaluation, customize the evaluation tool and identify important contextual factors, which can be labelled to the evaluations.

The final step of identifying contextual factors and labelling the evaluations, allows developers to analyze the impact of contextual factors on the user experience scores. A label that shows the prior professional experience of operators allows to analyze the difference between user experience scores of operators with different levels of experience. For the case company, the labels have been constructed and can be used for the analysis of the experience of MRI systems.

To accurately interpret the results of the user experience evaluation, developers should benchmark the scores to previous evaluations (Sauro & Lewis, 2016; Schrepp, 2015; Schrepp & Thomaschewski, 2019). Scores can be benchmarked internally to previous internal product evaluations or externally to an online database of previous user experience evaluations (Schrepp et al., 2014; Schrepp & Thomaschewski, 2019). This research found that both types of benchmarking have added value and can be combined to both have a comprehensive evaluation of the user experience, but also allow to benchmark the scores to an online database of previous evaluations to facilitate the interpretation.

Last, this research indicates that evaluating the user experience of both operators and representatives of the operators can have added value. In the long-term the scores between actual operators and representatives can be correlated to assess whether it is reliable to evaluate representatives of the operators. If the results of representatives are reliable, developers can decide to increase the involvement of the representatives in the innovation process, which will probably save costs, time and puts less burden on actual operators to evaluate innovations.

5.2.3 Limitations and future research

A pilot study was conducted at the case company to develop guidelines for the implementation of a user experience evaluation tool for MRI systems. As other medical imaging systems show large similarities to MRI systems, it is likely that the results will generalize to other medical imaging systems. The results might not be generalizable to other types of medical systems and products, as MRI systems are complex systems that need to be evaluated comprehensively and for simple products, a less comprehensive evaluation might be sufficient. Future research can use the solution design and design principles that resulted from this research to evaluate whether the results of this research can be generalized to other products and systems.

This research focused on the implementation of a user experience evaluation tool in the innovation process, but only the steps of the preparation for the evaluation have been conducted. Future research can use the steps described in the solution design and the customized tool for the case company of this research to evaluate whether the guidelines and the UEQ work in practice.

Furthermore, this research focused on the context of the FOK period of the innovation process of the case company which involves the evaluation of a full MRI system. Future research can use the results of this research to identify whether the guidelines are appropriate for earlier stages of the innovation process, where ideas, concepts, and prototypes are evaluated. Earlier stages of the innovation process might require different tools as often the full experience cannot be evaluated yet (Roto et al., 2009).

This research specifically focused on quantitative tools for evaluating the user experience. Both the theoretical and empirical analysis showed that the quantitative measures should always be combined with qualitative tools, as qualitative tools allow to understand what and how MRI systems need to be improved (Ardito, Buono, Costabile, De Angeli & Lanzilotti, 2008). Future research can focus on how qualitative tools for evaluating the user experience can be combined with the quantitative tool of this research to help developers in understanding how they can improve their MRI systems.

References

Agnisarman, S. O., Madathil, K. C., Smith, K., Ashok, A., Welch, B., & McElligott, J. T. (2017). Lessons learned from the usability assessment of home-based telemedicine systems. Applied ergonomics, 58, 424-434.

Aldoihi, S., & Hammami, O. (2020). Factors contributing to CT scan usability. In 2020 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) (pp. 1-4). IEEE.

Ardito, C., Buono, P., Costabile, M. F., De Angeli, A., & Lanzilotti, R. (2008). Combining quantitative and qualitative data for measuring user experience of an educational game. In International Workshop on (Vol. 27).

Asbach, P., Canda, V., Hermann, K. G. A., Krug, L., Hahn, H. K., Hamm, B., & Klessen, C. (2008). Efficient whole-body MRI interpretation: evaluation of a dedicated software prototype. Journal of digital imaging, 21(1), 50-58.

Assila, A., & Ezzedine, H. (2016). Standardized usability questionnaires: Features and quality focus. Electronic Journal of Computer Science and Information Technology: eJCIST, 6(1).

Badran, O., & Al-Haddad, S. (2018). The Impact of Software User Experience on Customer Satisfaction. Journal of Management Information and Decision Sciences, 21(1), 1-20.

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

Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. Journal of Usability Studies, 11(3), 89-109.

Bitkina, O. V., Kim, H. K., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. International Journal of Industrial Ergonomics, 76, 102932.

Bloch, P. H., Brunel, F. F., & Arnold, T. J. (2003). Individual differences in the centrality of visual product aesthetics: Concept and measurement. Journal of consumer research, 29(4), 551-565.

Blunck, K. (2020). Application and Validation of the UEQ KPI in a hedonic context: A context-only extension study in the form of an online survey on Netflix.

Boothe, C. S. (2020). Generalized User Experience Questionnaire (UEQ-G): A Holistic Tool for Measuring Multimodal User Experiences. Mississippi State University.

Borsci, S., Buckle, P., & Walne, S. (2020). Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology?. *Applied ergonomics*, *84*, 103007.

Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. International Journal of Human-Computer Interaction, 31(8), 484-495.

Bosley, J. J. (2013). Creating a short usability metric for user experience (UMUX) scale. Interacting with Computers, 25(4), 317–319.

Brade, J., Lorenz, M., Busch, M., Hammer, N., Tscheligi, M., & Klimant, P. (2017). Being there again–Presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. International Journal of Human-Computer Studies, 101, 76-87.

Brooke, J. (1996). Sus: a "quick and dirty'usability. Usability evaluation in industry, 189.

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

Brooks, P., & Hestnes, B. (2010). User measures of quality of experience: why being objective and quantitative is important. IEEE network, 24(2), 8-13.

Cairns, P. (2013). A commentary on short questionnaires for assessing usability. Interacting with Computers, 25(4), 312–316.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014, September). The Use of Triangulation in Qualitative Research. In Oncology Nursing Forum (Vol. 41, No. 5).

Chandran, S., Al-Sa'di, A., & Ahmad, E. (2020). Exploring User Centered Design in Healthcare: A Literature Review. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 213-218).

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in society, 1(1), 104-126.

Creusen, M. E., & Schoormans, J. P. (1998). The influence of observation time on the role of the product design in consumer preference. ACR North American Advances.

Denyer, D., Tranfield, D., & Van Aken, J. E. (2008). Developing design propositions through research synthesis. Organization studies, 29(3), 393-413.

Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).

Doria, L., Minge, M., Riedel, L., & Kraft, M. (2013). User-centred evaluation of lower-limb orthoses: a new approach. Biomedical Engineering/Biomedizinische Technik, 58(SI-1-Track-J), 000010151520134232.

Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). Design science research. In Design science research (pp. 67-102). Springer, Cham.

Ellsworth, M. A., Dziadzko, M., O'Horo, J. C., Farrell, A. M., Zhang, J., & Herasevich, V. (2017). An appraisal of published usability evaluations of electronic health records via systematic review. Journal of the American Medical Informatics Association, 24(1), 218-226.

Fajar, A. N. (2019). E-learning implementation using user experience questionnaire. In Journal of Physics: Conference Series (Vol. 1367, No. 1, p. 012015). IOP Publishing

Finstad, K. (2010). The usability metric for user experience. Interacting with Computers, 22(5), 323-327.

Fischer, P. T., Kuliga, S., Eisenberg, M., & Amin, I. (2018). Space is part of the product: Using attrakdiff to identify spatial impact on user experience with media facades. In Proceedings of the 7th ACM International Symposium on Pervasive Displays (pp. 1-8).

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: An introduction. Longman Publishing.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. The qualitative report, 8(4), 597-607.

Halkier, B. (2011). Methodological practicalities in analytical generalization. Qualitative inquiry, 17(9), 787-797.

Han, J. H., & Lee, J. Y. (2021). Digital Healthcare Industry and Technology Trends. In 2021 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 375-377). IEEE.

Handayani, P. W., Hidayanto, A. N., Pinem, A. A., Sandhyaduhita, P. I., & Budi, I. (2018). Hospital information system user acceptance factors: User group perspectives. Informatics for Health and Social Care, 43(1), 84-107.

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. Human–Computer Interaction, 19(4), 319-349.

Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. In Funology (pp. 31-42). Springer, Dordrecht.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Mensch & computer 2003 (pp. 187-196). Vieweg+ Teubner Verlag.

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. Behaviour & information technology, 25(2), 91-97.

Hatscher, B., Luz, M., & Hansen, C. (2017). Foot interaction concepts to support radiological interventions. Mensch und Computer 2017-Tagungsband.

Hevner, A. R. (2007). A three cycle view of design science research. Scandinavian journal of information systems, 19(2), 4.

Hinderks, A., Schrepp, M., Mayo, F. J. D., Cuaresma, M. J. E., & Thomaschewski, J. (2018). UEQ KPI Value Range based on the UEQ Benchmark. University of Seville, Tech. Rep.

Hinderks, A., Schrepp, M., Mayo, F. J. D., Escalona, M. J., & Thomaschewski, J. (2019). Developing a UX KPI based on the user experience questionnaire. Computer Standards & Interfaces, 65, 38-44.

Holmström, J., Ketokivi, M., & Hameri, A. P. (2009). Bridging practice and theory: A design science approach. Decision Sciences, 40(1), 65-87.

Hussain, J., Khan, W. A., Hur, T., Bilal, H. S. M., Bang, J., Hassan, A. U., ... & Lee, S. (2018). A multimodal deep log-based user experience (UX) platform for UX evaluation. Sensors, 18(5), 1622.

International Organization for Standardization. (2019). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (ISO Standard No. 9241-210). Retrieved from https://www.iso.org/standard/77520.html

Ishak, Z., Fong, S. L., & Shin, S. C. (2019). SMART KPI management system framework. In 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) (pp. 172-177). IEEE.

Keskin, D., & Romme, G. (2020). Mixing oil with water: How to effectively teach design science in management education?. BAR-Brazilian Administration Review, 17.

Khairat, S., Coleman, C., Newlin, T., Rand, V., Ottmar, P., Bice, T., & Carson, S. S. (2019). A mixed-methods evaluation framework for electronic health records usability studies. Journal of biomedical informatics, 94, 103175.

Klaassen, B., van Beijnum, B. J., & Hermens, H. J. (2016). Usability in telemedicine systems—A literature survey. International journal of medical informatics, 93, 57-69.

Klein, A. M., Hinderks, A., Schrepp, M., & Thomaschewski, J. (2020). Measuring User Experience Quality of Voice Assistants Voice Communication Scales for the UEQ+ Framework: Voice Communication Scales for the UEQ+ Framework. In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-4). IEEE.

Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the system usability scale. International Journal of Human-Computer Interaction, 29(2), 67-76.

Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. International Journal of Human–Computer Interaction, 34(2), 177-186.

Kortum, P., & Peres, S. C. (2014). The relationship between system effectiveness and subjective usability scores using the System Usability Scale. International Journal of Human-Computer Interaction, 30(7), 575-584.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. Interacting with computers, 23(5), 473-483.

Lachner, F., Naegelein, P., Kowalski, R., Spann, M., & Butz, A. (2016). Quantified UX: Towards a common organizational understanding of user experience. In Proceedings of the 9th Nordic conference on human-computer interaction (pp. 1-10).

Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. Computers in Human Behavior, 43, 35-48.

Lallemand, C., & Koenig, V. (2017). How Could an Intranet be Like a Friend to Me? Why Standardized UX Scales Don't Always Fit. In Proceedings of the European Conference on Cognitive Ergonomics 2017 (pp. 9-16).

Lallemand, C., & Koenig, V. (2020). Measuring the Contextual Dimension of User Experience: Development of the User Experience Context Scale (UXCS). In Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (pp. 1-13).

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and usability engineering group (pp. 63-76). Springer, Berlin, Heidelberg.

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.

Lewis, J. R. (2013). Critical review of “the usability metric for user experience.” Interacting with Computers, 25(4), 320–324

Lewis, J. R. (2018). Measuring perceived usability: The CSUQ, SUS, and UMUX. International Journal of Human–Computer Interaction, 34(12), 1148-1156.

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In International conference on human centered design (pp. 94-103). Springer, Berlin, Heidelberg.

Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. Journal of Usability Studies, 13(3).

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

Longhurst, R. (2003). Semi-structured interviews and focus groups. Key methods in geography, 3(2), 143-156.

Mahlke, S. (2008). User experience of interaction with technical systems.

Maia, C. L. B., & Furtado, E. S. (2016). A systematic review about user experience evaluation. In International Conference of Design, User Experience, and Usability (pp. 445-455). Springer, Cham.

McLellan, S., Muddimer, A., & Peres, S. C. (2012). The effect of experience on System Usability Scale ratings. Journal of usability studies, 7(2), 56-67.

Minge, M., Thüring, M., Wagner, I., & Kuhr, C. V. (2017). The meCUE questionnaire: a modular tool for measuring user experience. In Advances in Ergonomics Modeling, Usability & Special Populations (pp. 115-128). Springer, Cham.

Moumane, K., Idri, A., & Abran, A. (2016). Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards. SpringerPlus, 5(1), 1-15.

Mourouzis, A., Antona, M., Boutsakis, E., & Stephanidis, C. (2006). A user-orientation evaluation framework: Assessing accessibility throughout the user experience lifecycle. In International Conference on Computers for Handicapped Persons (pp. 421-428). Springer, Berlin, Heidelberg.

Pan, W., & Wei, H. (2012). Research on key performance indicator (KPI) of business process. In 2012 Second International Conference on Business Computing and Global Informatization (pp. 151-154). IEEE.

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

Pickup, L., Nugent, B., & Bowie, P. (2019). A preliminary ergonomic analysis of the MRI work system environment: Implications and recommendations for safety and design. Radiography, 25(4), 339-345.

QSR International Pty Ltd. (2018) NVivo (Version 12), https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137-141.

Randolph, J. (2009). A guide to writing the dissertation literature review. Practical Assessment, Research, and Evaluation, 14(1), 13.

Reichheld, F. F. (2003). The one number you need to grow. Harvard business review, 81(12), 46-55.

Roto, V., Obrist, M., & Väänänen-Vainio-Mattila, K. (2009). User experience evaluation methods in academic and industrial contexts. In Proceedings of the Workshop UXEM (Vol. 9, pp. 1-5).

Roto, V., Rantavuo, H., & Väänänen-Vainio-Mattila, K. (2009). Evaluating user experience of early product concepts. In Proc. DPPI (Vol. 9, pp. 199-208).

Saldaña, J. (2009). The coding manual for qualitative researchers.

Santoso, H. B., & Schrepp, M. (2018). Importance of User Experience Aspects for Different Software Product Categories. In International Conference on User Science and Engineering (pp. 231-241). Springer, Singapore.

Santoso, H., Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Cultural differences in the perception of user experience. Mensch und Computer 2017-Tagungsband.

Sauer, J., Hein, A., Muenzberg, A., & Roesch, N. (2019). Simplify testing of mobile medical applications by using timestamps for remote, automated evaluation. In 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 203-206). IEEE.

Sauro, J., & Lewis, J. R. (2016). Quantifying the user experience: Practical statistics for user research. Morgan Kaufmann.

Schrepp, M. (2015). User experience questionnaire handbook. All you need to know to apply the UEQ successfully in your project.

Schrepp, M., Held, T., & Laugwitz, B. (2006). The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers, 18(5), 1055-1069.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In International Conference of Design, User Experience, and Usability (pp. 383-392). Springer, Cham.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017a). Construction of a Benchmark for the User Experience Questionnaire (UEQ). IJIMAI, 4(4), 40-44.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017b). Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). Ijimai, 4(6), 103-108.

Schrepp, M., & Thomaschewski, J. (2019). Handbook for the modular extension of the User Experience Questionnaire. All you need to know to apply the UEQ+ to create your own UX questionnaire.

Shah, S. G. S., & Robinson, I. (2008). Medical device technologies: who is the user?. International Journal of Healthcare Technology and Management, 9(2), 181-197.

Shah, S. G. S., Robinson, I., & AlShawi, S. (2009). Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process.

Sharma, A., Malviya, R., Awasthi, R., & Sharma, P. K. (2020). Artificial Intelligence, Blockchain, and Internet of Medical Things: New Technologies in Detecting, Preventing, and Controlling of Emergent Diseases. In Advances in Multidisciplinary Medical Technologies– Engineering, Modeling and Findings (pp. 127-154). Springer, Cham.

Shih, P. Y., & Zheng, M. C. (2020). Usability Study of a Pre-anesthesia Evaluation App in an University Hospital: Before the Revision of User Interface. In International Conference on Human-Computer Interaction (pp. 480-487). Springer, Cham.

Sittig, D. F., Kuperman, G. J., & Fiskio, J. (1999). Evaluating physician satisfaction regarding user interactions with an electronic medical record system. In Proceedings of the AMIA Symposium (p. 400). American Medical Informatics Association.

Sousa, V. E., & Lopez, K. D. (2017). Towards usable e-health: A systematic review of usability questionnaires. Applied clinical informatics, 8(2), 470.

Sproll, S., Peissner, M., & Sturm, C. (2010). From product concept to user experience: exploring UX potentials at early product stages. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (pp. 473-482).

Stauder, R., Belagiannis, V., Schwarz, L. A., Bigdelou, A., Söhngen, E., Ilic, S., & Navab, N. (2012). A user-centered and workflow-aware unified display for the operating room. In MICCAI Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI), Nice, France.

Stola, K. (2018). User experience and design thinking as a global trend in healthcare. Journal of Medical Science, 87(1), 28-33.e

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. International journal of psychology, 42(4), 253-264.

Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (pp. 115-122).

Tullis, T. S., & Stetson, J. N. (2004, June). A comparison of questionnaires for assessing website usability. In Usability professional association conference (Vol. 1, pp. 1-12).

van Aken, J. E., Berends, J. J., & van der Bij, H. (2012). Problem Solving in Organizations: A Methodological Handbook for Business and Management Students. Cambridge University Press.

Van Burg, E., & Romme, A. G. L. (2014). Creating the future together: Toward a framework for research synthesis in entrepreneurship. Entrepreneurship Theory and Practice, 38(2), 369-397.

Van Burg, E., Romme, A. G. L., Gilsing, V. A., & Reymen, I. M. (2008). Creating university spin-offs: A science-based design perspective. Journal of Product Innovation Management, 25(2), 114-128.

Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008a). Now let's do it in practice: user experience evaluation methods in product development. In CHI'08 extended abstracts on Human factors in computing systems (pp. 3961-3964).

Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008b). Towards practical user experience evaluation methods. Meaningful measures: Valid useful user experience measurement (VUUM), 19-22.

Valadi, S., & Broneske, (2020). Analysis of Current Usability and User Experience Questionnaires and Creating an Optimized Usability Questionnaire.

Velimirović, D., Velimirović, M., & Stanković, R. (2011). Role and importance of key performance indicators measurement. Serbian Journal of Management, 6(1), 63-72.

Verint. (2018). EFM Verint (Version 15.2).

Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries (pp. 521-530).

von Wilamowitz-Moellendorff, M., Hassenzahl, M., & Platz, A. (2006). Dynamics of user experience: How the perceived quality of mobile phones changes over time. In User Experience-Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer Interaction (pp. 74-78).

Walsh, T., Varsaluoma, J., Kujala, S., Nurkka, P., Petrie, H., & Power, C. (2014). Axe UX: exploring long-term user experience with iScale and AttrakDiff. In Proceedings of the 18th international academic mindtrek conference: Media business, management, content & services (pp. 32-39).

Web of Science. (2021). Retrieved 5 August 2021, from https://www-webofscience-com.dianus.libr.tue.nl/wos/woscc/basic-search

Wigelius, H., & Väätäjä, H. (2009). Dimensions of context affecting user experience in mobile work. In IFIP Conference on Human-Computer Interaction (pp. 604-617). Springer, Berlin, Heidelberg.

Yin, R. K. (2013). Validity and generalization in future case study evaluations. Evaluation, 19(3), 321-332.

Yin, R. K. (2017). Case Study Research and Applications: Design and Methods. SAGE Publications.

Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.

Zhou, L., Bao, J., & Parmanto, B. (2017). Systematic review protocol to assess the effectiveness of usability questionnaires in mhealth app studies. JMIR research protocols, 6(8), e151.

Appendix 1. Search string and Inclusion criteria

User experience framework of aspects search string	Criteria
TI= (user experience OR UX) AND TI= (framework OR parameter OR definition) AND AB= (product* OR service* OR system*)	English only, Articles and conference papers only, (Medical context included)
User experience tools and questionnaires search string	
TI= (user experience OR UX OR usability) AND TI= (questionnaire* OR tool*) AND TI= (product* OR service* OR system*)	English only, Articles and conference papers only, (Medical context included)
User experience process implementation search string	
TI= (user experience OR UX OR usability) AND AB= (innovation process OR product development OR NPD) AND AB= (product* OR service* OR system*) AND AB= (questionnaire* OR tool*)	English only, Articles and conference papers only, (Medical context included)

Appendix 2. Method of finding articles

Article Reference	Search method	Lens number
Agnisarman, S. O., Madathil, K. C., Smith, K., Ashok, A., Welch, B., & McElligott, J. T. (2017). Lessons learned from the usability assessment of home-based telemedicine systems. Applied ergonomics, 58, 424-434.	**Backward snowballing** from Borsci, S., Buckle, P., & Walne, S. (2020). Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology?. Applied ergonomics, 84, 103007.	3
van Aken, J. E., Berends, J. J., & van der Bij, H. (2012). Problem Solving in Organizations: A Methodological Handbook for Business and Management Students. Cambridge University Press.	**Expert recommendation**	-
Aldoihi, S., & Hammami, O. (2020). Factors contributing to CT scan usability. In 2020 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) (pp. 1-4). IEEE.	**Google Scholar:** "Usability CT scan"	1
Ardito, C., Buono, P., Costabile, M. F., De Angeli, A., & Lanzilotti, R. (2008). Combining quantitative and qualitative data for measuring user experience of an educational game. In International Workshop on (Vol. 27).	**Google Scholar :** "Quantitative user experience »	3
Asbach, P., Canda, V., Hermann, K. G. A., Krug, L., Hahn, H. K., Hamm, B., & Klessen, C. (2008). Efficient whole-body MRI interpretation: evaluation of a dedicated software prototype. Journal of digital imaging, 21(1), 50-58.	**Forward snowballing** Brooke, J. (1996). Sus: a "quick and dirty'usability. Usability evaluation in industry, 189.	3
Assila, A., & Ezzedine, H. (2016). Standardized usability questionnaires: Features and quality focus. Electronic Journal of Computer Science and Information Technology: eJCIST, 6(1).	**Forward snowballing** Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. International Journal of Human-Computer Interaction, 31(8), 484-495.	1
Badran, O., & Al-Haddad, S. (2018). The Impact of Software User Experience on Customer Satisfaction. Journal of Management Information and Decision Sciences, 21(1), 1-20.	**Forward snowballing** Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries (pp. 521-530).	-

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies, 4(3), 114-123.	Web of science search string 3	3
Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. Journal of Usability Studies, 11(3), 89-109.	**Forward snowballing** Lewis, J. R. (2013). Critical review of "the usability metric for user experience." Interacting with Computers, 25(4), 320–324	2 & 3
Bitkina, O. V., Kim, H. K., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. International Journal of Industrial Ergonomics, 76, 102932.	**Backward snowballing** Chandran, S., Al-Sa'di, A., & Ahmad, E. (2020). Exploring User Centered Design in Healthcare: A Literature Review. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.	1, 2 & 3
Bloch, P. H., Brunel, F. F., & Arnold, T. J. (2003). Individual differences in the centrality of visual product aesthetics: Concept and measurement. Journal of consumer research, 29(4), 551-565.	**Backward snowballing** Mahlke, S. (2008). User experience of interaction with technical systems.	3
Blunck, K. (2020). Application and Validation of the UEQ KPI in a hedonic context: A context-only extension study in the form of an online survey on Netflix.	**Forward snowballing** Hinderks, A., Schrepp, M., Mayo, F. J. D., Escalona, M. J., & Thomaschewski, J. (2019). Developing a UX KPI based on the user experience questionnaire. Computer Standards & Interfaces, 65, 38-44.	3
Boothe, C. S. (2020). Generalized User Experience Questionnaire (UEQ-G): A Holistic Tool for Measuring Multimodal User Experiences. Mississippi State University.	**Forward snowballing** Lewis, J. R. (2013). Critical review of "the usability metric for user experience." Interacting with Computers, 25(4), 320–324	2
Borsci, S., Buckle, P., & Walne, S. (2020). Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology?. *Applied ergonomics*, *84*, 103007.	**Forward snowballing** Lewis, J. R. (2013). Critical review of "the usability metric for user experience." Interacting with Computers, 25(4), 320–324	1, 2 & 3
Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. International Journal of Human-Computer Interaction, 31(8), 484-495.	**Backward snowballing** Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. Journal of Usability Studies, 11(3), 89-109.	3
Bosley, J. J. (2013). Creating a short usability metric for user experience (UMUX) scale. Interacting with Computers, 25(4), 317–319.	**Backward snowballing** Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. Journal of Usability Studies, 11(3), 89-109.	2

Brooke, J. (1996). Sus: a "quick and dirty'usability. Usability evaluation in industry, 189.	**Backward snowballing** Sousa, V. E., & Lopez, K. D. (2017). Towards usable e-health: A systematic review of usability questionnaires. Applied clinical informatics, 8(2), 470.	2
Brooke, J. (2013). SUS: a retrospective. Journal of usability studies, 8(2), 29-40.	**Backward snowballing** Sousa, V. E., & Lopez, K. D. (2017). Towards usable e-health: A systematic review of usability questionnaires. Applied clinical informatics, 8(2), 470.	2
Brooks, P., & Hestnes, B. (2010). User measures of quality of experience: why being objective and quantitative is important. IEEE network, 24(2), 8-13.	**Google Scholar :** "Quantitative user experience »	1
Cairns, P. (2013). A commentary on short questionnaires for assessing usability. Interacting with Computers, 25(4), 312–316.	**Backward snowballing** Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. Journal of Usability Studies, 11(3), 89-109.	2
Chandran, S., Al-Sa'di, A., & Ahmad, E. (2020). Exploring User Centered Design in Healthcare: A Literature Review. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.	**Backward snowballing** Bitkina, O. V., Kim, H. K., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. International Journal of Industrial Ergonomics, 76, 102932.	-
Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 213-218).	**Backward snowballing** Sousa, V. E., & Lopez, K. D. (2017). Towards usable e-health: A systematic review of usability questionnaires. Applied clinical informatics, 8(2), 470.	2
Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in society, 1(1), 104-126.	**Expert recommendation**	-
Creusen, M. E., & Schoormans, J. P. (1998). The influence of observation time on the role of the product design in consumer preference. ACR North American Advances.	**Backward snowballing** Mahlke, S. (2008). User experience of interaction with technical systems.	3
Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).	**Backward snowballing** Blunck, K. (2020). Application and Validation of the UEQ KPI in a hedonic context: A context-only extension study in the form of an online survey on Netflix	2 & 3
Doria, L., Minge, M., Riedel, L., & Kraft, M. (2013). User-centred evaluation of lower-limb orthoses: a new approach. Biomedical	**Backward snowballing** Minge, M., Thüring, M., Wagner, I., & Kuhr, C. V. (2017). The meCUE questionnaire: a modular tool for measuring user experience. In Advances	3

Engineering/Biomedizinische Technik, 58(SI-1-Track-J), 000010151520134232.	in Ergonomics Modeling, Usability & Special Populations (pp. 115-128). Springer, Cham.	
Ellsworth, M. A., Dziadzko, M., O'Horo, J. C., Farrell, A. M., Zhang, J., & Herasevich, V. (2017). An appraisal of published usability evaluations of electronic health records via systematic review. Journal of the American Medical Informatics Association, 24(1), 218-226.	**Backward snowballing** Chandran, S., Al-Sa'di, A., & Ahmad, E. (2020). Exploring User Centered Design in Healthcare: A Literature Review. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.	2 & 3
Fajar, A. N. (2019). E-learning implementation using user experience questionnaire. In Journal of Physics: Conference Series (Vol. 1367, No. 1, p. 012015). IOP Publishing	**Forward snowballing** Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and usability engineering group (pp. 63-76). Springer, Berlin, Heidelberg.	3
Finstad, K. (2010). The usability metric for user experience. Interacting with Computers, 22(5), 323-327.	**Forward snowballing** Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In International conference on human centered design (pp. 94-103). Springer, Berlin, Heidelberg.	2
Fischer, P. T., Kuliga, S., Eisenberg, M., & Amin, I. (2018). Space is part of the product: Using attrakdiff to identify spatial impact on user experience with media facades. In Proceedings of the 7th ACM International Symposium on Pervasive Displays (pp. 1-8).	**Forward snowballing** Hassenzahl, M., et al. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Proc. Of Mensch & Computer 2003: Interaktion in Bewegung. (2003). 187 - 196.	3
Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: An introduction. Longman Publishing.	**Expert recommendation**	-
Han, J. H., & Lee, J. Y. (2021). Digital Healthcare Industry and Technology Trends. In 2021 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 375-377). IEEE.	**Google scholar:** "trends in healthcare technology"	-
Handayani, P. W., Hidayanto, A. N., Pinem, A. A., Sandhyaduhita, P. I., & Budi, I. (2018). Hospital information system user acceptance factors: User group perspectives. Informatics for Health and Social Care, 43(1), 84-107.	**Google scholar:** "factors of the user experience hospital"	1
Hassenzahl, M. (2003). The thing and I: understanding the relationship between user and product. In Funology 2 (pp. 301-313). Springer, Cham.	**Backward snowballing** Mahlke, S. (2008). User experience of interaction with technical systems.	1 & 3
Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. Human–Computer Interaction, 19(4), 319-349.	**Backward snowballing from** Roto, V., Obrist, M., & Väänänen-Vainio-Mattila, K. (2009). User experience evaluation methods in academic and industrial contexts. In Proceedings of the Workshop UXEM (Vol. 9, pp. 1-5).	3

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Mensch & computer 2003 (pp. 187-196). Vieweg+ Teubner Verlag.	**Backward snowballing** Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).	2
Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. Behaviour & information technology, 25(2), 91-97.	**Backward snowballing** Bitkina, O. V., Kim, H. K., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. International Journal of Industrial Ergonomics, 76, 102932.	3
Hatscher, B., Luz, M., & Hansen, C. (2017). Foot interaction concepts to support radiological interventions. Mensch und Computer 2017-Tagungsband.	**Forward snowballing** Minge, M., Thüring, M.,Wagner, I., & Kuhr, C. V. (2017). The meCUE Questionnaire:A Modular Tool for Measuring User Experience. In Advances in Ergonomics Modeling, Usability & Special Populations (pp. 115–128). Springer	3
Hinderks, A., Schrepp, M., Mayo, F. J. D., Cuaresma, M. J. E., & Thomaschewski, J. (2018). UEQ KPI Value Range based on the UEQ Benchmark. University of Seville, Tech. Rep.	**Foward snowballing** Schrepp, M. (2015). User experience questionnaire handbook. All you need to know to apply the UEQ successfully in your project.	3
Hinderks, A., Schrepp, M., Mayo, F. J. D., Escalona, M. J., & Thomaschewski, J. (2019). Developing a UX KPI based on the user experience questionnaire. Computer Standards & Interfaces, 65, 38-44.	**Forward snowballing** Schrepp, M. (2015). User experience questionnaire handbook. All you need to know to apply the UEQ successfully in your project.	2 & 3
Hussain, J., Khan, W. A., Hur, T., Bilal, H. S. M., Bang, J., Hassan, A. U., ... & Lee, S. (2018). A multimodal deep log-based user experience (UX) platform for UX evaluation. Sensors, 18(5), 1622.	**Forward snowballing** Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries (pp. 521-530).	2 & 3
International Organization for Standardization. (2019). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (ISO Standard No. 9241-210). Retrieved from https://www.iso.org/standard/77520.html	**Google search:** "ISO standards"	1
Ishak, Z., Fong, S. L., & Shin, S. C. (2019). SMART KPI management system framework. In 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) (pp. 172-177). IEEE.	**Google scholar:** "SMART KPI"	3
Khairat, S., Coleman, C., Newlin, T., Rand, V., Ottmar, P., Bice, T., & Carson, S. S. (2019). A	**Backward snowballing** Chandran, S., Al-Sa'di, A., & Ahmad, E. (2020). Exploring User Centered	-

mixed-methods evaluation framework for electronic health records usability studies. Journal of biomedical informatics, 94, 103175.	Design in Healthcare: A Literature Review. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.	
Klaassen, B., van Beijnum, B. J., & Hermens, H. J. (2016). Usability in telemedicine systems—A literature survey. International journal of medical informatics, 93, 57-69.	**Backward snowballing** from Borsci, S., Buckle, P., & Walne, S. (2020). Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology?. *Applied ergonomics*, *84*, 103007.	3
Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the system usability scale. International Journal of Human-Computer Interaction, 29(2), 67-76.	**Backward snowballing** Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. Journal of Usability Studies, 13(3).	3
Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. International Journal of Human–Computer Interaction, 34(2), 177-186.	**Forward snowballing** Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the system usability scale. International Journal of Human-Computer Interaction, 29(2), 67-76**.**	3
Kortum, P., & Peres, S. C. (2014). The relationship between system effectiveness and subjective usability scores using the System Usability Scale. International Journal of Human-Computer Interaction, 30(7), 575-584.	**Backward snowballing** Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. Journal of Usability Studies, 13(3).	1
Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. Interacting with computers, 23(5), 473-483.	**Forward snowballing** Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries (pp. 521-530).	3
Lachner, F., Naegelein, P., Kowalski, R., Spann, M., & Butz, A. (2016). Quantified UX: Towards a common organizational understanding of user experience. In Proceedings of the 9th Nordic conference on human-computer interaction (pp. 1-10).	Web of science search string 3	3
Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. Computers in Human Behavior, 43, 35-48.	**Backward snowballing** from Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. International Journal of Human-Computer Interaction, 31(8), 484-495.	1
Lallemand, C., & Koenig, V. (2017). How Could an Intranet be Like a Friend to Me? Why	**Backward snowballing** Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019).	2 & 3

Standardized UX Scales Don't Always Fit. In Proceedings of the European Conference on Cognitive Ergonomics 2017 (pp. 9-16).	Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).	
Lallemand, C., & Koenig, V. (2020). Measuring the Contextual Dimension of User Experience: Development of the User Experience Context Scale (UXCS). In Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (pp. 1-13).	**Forward snowballing** Lallemand, C., & Koenig, V. (2017). How Could an Intranet be Like a Friend to Me? Why Standardized UX Scales Don't Always Fit. In Proceedings of the European Conference on Cognitive Ergonomics 2017 (pp. 9-16).	3
Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and usability engineering group (pp. 63-76). Springer, Berlin, Heidelberg.	**Backward snowballing** Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).	2 & 3
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.	**Backward snowballing** from Tullis, T. S., & Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. In Usability professional association conference (Vol. 1, pp. 1-12).	2
Lewis, J. R. (2013). Critical review of "the usability metric for user experience." Interacting with Computers, 25(4), 320–324	**Backward snowballing** Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. Journal of Usability Studies, 11(3), 89-109.	2
Lewis, J. R. (2018). Measuring perceived usability: The CSUQ, SUS, and UMUX. International Journal of Human–Computer Interaction, 34(12), 1148-1156.	**Forward snowballing** Lewis, J. R. (2013). Critical review of "the usability metric for user experience." Interacting with Computers, 25(4), 320–324	2
Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In International conference on human centered design (pp. 94-103). Springer, Berlin, Heidelberg.	**Backwards snowballing** Brooke, J. (2013). SUS: a retrospective. Journal of usability studies, 8(2), 29-40.	2
Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. Journal of Usability Studies, 13(3).	**Forward snowballing** Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In International conference on human centered design (pp. 94-103). Springer, Berlin, Heidelberg.	2 & 3
Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE: when there's no time for the SUS. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2099-2102).	**Forward snowballing** Lewis, J. R. (2013). Critical review of "the usability metric for user experience." Interacting with Computers, 25(4), 320–324	2

Mahlke, S. (2008). User experience of interaction with technical systems.	**Backward snowballing** Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. Interacting with computers, 23(5), 473-483.	1, 2 & 3
McLellan, S., Muddimer, A., & Peres, S. C. (2012). The effect of experience on System Usability Scale ratings. Journal of usability studies, 7(2), 56-67.	**Backward snowballing** Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. International Journal of Human–Computer Interaction, 34(2), 177-186.	3
Minge, M., Thüring, M., Wagner, I., & Kuhr, C. V. (2017). The meCUE questionnaire: a modular tool for measuring user experience. In Advances in Ergonomics Modeling, Usability & Special Populations (pp. 115-128). Springer, Cham.	**Backward snowballing** Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).	2 & 3
Moumane, K., Idri, A., & Abran, A. (2016). Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards. SpringerPlus, 5(1), 1-15.	**Backward snowballing** Valadi, S., & Broneske, (2020). Analysis of Current Usability and User Experience Questionnaires and Creating an Optimized Usability Questionnaire.	2
Mourouzis, A., Antona, M., Boutsakis, E., & Stephanidis, C. (2006). A user-orientation evaluation framework: Assessing accessibility throughout the user experience lifecycle. In International Conference on Computers for Handicapped Persons (pp. 421-428). Springer, Berlin, Heidelberg.	Web of science search string 1	1
Pan, W., & Wei, H. (2012). Research on key performance indicator (KPI) of business process. In 2012 Second International Conference on Business Computing and Global Informatization (pp. 151-154). IEEE.	**Google scholar:** "Key performance indicators kpi"	3
Petrie, H., & Bevan, N. (2009). The Evaluation of Accessibility, Usability, and User Experience. The universal access handbook, 1, 1-16.	**Backward snowballing** Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	3
Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137-141.	**Google scholar:** "Black box AI"	3
Randolph, J. (2009). A guide to writing the dissertation literature review. Practical Assessment, Research, and Evaluation, 14(1), 13.	**Forward snowballing** Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in society, 1(1), 104-126.	-

Reichheld, F. F. (2003). The one number you need to grow. Harvard business review, 81(12), 46-55.	**Backward snowballing** Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. International Journal of Human–Computer Interaction, 34(2), 177-186.	2
Roto, V., Obrist, M., & Väänänen-Vainio-Mattila, K. (2009). User experience evaluation methods in academic and industrial contexts. In Proceedings of the Workshop UXEM (Vol. 9, pp. 1-5).	**Backward snowballing** Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	-
Roto, V., Rantavuo, H., & Väänänen-Vainio-Mattila, K. (2009). Evaluating user experience of early product concepts. In Proc. DPPI (Vol. 9, pp. 199-208).	**Backward snowballing** Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	3
Santoso, H. B., & Schrepp, M. (2018). Importance of User Experience Aspects for Different Software Product Categories. In International Conference on User Science and Engineering (pp. 231-241). Springer, Singapore.	**Forward snowballing** Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. Computers in Human Behavior, 43, 35-48.	3
Santoso, H., Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Cultural differences in the perception of user experience. Mensch und Computer 2017-Tagungsband.	**Backward snowballing** Santoso, H. B., & Schrepp, M. (2018). Importance of User Experience Aspects for Different Software Product Categories. In International Conference on User Science and Engineering (pp. 231-241). Springer, Singapore.	3
Sauer, J., Hein, A., Muenzberg, A., & Roesch, N. (2019). Simplify testing of mobile medical applications by using timestamps for remote, automated evaluation. In 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 203-206). IEEE.	**Forward snowballing** Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and usability engineering group (pp. 63-76). Springer, Berlin, Heidelbe	3
Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.	**Backward snowballing** Valadi, S., & Broneske, (2020). Analysis of Current Usability and User Experience Questionnaires and Creating an Optimized Usability Questionnaire.	2 & 3
Schrepp, M. (2015). User experience questionnaire handbook. All you need to know to apply the UEQ successfully in your project.	**Backward snowballing** Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017b). Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). Ijimai, 4(6), 103-108.	3
Schrepp, M., Held, T., & Laugwitz, B. (2006). The influence of hedonic quality on the attractiveness of user interfaces of business	**Backward snowballing** Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In	2

management software. Interacting with Computers, 18(5), 1055-1069.	Symposium of the Austrian HCI and usability engineering group (pp. 63-76). Springer, Berlin, Heidelbe	
Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In International Conference of Design, User Experience, and Usability (pp. 383-392). Springer, Cham.	***Backward snowballing*** Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).	2 & 3
Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017a). Construction of a Benchmark for the User Experience Questionnaire (UEQ). IJIMAI, 4(4), 40-44.	**Forward snowballing** Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In International Conference of Design, User Experience, and Usability (pp. 383-392). Springer, Cham.	2 & 3
Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017b). Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). Ijimai, 4(6), 103-108.	**Backward snowballing** Valadi, S., & Broneske, (2020). Analysis of Current Usability and User Experience Questionnaires and Creating an Optimized Usability Questionnaire.	3
Schrepp, M., & Thomaschewski, J. (2019). Handbook for the modular extension of the User Experience Questionnaire. All you need to know to apply the UEQ+ to create your own UX questionnaire. Google Scholar Google Scholar Cross Ref Cross Ref.	**Forward snowballing** Hinderks, A., Schrepp, M., Mayo, F. J. D., Escalona, M. J., & Thomaschewski, J. (2019). Developing a UX KPI based on the user experience questionnaire. Computer Standards & Interfaces, 65, 38-44.	2 & 3
Shah, S. G. S., & Robinson, I. (2008). Medical device technologies: who is the user?. International Journal of Healthcare Technology and Management, 9(2), 181-197.	**Backward snowballing** Shah, S. G. S., Robinson, I., & AlShawi, S. (2009). Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process.	-
Shah, S. G. S., Robinson, I., & AlShawi, S. (2009). Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process.	**Google scholar:** "user involvement medical devices"	-
Sharma, A., Malviya, R., Awasthi, R., & Sharma, P. K. (2020). Artificial Intelligence, Blockchain, and Internet of Medical Things: New Technologies in Detecting, Preventing, and Controlling of Emergent Diseases. In Advances in Multidisciplinary Medical Technologies–Engineering, Modeling and Findings (pp. 127-154). Springer, Cham.	**Google scholar:** "Artificial intelligence medical technologies"	-
Shih, P. Y., & Zheng, M. C. (2020). Usability Study of a Pre-anesthesia Evaluation App in an University Hospital: Before the Revision of User Interface. In International Conference on Human-Computer Interaction (pp. 480-487). Springer, Cham.	**Forward snowballing** Schrepp, M., Held, T., & Laugwitz, B. (2006). The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers, 18(5), 1055-1069.	3

Sittig, D. F., Kuperman, G. J., & Fiskio, J. (1999). Evaluating physician satisfaction regarding user interactions with an electronic medical record system. In Proceedings of the AMIA Symposium (p. 400). American Medical Informatics Association.	**Backward snowballing** Ellsworth, M. A., Dziadzko, M., O'Horo, J. C., Farrell, A. M., Zhang, J., & Herasevich, V. (2017). An appraisal of published usability evaluations of electronic health records via systematic review. Journal of the American Medical Informatics Association, 24(1), 218-226.	3
Sousa, V. E., & Lopez, K. D. (2017). Towards usable e-health: A systematic review of usability questionnaires. Applied clinical informatics, 8(2), 470.	**Forward snowballing** Lewis JR. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Int J Hum-Comput Int. 1995;7(01):57–78.	2 & 3
Sproll, S., Peissner, M., & Sturm, C. (2010). From product concept to user experience: exploring UX potentials at early product stages. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (pp. 473-482).	**Backward snowballing** Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	3
Stauder, R., Belagiannis, V., Schwarz, L. A., Bigdelou, A., Söhngen, E., Ilic, S., & Navab, N. (2012). A user-centered and workflow-aware unified display for the operating room. In MICCAI Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI), Nice, France.	**Google scholar:** "Attrakdiff MRI"	3
Stola, K. (2018). User experience and design thinking as a global trend in healthcare. Journal of Medical Science, 87(1), 28-33.e	**Google scholar:** "user experience healthcare"	-
Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. International journal of psychology, 42(4), 253-264.	**Backward snowballing** *"Minge, M., Thüring, M., Wagner, I., & Kuhr, C. V. (2017). The meCUE questionnaire: a modular tool for measuring user experience. In Advances in Ergonomics Modeling, Usability & Special Populations (pp. 115-128). Springer, Cham.*	1 & 2
Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (pp. 115-122).	**Backward snowballing from:** McLellan, S., Muddimer, A., & Peres, S. C. (2012). The effect of experience on System Usability Scale ratings. Journal of usability studies, 7(2), 56-67.	3
Tullis, T. S., & Stetson, J. N. (2004, June). A comparison of questionnaires for assessing website usability. In Usability professional association conference (Vol. 1, pp. 1-12).	Backwards snowballing Brooke (2013)	3
von Wilamowitz-Moellendorff, M., Hassenzahl, M., & Platz, A. (2006). Dynamics of user	**Forward snowballing** Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research	3

experience: How the perceived quality of mobile phones changes over time. In User Experience-Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer Interaction (pp. 74-78).	agenda. Behaviour & information technology, 25(2), 91-97.	
Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008). Towards practical user experience evaluation methods. Meaningful measures: Valid useful user experience measurement (VUUM), 19-22.	**Backward snowballing** Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	3
Valadi, S., & Broneske, (2020). Analysis of Current Usability and User Experience Questionnaires and Creating an Optimized Usability Questionnaire.	**Forward snowballing** Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14).	3
Velimirović, D., Velimirović, M., & Stanković, R. (2011). Role and importance of key performance indicators measurement. Serbian Journal of Management, 6(1), 63-72.	**Google scholar:** "Key performance indicators kpi"	3
Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries (pp. 521-530).	**Backward snowballing** Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	3
Walsh, T., Varsaluoma, J., Kujala, S., Nurkka, P., Petrie, H., & Power, C. (2014). Axe UX: exploring long-term user experience with iScale and AttrakDiff. In Proceedings of the 18th international academic mindtrek conference: Media business, management, content & services (pp. 32-39).	**Backward snowballing** Chandran, S., Al-Sa'di, A., & Ahmad, E. (2020). Exploring User Centered Design in Healthcare: A Literature Review. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.	2 & 3
Wigelius, H., & Väätäjä, H. (2009). Dimensions of context affecting user experience in mobile work. In IFIP Conference on Human-Computer Interaction (pp. 604-617). Springer, Berlin, Heidelberg.	**Backward snowballing** Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	3
Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering, 4(1), 1421006.	Web of science search string 1	1
Zhou, L., Bao, J., & Parmanto, B. (2017). Systematic review protocol to assess the effectiveness of usability questionnaires in	Web of science search string 2	2

mhealth app studies. JMIR research protocols, 6(8), e151.		

Appendix 3. Informal meetings field notes

<Removed for confidentiality>

Appendix 4. Interview guides and PowerPoint slides

Interview guide A - identification of needs and requirements for measuring the user experience (English)

Phase 01 – Introduction

- *Tu Eindhoven*
- *Developing a tool to quantitatively measuring the user experience of MRI innovations*
- *Goal of the interview:*

Start recording (check microphone)

Phase 02 – General questions

1. What is the role of your department in the innovation process of MRI systems?
    a. What is your role in this process?
2. Is your department involved in measuring the end-user experience?
    a. If so, how is your department involved?
3. Is it useful for your department to measure the user experience of MRI systems?
    a. If yes: For what purpose would it be useful?
    b. If not: Why is it not useful?
4. Would it be useful to measure the end-user experience quantitatively (with a score)?
    a. Why?
    b. Why not?
5. Which aspects or parameters of the end-user experience are useful for your department to measure? (e.g. efficiency)

Phase 03 – Discussion of UEQ tool

6. Show the dimensions that are available in the UEQ
7. Which of these dimensions do you think that are useful to measure the end-user experience of MRI systems?

    a. Are there any that you would like to add in terms of dimensions?
8. Show the questionnaire
9. Do you think such a questionnaire would be useful for measuring the end-user experience?
    a. Why?
    b. Why not?
10. Are there any limitations of the UEQ?
11. When do you think such a tool can be used in the innovation process?

Phase 03 – End

12. Do you have any other remarks, comments or suggestions for this process?
13. Thank for participating and ask for possible follow-up interview when a more detailed solution is created.

Interview guide A - Identification of needs and requirements for measuring the user experience (Dutch)

Phase 01 – Introduction

- *Tu Eindhoven*
- *Developing a tool to quantitatively measuring the user experience of MRI innovations*
- *Goal of the interview:*

Ask consent form

Ask for recording (check microphone) (sommige vragen herhaling)

Phase 02 – General questions

1. Wat is de rol van jouw afdeling in het innovatieproces van MRI systemen?
    a. Wat is jouw rol in dit proces?
2. Is jouw afdeling betrokken bij het meten van de eindgebruikers ervaring (end-user experience)?
    a. Hoe is je afdeling hier bij betrokken?
    b. Op welke momenten in het innovatieproces?
3. Is het nuttig voor jouw afdeling om de end-user experience van MRI system te meten?
    a. Zo ja, waarom zou het nuttig zijn?
    b. Zo nee, waarom zou het niet nuttig zijn?
4. Zou het nuttig zijn om de end-user experience met een score (kwantitatief) te meten?
    a. Waarom?
    b. Waarom niet?
5. Welke aspecten of parameters van de end-user experience zijn nuttig om te meten voor jouw afdeling? (bijvoorbeeld efficiency)

Phase 03 – Discussion of UEQ tool

6. Show the dimensions that are available in the UEQ
7. Welke van deze dimensies denk je dat nuttig zijn om te meten om de end-user experience te bepalen?

    a. Zijn er dimensies die je hier nog mist voor het meten van de end-user experience?

8. Show the questionnaire
9. Denk je dat deze questionnaire nuttig zou zijn voor het meten van de end-user experience?
    a. Waarom?
    b. Waarom niet?

10. Wat zijn de nadelen/ limitaties van de UEQ?
11. Wanneer denk je dat de UEQ gebruikt zou kunnen worden in het innovatieproces?

Phase 03 – End

12. Heb je andere opmerkingen of suggesties voor dit proces?
13. Bedankt voor de deelname en zou je eventueel beschikbaar zijn om verder in het proces beschikbaar te zijn voor een follow-up interview?

Interview guide B Innovation process of the case company (English)

Phase 01 – Introduction:

Start recording (make sure that microphone is recording)

Phase 02 – Process questions:

1. How does the current development OR innovation process of The case company of MRI systems look like in terms of phases or steps?
    a. Are there any documents that clearly describe this process?
    b. Does the PDLM process best describe the innovation process?

2. How is the perspective of the end-user incorporated in this innovation process?
    a. Are there any documents that clearly describe this process?
    b. Does the handbook for usability engineering document best describe the innovation process?

3. What is the role of clinical application specialists in this innovation process?
    a. What other departments are involved in the innovation process?

4. How is decided which designs or products should be developed? (which are the best options?)

5. How is the SUS score currently incorporated in the innovation process?

    a. At which moments is the SUS used? (early/later stages?)
    b. How often is the SUS repeated?
    c. How many people fill in the SUS?
    d. Is the SUS used with clinical application specialists?
    e. How are the results communicated?

Phase 03 – Needs and requirements questions: (optional)

1. Is it useful for your department to measure the user experience of MRI systems?
    a. If yes: For what purpose would it be useful?
    b. If not: Why is it not useful?
2. Would it be useful to measure the end-user experience quantitatively (with a score)?
    a. Why?
    b. Why not?
3. Which aspects or parameters of the end-user experience are useful for your department to measure? (e.g. efficiency)

Phase 03 – Discussion of UEQ tool

4. Show the dimensions that are available in the UEQ
5. Which of these dimensions do you think that are useful to measure the end-user experience of MRI systems?

    a. Are there any that you would like to add in terms of dimensions?
6. Show the questionnaire
7. Do you think such a questionnaire would be useful for measuring the end-user experience?
    a. Why?
    b. Why not?
8. Are there any limitations of the UEQ?
9. When do you think such a tool can be used in the innovation process?

Phase 03 – End

10. Do you have any other remarks, comments or suggestions for this process?
11. Thank for participating and ask for possible follow-up interview when a more detailed solution is created.
12.

Interview guide B Innovation process of the case company (Dutch)

Introductie

Vraag om opnemen

Opnemen (check microfoon)

Phase 02 – Process questions :

1. Hoe ziet het innovatie proces voor MRI systemen van The case company er nu uit in grote lijnen en fasen ?
    a. Zijn er documenten die dit duidelijk beschrijven?
    b. Beschrijft het PDLM proces dit goed?

2. Hoe wordt het perspectief van de end-user momenteel meegenomen in het innovatie proces?
    a. Zijn er documenten die dit proces duidelijk beschrijven?
    b. Beschrijft het handbook for usability engineering document dit proces goed?
    c. In welke stap begint FOK testing?

3. Wat is de rol van klinische applicatie specialisten in het meten van de end-user experience?
    a. Welke andere afdelingen binnen The case company zijn betrokken in het meten van de end-user experience?

4. Hoe wordt er momenteel gekozen welke solution direction ontwikkeld moet worden?
    a. Wie bepaalt dit?

5. Hoe wordt de SUS score momenteel meegenomen in het innovatie proces?

    a. Op welke momenten wordt de SUS gebruikt in het innovatie proces? (early development or later stages?
    b. Hoe vaak wordt de SUS herhaald?
    c. Hoeveel mensen vullen de SUS dan ongeveer in?
    d. Wordt de SUS gebruikt met klinische applicatie specialisten?
    e. Hoe worden de resultaten van de SUS gecommuniceerd?

Phase 03 – Needs and requirements questions: (optional)

14. Is het nuttig voor jouw afdeling om the end-user experience van MRI system te meten?
    a. Zo ja, waarom zou het nuttig zijn?
    b. Zo nee, waarom zou het niet nuttig zijn?
15. Zou het nuttig zijn om de end-user experience met een score (kwantitatief) te meten?
    a. Waarom?
    b. Waarom niet?
16. Welke aspecten of parameters van de end-user experience zijn nuttig om te meten voor jouw afdeling? (bijvoorbeeld efficiency)

Phase 03 – Discussion of UEQ tool

17. Show the dimensions that are available in the UEQ
18. Welke van deze dimensies denk je dat nuttig zijn om te meten om de end-user experience te bepalen?

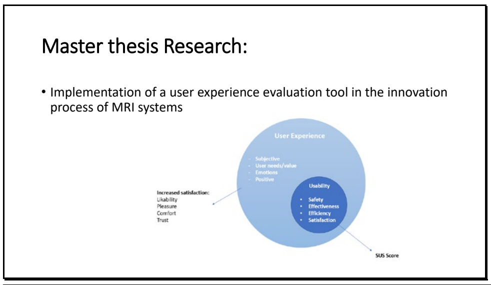
    a. Zijn er dimensies die je hier nog mist voor het meten van de end-user experience?
19. Show the questionnaire

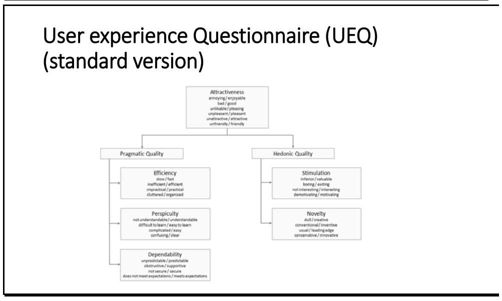
20. Denk je dat deze questionnaire nuttig zou zijn voor het meten van de end-user experience?
    a. Why?
    b. Why not?
21. Wat zijn de nadelen van de UEQ?
22. Wanneer denk je dat de UEQ gebruikt zou kunnen worden in het innovatieproces?

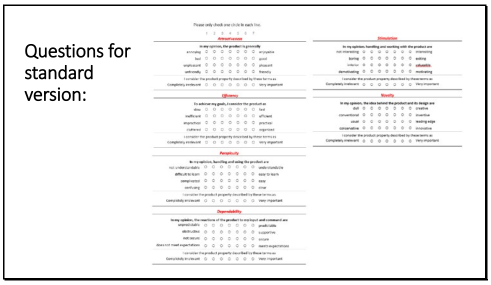
Phase 03 – End

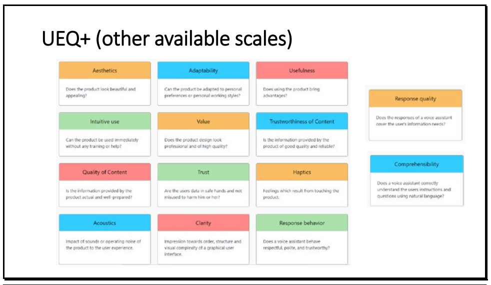
23. Heb je andere opmerkingen of suggesties voor dit proces?
24. Bedankt voor de deelname en zou je eventueel beschikbaar zijn om verder in het proces beschikbaar te zijn voor een follow-up interview?

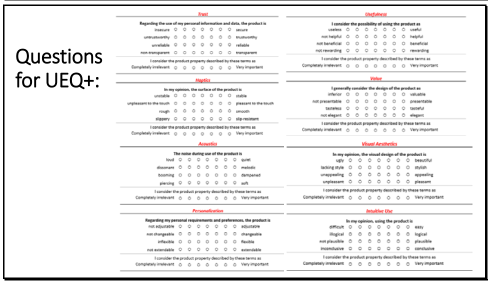
Powerpoint slides Interview Requirements and needs

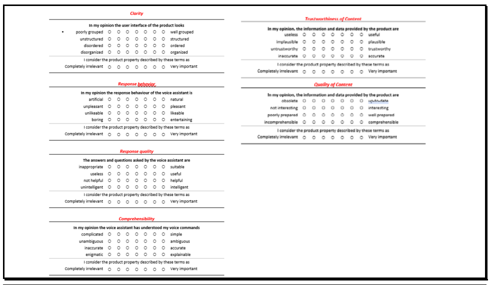


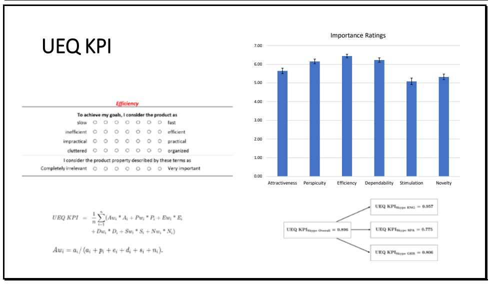


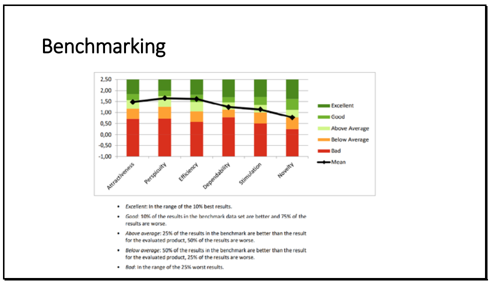



UEQ+ (other available scales)
Aesthetics
Adaptability
Usefulness
Intuitive use
Value
Trustworthiness of Content
Quality of Content
Trust
Haptics
Acoustics
Clarity
Response behavior
Response quality
Comprehensibility

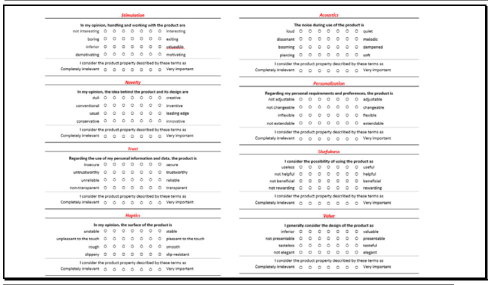

Questions for UEQ+:

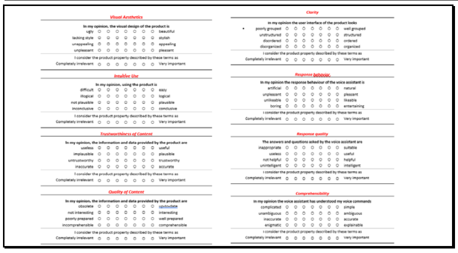


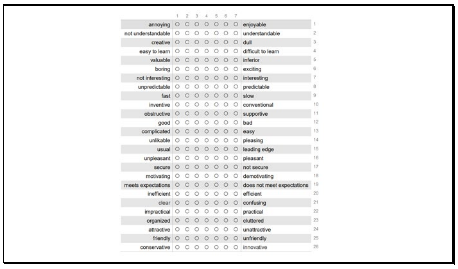

UEQ KPI
Importance Ratings
Attractiveness
Perspicuity
Efficiency
Dependability
Stimulation
Novelty


Benchmarking
Excellent
Good
Above Average
Below Average
Bad
Mean

All available scales:







	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Consent form:

Subject information for participation in scientific research

User experience measurement approach of MRI innovations

Official title: Development of a quantitative user experience measurement approach for MRI innovations

Introduction

Dear Sir/Madam,

You are asked to take part in a scientific study.

Participation is voluntary. Participation requires your written consent. Before you decide whether you want to participate in this study, you will be given an explanation about what the study involves. Please read this information carefully and ask the investigator for an explanation if you have any questions. You may also discuss it with your partner, friends or family.

1. General information

Initiator	
Initiated by Eindhoven University of Technology in collaboration with the case company	This study has been designed by the case company and is being carried out by Bas de Groot at the Eindhoven University of Technology.

2. Purpose of the study

The purpose of this study is to identify how a user experience measurement approach can be integrated in the innovation process of the case company. Requirements and needs for this approach will be discussed with employees of the case company to develop an approach that fits the context of the case company.

3. What participation involves

During the study, the following will happen:

- data is collected about the requirements and needs for a user experience measurement approach

- you will participate in an interview that takes approximately one hour

- an audio recording will be made

4.If you do not want to participate or you want to stop participating in the study

It is up to you to decide whether or not to participate in the study. Participation is voluntary.

If you do participate in the study, you can always change your mind and decide to stop, at any time during the study. You do not have to say why you are stopping, but you do need to tell the investigator immediately.

The data collected until that time will still be used for the study.

If there is any new information about the study that is important for you, the investigator will let you know. You will then be asked whether you still want to continue your participation.

5. End of the study

Your participation in the study stops when

- you choose to stop
- the end of the entire study has been reached
- the investigator considers it best for you to stop
- The case company, the government or the Ethical Review Board decides to stop the study.

The study is concluded once all the participants have completed the study.

6. Usage and storage of your data

Your personal data will be collected, used and stored for this study. This concerns data such as your name. The collection, use and storage of your data is required to answer the questions asked in this study and to publish the results. We ask your permission for the use of your data

Confidentiality of your data To protect your privacy, your data will be given a code. Your name and other information that can directly identify you, will be omitted. Data can only be traced back to you with the encryption key. The encryption key remains safely stored in the local research institute. The data that is sent to the Eindhoven University of Technology will only contain the code, not your name or other data with which you can be identified. The data cannot be traced back to you in reports and publications about the study.

Access to your data for verification

Some people can access all your data at the research location. Including the data without a code. This is necessary to check whether the study is being conducted in a good and reliable manner. Persons who have access to your data for review are: the committee that monitors the safety of the study and a supervisor from the Eindhoven University of Technology. They will keep your data confidential. We ask you to consent to this access.

Withdrawing consent

You can withdraw your consent to the use of your personal data at any time. The study data collected until the moment you withdraw your consent will still be used in the study.

More information about your rights when processing data

For general information about your rights when processing your personal data, you can consult the website of the Dutch Data Protection Authority.

If you have questions about your rights, please contact the person responsible for the processing of your personal data. For this study, that is:

Bas de Groot – Eindhoven University of Technology. See Appendix A for contact details.

If you have questions or complaints about the processing of your personal data, we advise you to first contact the research location. You can also contact the Data Protection Officer of the Eindhoven University of Technology or the Dutch Data Protection Authority.

7. Any questions?

If you have any questions, please contact Bas de Groot.

If you have any complaints about the study, you can discuss this with the investigator. If you prefer not to do this, you may contact Rens Kruisbrink. All the relevant details can be found in **Appendix A**: Contact details.

8. Signing the consent form

When you have had sufficient time for reflection, you will be asked to decide on participation in this study. If you give permission, we will ask you to confirm this in writing on the appended consent form. By your written permission you indicate that you have understood the information and consent to participation in the study. The signature sheet is kept by the investigator. Both the Investigator and yourself receive a signed version of this consent form.

Thank you for your attention.

9. Appendices to this information

A. Contact details

B. Informed Consent Form

Appendix A: contact details for Bas de Groot of the Eindhoven University of Technology

Bas de Groot

Mail: (b.d.groot.1@student.tue.nl)

Phone: 0613284316

Alternative contact:

Rens Kruisbrink

Mail: *confidential*

Complaints: Bas de Groot

Data Protection Officer of the institution: dataprotectionofficer@tue.nl.

For more information about your rights: Bas de Groot

Appendix B: Subject Consent Form

User experience measurement approach of MRI innovations

- I have read the subject information form. I was also able to ask questions. My questions have been answered to my satisfaction. I had enough time to decide whether to participate.
- I know that participation is voluntary. I know that I may decide at any time not to participate after all or to withdraw from the study. I do not need to give a reason for this.
- I give permission for the collection and use of my data to answer the research question in this study.
- I know that some people may have access to all my data to verify the study. These people are listed in this information sheet. I consent to the inspection by them.

- I □ **do**

  □ **do not**
  consent to being contacted again after this study for a follow-up study.

- I want to participate in this study.

Name of study subject:

Signature: Date: __ / __ / __

I hereby declare that I have fully informed this study subject about this study.

If information comes to light during the course of the study that could affect the study subject's consent, I will inform him/her of this in a timely fashion.

Name of investigator (or his/her representative):

Signature: Date: __ / __ / __

The study subject will receive the full information sheet, together with an original of the signed consent form.

Approval Ethical Review Board

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Date
May 5, 2021

Reference
ERB2021IEIS6

Ethical Review Board TU/e

T +31 (0)40 247 6259
ethics@tue.nl

intranet.tue.nl/ethics

Ethical review research proposal

Dear B. de Groot,

It is a pleasure to inform you that the Ethical Review Board (ERB) has discussed and approved your application "Philips MR Applications master thesis".

The Board wants to draw your attention to the terms and conditions in the appendix.

Success with your research!

Sincerely,

Prof. dr. ir. A.W.M. Meijers
Chair Ethical Review Board TU/e

Enclosures
1

The ERB retains the right to revise its decision regarding the implementation and the WMO[1]/WMH[2] status of any research study in response to changing regulations, research activities, or other unforeseen circumstances that are relevant to reviewing any such study. The ERB shall notify the principal researcher of its revised decision and of the reasons for having revised its decision.

[1]WMO: Law on Medical Scientific Research involving Human Beings (in Dutch: Wet medisch-wetenschappelijk onderzoek met mensen)

[2]WMH: Medical Device Directive (in Dutch: Wet op de medische hulpmiddelen)

TU/e

APPENDIX 1

Terms and conditions

Amendments
When considerable amendments are made to the design of the study or educational activity, or when the time period between ERB approval and start of the study is longer than one year, please consult the ERB.

Privacy and research data management
The ERB would like to point out that collecting, handling and storing personal information is subject to the General Data Protection Regulation. Please visit TU/e intranet for the latest information and regulations on www.tue.nl/rdm

Appendix 6 Coding

Start list codes:

Name
Innovation process
Application specialists
Design
Usability
Marketing
Planning Development Launch and Maintentance (PDLM)
Early
Overall process
Quality and regulatory
SUS (System usability score)
Requirements and needs
Benchmarking
Communication
NPS (Net promoter score)
Quantification
Subjective and Objective measures
UEQ (User Experience Questionnaire)
Number of parameters
Limitations
Important Parameters
Missing scales
Evaluating usability or User Experience

Final codes:

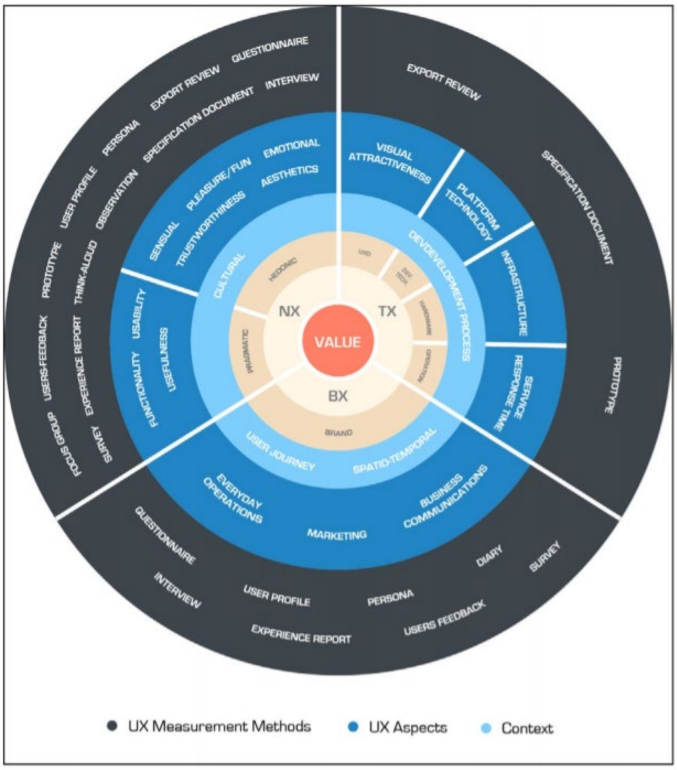
Name	Files	References
Innovation process		
Advanced development	3	8
Department		
Application specialists (Proxies)	4	13
Design	3	7
Usability	1	2
Marketing	2	17
Quality and regulatory	1	4
PDLM	0	0
Early phase of the innovation process	2	4
Develop user and business needs	3	13
Later phases of the innovation process		
Integration	4	22
Monitor launch	5	24
First of a Kind (FOK)	5	20
Validation	4	9
Testathons	1	1
Overall process	5	16
Stage - Gate	3	6

SUS (System usability score)	6	22
Content of tool		
Benchmarking	6	12
Standardization	1	1
Trend over time	2	2
Communication and interpretation	7	14
NPS (Net promoter score)	1	1
Quantification	6	19
Scalability	1	1
Subjective and Objective measures	2	7
UEQ (User Experience Questionnaire)	3	5
Applicable innovation stage	4	4
Business ownership	1	1
Importance (weighing factor)	4	4
Language	5	14
Length and time	3	7
Limitations	5	16
Sample size	1	1
Parameters	8	35
Indirect or direct evaluation	1	3
Missing scales	4	8
Key Performance Indicators (KPI)	3	5
Usability or UX	4	6

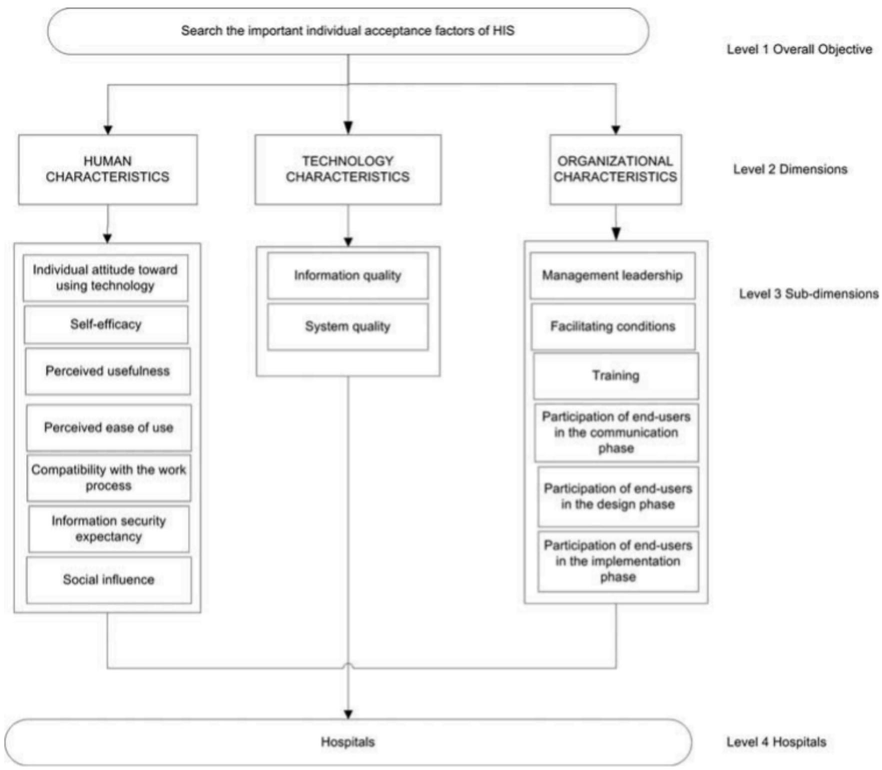
Appendix 7. Transcriptions

<Removed for confidentiality>

Appendix 8. Frameworks of the user experience



User experience framework (Zarour, & Alharbi, 2017)



Framework for user acceptance for medical technologies (Handayani et al., 2018)

Attributes	Number of coding references	Number of sources
Context of use	13	7
Easiness	1	1
Effectiveness	4	4
Efficiency	3	2
Efficient to use	1	1
Error prevention	2	2
Functionally correct	2	1
Helpfulness	2	2
Image Quality	4	3
Information communicativeness	14	5
Learnability	5	3
Operability	4	3
Productivity	4	3
Safety	3	3
Speed of performance	4	4
Standardization	6	3
System functions	8	3
System performance	2	2
Training	7	3
Useableness	2	1
Usefulness	6	4
User satisfaction	1	1

Usability attributes for CT (Aldoihi & Hammami, 2020)

Appendix 9. Usability tools, other UX tools and related tools

3.3.1 Usability tools

There are many different tools in academic literature that either focus on the user experience or a subset of the user experience (Lewis, 2018). This section will focus on tools for measuring usability, which is a subset of the user experience. This section will only discuss the most used and standardized usability tools, which are shown in Table 3.

There have been several academic literature reviews that identified the most used tools in contexts such as general software systems and Mobile-Health applications (Assila & Ezzedine, 2016; Sousa & Lopez, 2017; Zhou, Bao & Parmanto, 2017). The researchers mention different tools as the most used tools, which again shows that it is a fragmented field, and the use of tools differs in different contexts.

Lewis (2018) mentioned that the most used standardized tool for assessing the perceived usability is the System Usability Scale (SUS), followed by the Computer System Usability Questionnaire (CSUQ). In the context of electronic health records, the SUS and the Questionnaire for User Interface Satisfaction (QUIS) are the most used standardized tools (Ellsworth et al., 2017).

Recently, the Usability Metric for User Experience (UMUX) and its shorter version the UMUX-LITE have gotten increased attention (Borsci Buckle & Walne, 2020; Lewis, 2018). Furthermore, the UMUX and UMUX-LITE have previously been discussed in relation to new healthcare technologies.

If developers of products or systems merely want to evaluate the usability of their products, they can choose one of the tools from the table below. As the focus of this research is on tools that can measure the full user experience, the usability tools will not be discussed in detail here. The usability tools that were mentioned are described **in more detail** in

Usability tools	Description
System usability scale (SUS)	The systems usability scale (SUS) is a tool developed by John Brooke to measure the subjective usability of products and systems (Brooke, 1996; Brooke, 2013). The score is referred to as "quick and dirty" as it reduces usability to a single score based on just 10 questions on a 5-point Likert scale.
Computer System Usability Questionnaire (CSUQ)	*The CSUQ consists of 16 items on a 7-point Likert scale. Together the items produce four scores, the overall score (all the items), system usefulness (items 1-6), information quality (items 7-12) and interface quality (items 13-15) (Lewis, 2018).*

UMUX / UMUX-LITE	*The usability metric for user experience (UMUX) is a tool introduced by Kraig Finstad with the aim to reduce the length of the SUS score (Finstad, 2010). The UMUX-LITE is a shorter version of the UMUX.*
Questionnaire for User Interface Satisfaction (QUIS)	*The Questionnaire for User Interface Satisfaction (QUIS) focuses on being comprehensive (Chin, Diehl & Norman, 1988). The QUIS was specifically made to assess user interfaces, which does not allow the tool to be used for other systems. The tool measures aspects on a 9-point Likert scale (Chin et al., 1988).*

Standardized usability tools

System Usability Scale (SUS)

The systems usability scale (SUS) is a tool developed by John Brooke to measure the subjective usability of products and systems (Brooke, 1996; Brooke, 2013). The score is referred to as "quick and dirty" as it reduces usability to a single score based on just 10 questions on a 5-point Likert scale. Following a simple formula, these 10 questions are scored on a 0 to 100 scale for easy interpretation. As the SUS score is the most used usability score, there is a lot of benchmark data available to compare the score with other products (Sauro & Lewis, 2016). The full system usability scale can be seen in the figure below.

Brooke (1996; 2013) mentioned that important factors for evaluating the usability of a system are the context in which the system is used, the background and experience of the users and the tasks the users will perform with the system. In other words, the usability evaluation only provides the usability of that specific context (Brooke, 2013).

The shortness of the questionnaire is an added value in practice, as there is often no time to ask many questions about every detail of the product. As it only consists of 10 questions, the SUS score does not ask a lot of effort from the users and can therefore be repeated frequently to track changes throughout the innovation process. At the same time, reducing the number of questions for usability also reduces the amount of information developers can use to further develop their products to the needs of their users.

The tool was constructed to measure subjective usability by using statements that best allow to distinguish between usable and unusable systems (Brooke, 2013). Further research has shown that, in contrary to the initial purpose, the SUS measures both learnability (with items 4 and 10 of the SUS scale) and usability (the other 8 items), which indicates that this information can be retrieved separately to have more specific knowledge (Lewis & Sauro, 2009). Critics of the SUS mention that the word cumbersome in statement 8 of the SUS score is not always understood by users but Lewis and Sauro (2009) have shown that the word can be replaced by the word awkward without significantly impacting the results.

Brooke (1996) also mentioned that the respondents should immediately respond to the questions after using the system that is being evaluated and if they feel that they cannot respond to the question, they should mark the center point of the scale.

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

The System Usability Scale

For the items 1,3,5,7 and 9 the score is the scale position (1 to 5) minus 1. For the items 2,4,6,8 and 10, the score is 5 minus the scale position (1 to 5). The sum of all the scores is then multiplied by 2.5 to obtain the overall SUS score on a range of 0 to 100 (Brooke, 1996).

As the SUS score is calculated on a range from 0 to 100, it might seem that the score gives a percentage of the subjective usability for users with the product or system. However, this would be a too optimistic interpretation as it is unlikely for SUS scores to be very low (Sauro & Lewis, 2016). For this reason, the curved grading scale was developed, which will be discussed in Chapter 4.

Computer System Usability Questionnaire (CSUQ)

The Computer System Usability Questionnaire (CSUQ) is a variant of the Post-Study System Usability Questionnaire (PSSUQ) which was designed to assess users' perceived satisfaction with computer systems or applications (Lewis, 1995). The tools are identical apart from the choice of words in some sentences of the CSUQ to allow the tool to be used in a real-life setting, where the PSSUQ is only applicable in lab research (Sauro & Lewis, 2016). For this reason, only the CSUQ will be discussed here which is shown in the figure below.

The CSUQ consists of 16 items on a 7-point Likert scale. Together the items produce four scores, the overall score (all the items), system usefulness (items 1-6), information quality (items 7-12) and interface quality (items 13-15) (Lewis, 2018). Each score can be calculated by averaging their respective items which results in a score between 1 and 7.

The CSUQ has previously shown to be sensitive to several independent variables including the number of years of experience with the system (Berkman & Karahoca, 2016). The sensitivity of the results might impact the implications for practitioners to use the tool.

The main downside of this tool is that the tool was made to assess computer interface systems. This limits the tools in the use for other products. If developers want to use one standardized tool to assess the usability of different types of products this tool might not be sufficient. As this tool only allows to test computer interface systems, developers might have to use other tools next to this tool.

The Computer System Usability Questionnaire Version 3		Strongly agree						Strongly disagree	
		1	2	3	4	5	6	7	NA
1	Overall, I am satisfied with how easy it is to use this system.	○	○	○	○	○	○	○	○
2	It is simple to use this system.	○	○	○	○	○	○	○	○
3	I am able to complete my work quickly using this system.	○	○	○	○	○	○	○	○
4	I feel comfortable using this system.	○	○	○	○	○	○	○	○
5	It was easy to learn to use this system.	○	○	○	○	○	○	○	○
6	I belive I became productive quickly using this system.	○	○	○	○	○	○	○	○
7	The system gives error messages that clearly tell me how to fix problems.	○	○	○	○	○	○	○	○
8	Whenever I make a mistake using the system, I recover easily and quickly.	○	○	○	○	○	○	○	○
9	The information (such as online help, on-screen messages and other documentation) provided with this system is clear.	○	○	○	○	○	○	○	○
10	It is easy to find the information I needed.	○	○	○	○	○	○	○	○
11	The information provided with the system is effective in helping me complete my work.	○	○	○	○	○	○	○	○
12	The organization of information on the system screens is clear.	○	○	○	○	○	○	○	○
13	The interface* of this system is pleasant.	○	○	○	○	○	○	○	○
14	I like using the interface of this system.	○	○	○	○	○	○	○	○
15	This system has all the functions and capabilities I expect it to have.	○	○	○	○	○	○	○	○
16	Overall, I am satisfied with this system.	○	○	○	○	○	○	○	○

**The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens* (*including thier graphics and language*).

CSUQ (computer system usability questionnaire) (Sauro & Lewis, 2016)

Usability metric for user experience (UMUX)

The usability metric for user experience (UMUX) is a tool introduced by Kraig Finstad with the aim to reduce the length of the SUS score (Finstad, 2010). The aim is the same as the SUS score, to measure the subjective usability of systems. The difference is that it only consists of four items on a 7-point Likert scale instead of 10 for the SUS score, therefore allowing for quicker evaluations. Like the SUS score, the items are recoded and calculated to a single score on a range of 0 to 100. The four items can be seen in the figure below and the full is scale is added to the appendix .

The items were made with the aim to fit the definition of the ISO standard of usability, which was discussed before and consists of efficiency, effectiveness and satisfaction. One item was added to measure the overall usability. The links between the items in the tool and the aspects of usability can be seen in the figure below.

As discussed before, usability is only a subset of the user experience. The aim of UMUX was to create a shorter scale for the usability as other elements also need to be tested in a broader user experience measurement with users and 10 items for only usability was experienced to be too much effort combined with other questionnaires (Finstad, 2010).

Usability component	Candidate UMUX item
Effectiveness	[This system's] capabilities meet my requirements.
Satisfaction	Using [this system] is a frustrating experience.
Overall	[This system] is easy to use.
Efficiency	I have to spend too much time correcting things with [this system].

UMUX-tool

The UMUX-tool has been criticized by several studies regarding its validity, reliability and sensitivity (Bosley, 2013; Cairns, 2013; Lewis, 2013). Berkman and Karahoca (2016) did show that the criticism was based on the results of the initial study by Finstad (2010) and when re-evaluating the scale, they concluded that there is evidence for its reliability, sensitivity and concurrent validity. Berkman and Karahoca (2016) also mentioned that it was difficult to evaluate scores that were close to each other and advised using an additional tool such as the SUS or CSUQ as they are more sensitive.

UMUX-LITE

The UMUX-LITE is a shorter version of the UMUX that only consists of two items on a 7-point Likert scale to allow for even quicker subjective usability evaluations of systems (Lewis, Utesch & Maher, 2013). Again, the items are recoded and transformed to obtain a range between 0 and 100. Specifically, 1 is subtracted from each item and multiplied by 100/12.

A regression formula was made so the score can be used for comparison to the SUS score:

$$\text{UMUX-LITE} = .65(\text{UMUX}_{(1,3)}) + 22.9$$

The $\text{UMUX}_{(1,3)}$ in the formula refers to the items 1 of 3 of the UMUX questionnaire, which means that this questionnaire uses two of the same questions from the UMUX. The items 1 and 3 consists of the previously discussed parameter effectiveness and the overall usability, as can be seen in the figure above.

Although the results of the tool are promising, the authors mention that they do not recommend using it independent of the SUS as there is not a lot of empirical evidence yet (Lewis, Utesch & Maher, 2013). Recently, the tool has been tested in the context of new healthcare technologies, specifically point of care tools, by Borsci, Buckle and Walne (2020). They found that there is a significant positive correlation between UMUX-LITE and the net promoter score (NPS) in the healthcare context. This means that the higher the satisfaction score of UMUX is, the more people are willing to promote the product or service (Borsci et al., 2020). The authors do mention that the UMUX-LITE tool should not be used solely in the innovation process, but only to prepare for more formal usability testing (Borsci et al., 2020).

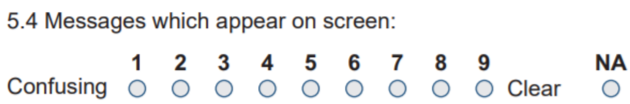
Borsci, Buckle and Walne (2020) also mention that more studies are needed to research the use of usability questionnaires in the context of healthcare technologies and that their research only discussed point of care tools, which cannot be generalized to all healthcare technologies.

Questionnaire for user interface satisfaction (QUIS)

Where the previously discussed tools focused on being fast and simple, the Questionnaire for User Interface Satisfaction (QUIS) focuses on being comprehensive (Chin, Diehl & Norman, 1988). The QUIS allows to derive more detailed information from the users but also requires more effort from the users. The tool focuses on satisfaction, but satisfaction is one of the aspects of usability according to the ISO standard discussed before. Like the CSUQ, the QUIS was specifically made to assess user interfaces, which does not allow the tool to be used for other systems. Other than the previously discussed tools, this tool measures aspects on a 9-point Likert scale (Chin et al., 1988). An additional option is added which allows users to select that the question is not applicable (NA). A sample question is included in the figure below.

The tool has seen several iterations and the later versions are not freely available, which is a downside of the questionnaire. More questions of QUIS 5.0 that were available are included in the appendix .

The latest version of the QUIS, is version 7.0 which includes a demographic questionnaire and the evaluation of 9 aspects: *screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, online tutorials, multimedia, teleconferencing, and software installation* (Moumane, Idri & Abran 2016). The QUIS 7.0 is available in two versions, short (41 items) and long (122) items (Sauro & Lewis, 2016). Compared to the previously discussed questionnaires, the short version of the QUIS is already substantially longer.



Sample question QUIS (Sauro & Lewis, 2016)

Previous use of usability tools in the medical context

Questionnaires that measure the usability of systems and products are widely adopted for technologies in the medical context. For example, in the context of telemedicine, standardized questionnaires for usability were the most used method to measure usability (Klaassen, van Beijnum & Hermens, 2016).

The QUIS and the CSUQ both focus on the user interface of systems and can therefore probably be used for the context of user interface systems of medical systems. The computer system usability questionnaire (CSUQ) has been used for telemedicine technologies and the QUIS has been used in the context of electronic health records, which showed that there were no issues with these tools in the medical context (Agnisarman et al., 2017; Sittig, Kuperman & Fiskio, 1999).

The UMUX-LITE questionnaire has been shown to be reliable for the medical context of diagnostic point-of-care tools (POCT) at each stage of the development process (Borsci et al., 2020). The UMUX-LITE should however be done in combination with other tools due to its limited scope and other application in the medical context still need to be validated (Borsci et al., 2020).

SUS has been identified as the most-used usability questionnaire in the medical context for electronic health records by a previous literature review (Ellsworth et al., 2017). Furthermore, the SUS has also been used for evaluating medical application systems as well as the specific context MRI (Asbach et al., 2008; Shih & Zheng, 2020). These cases show that the SUS score can be used for medical systems.

Of the usability tools, only the SUS was found to be used in the specific context of MRI. Most tools were used in the context of electronic health records, which differ from medical sytems as they are not involved in the direct diagnosis or treatment of patients. To determine the appropriateness of these tools for MRI systems, they need to be assessed in that specific context (Borsci et al., 2020).

All the previously discussed usability questionnaires seem to be validated and used in many different contexts, including medical contexts. None of the usability questionnaires seem to be completely inappropriate in the medical context, however there is limited information for the specific context of MRI. The next section will compare the usability tools in terms of appropriateness for MRI systems.

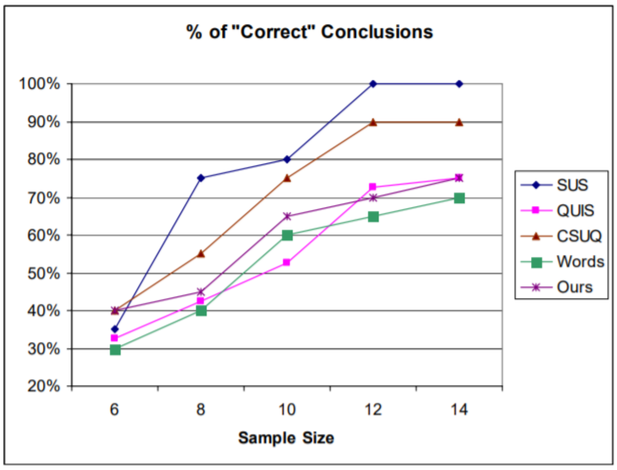
Adopting a usability tool for the context of developing MRI systems

The QUIS and the CSUQ were specifically designed for user interfaces and can therefore only be used for user interfaces of systems. This is a major disadvantage of these tools for developers that do not only want to evaluate interfaces. Developers that want to evaluate different products and systems would still have to adopt other tools, which makes the comparison between different products and systems more difficult and would require more effort. It would be easier for developers to only implement one tool. For this reason, developers of medical systems that not only want to test user-interfaces and do not want to use multiple different tools for measuring usability should consider other tools than the QUIS and CSUQ.

Although the SUS, UMUX and UMUX-LITE each provide reliable results, the SUS is the most reliable option out of the three (Borsci et al., 2015). The questionnaires UMUX and UMUX-LITE should only be used when there is not enough time to use SUS, as they are less reliable due to the reduction of the number of questions (Borsci et al., 2020). However, the SUS only consists of 10 statements thus there is no large time difference for the evaluation between the SUS and either the UMUX or UMUX-LITE, as these tools respectively consist of 4 and 2 statements.

SUS is the most widely used tool for assessing usability, but the QUIS and CSUQ have also shown to be reliable tools (Sousa & Lopez, 2017). A literature review showed that SUS was rated the highest in terms of quality but only provides one score and does not cover different aspects such as efficiency, memorability or errors that other usability questionnaires do measure (Sousa & Lopez, 2017). The choice of tool for developers can thus depend on the specific aspects that developers want to measure.

Tullis and Stetson (2004) compared the SUS, QUIS, CSUQ and two other methods that will not be discussed here as they are not standardized tools. In the comparison, two different websites were evaluated in terms of usability and each of the usability tools showed the same preferred website. This website was then assumed to be the preferred website with higher usability. Subsampling with each of the tools was done to assess how well the tools can predict the preferred website and thus the “correct” conclusion. Although it is arbitrary whether the first website can be seen as the correct conclusion, the results indicate that the SUS, followed by the QUIS were the best tools for predicting the preferred website, as can be seen in the figure below (Tullis & Stetson, 2004).



Comparison of different usability tools (Tullis & Stetson, 2004)

Concluding, different evaluations showed that the most reliable available questionnaire for measuring the subjective usability is the SUS questionnaire (Borsci et al., 2015; Sousa & Lopez, 2017; Tullis & Stetson, 2004). The SUS can be used for different types of systems other than user interfaces and has available data which allows for benchmarking the evaluations to other products (Sousa & Lopez, 2017). If there is very limited time with the user, the UMUX or UMUX-LITE can be used as they have also shown to be reliable (Borsci et al., 2015).

Developers could choose to design a new questionnaire like other researchers have done in the context of electronic health, but using a standardized and validated questionnaire has the advantage that the scores can be benchmarked to other products and do not require the effort of developing and validating a new tool (Klaassen, van Beijnum & Hermens, 2016; Sousa & Lopez, 2017). As Sousa and Lopez (2017) showed, the SUS does not provide a comprehensive view of the user experience and if a more comprehensive understanding is desired, other tools might be needed. Other tools that capture a broader range of the user experience will be discussed in the next section.

Attrakdiff

The attrakdiff is the oldest and most used user experience questionnaire of the three user experience questionnaires that will be discussed here (Díaz-Oreiro et al., 2019). Hassenzahl, Burmester and Koller (2003) developed the Attrakdiff questionnaire and focused on measuring both hedonic and pragmatic qualities. This is in contrary to the previously discussed usability questionnaires, which focus merely on the pragmatic quality of a system. The Attrakdiff was iteratively developed and the latest version, the Attrakdiff3, consists of 28 statements on a 7-step scale whose poles are opposite adjectives (for example “confusing-clear” or “good-bad”).

The Attrakdiff scale distinguishes between four aspects: *(1) pragmatic quality, (2) hedonic quality – identity, (3) hedonic quality – stimulation, and (4) attractiveness* (Walsh et al., 2014). Three sample questions can be seen in the figure below.

A mean score for the four different aspects is created which can either be compared with other systems or with previous iterations of the same system. The focus of this questionnaire is to derive specific knowledge on the pragmatic quality, hedonic quality and the attractiveness of a system and the aim is not to create a single score like the SUS score. Although the aim of the questionnaire is not to create a single score, the items can all be averaged to create a single score on a range between 1 and 7.

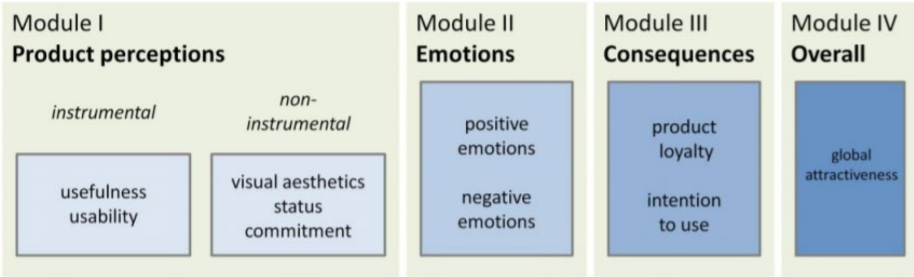
Sample questions Attrakdiff questionnaire

Modular evaluation of key components of user experience (meCUE)

The modular evaluation of key components of user experience (meCUE) is a questionnaire based on the Components model of User Experience (CUE) developed by Thüring and Mahlke (2007) which was discussed before (Minge, Thüring, Wagner & Kuhr, 2017).

The tool was developed to be more comprehensive in the user experience measurement as the developers mention that the previously discussed Attrakdiff and UEQ focus on pragmatic and hedonic qualities but not on the emotional component, both verbal and non-verbal (Minge et al., 2017).

The meCUE addresses this gap and includes the four separate modules: *(1) instrumental and non-instrumental product perceptions, (2) user emotions, (3) consequences of usage and (4) an overall judgment of the attractiveness,* which can be seen in the figure below (Minge et al., 2017). The authors mention that meCUE is particularly useful for comparing different product or design options and detecting changes of experience for long-term usage (Minge et al., 2017).



meCUE (Minge et al., 2017)

The questionnaire consists of 34 items on a 7-point Likert scale ranging from strongly disagree to strongly agree. There is one overall evaluation question that ranges from –5 to 5 with a scale interval of 0.5. Three sample questions can be seen in the figure below. The developers offer simple guidelines and an excel template that only requires the input data, which allows for quick and easy analysis (Minge et al., 2017).

	strongly disagree	disagree	somewhat disagree	neither agree nor disagree	somewhat agree	agree	strongly agree
The product is creatively designed.	O	O	O	O	O	O	O
The product would enhance my standing among peers.	O	O	O	O	O	O	O
I could not live without this product.	O	O	O	O	O	O	O

Sample questions meCUE

As the tool is modular, developers can choose to select the relevant modules for their specific case or remove some to save time (Minge et al., 2017). However, as can be seen in the figure above the statements do not seem to fit expert or medical systems. Statements such as *"I could not live without this product"* or *"The product is like a friend to me"* do not seem to fit the context of expert or medical systems. Lallemand & Koenig (2017) showed that there are issues with the face validity of the items of the meCUE and this can lead to frustration by participants as they feel they are wasting their time. This means that the tool can give difficulties in practice and cannot be used in every context.

Loyalty

Developers of products and systems want to assess what the loyalty of their customers is and how likely their customers are to recommend the product or system (Sauro & Lewis, 2016). A popular metric of customer loyalty is the Net Promoter Score (NPS) introduced by Fred Reichheld (2003) which uses a single likelihood to recommend question. Users rate the NPS question *"How likely is it that you would recommend this system to a friend or colleague?"* on an 11-step scale from 0 (not at all likely) to 10 (extremely likely). The respondents who select a 9 or 10 are referred to as *"promoters"* the respondents who select 0 through 6 are *"detractors"* and the others are *"passives".*

The percentage of detractors is subtracted from the number of promoters, making it a net promoter score.

The NPS is often used together with usability measurements as developers want to know what the impact of the change in usability or user experience is on the likelihood to recommend the product or system (Sauro and Lewis, 2016). The SUS score is often adopted by developers as the relation between SUS and NPS has been researched and when people rate a system or product higher than 82 for the SUS they also tend to be a promoter (Brooke, 2013). The relation with the NPS can help in the communication of results.

Other Tools

There are many different tools that either measure the usability, the user experience, or the consequences that result from it (Hussain et al., 2018). This section will briefly discuss alternatives to self-reported measures in the form of observational and physiological measures.

Observational measures

The previously discussed tools were all self-reported measures for evaluating the usability or user experience of a product or system. An alternative to the self-reported measures are the observational measures. Situations where the user is nonverbal or limited in verbal or cognitive ability, are situations where observational measures can help in understanding the usability or user experience of a product or system (Hussain et al., 2018).

Examples of observational measures are video-based facial expression analysis (FEA), emotion from the human voice and tracking user interactions by means of logging (Hussain et al., 2018). These observational measures can be useful for interpretation of the user experience but are unable to determine the psychological state of the user (Hussain et al., 2018).

Physiological measures

An alternative to the previously described methods to evaluate the user experience are physiological measures. With the use of biometric sensors such as eye trackers, electroencephalography and electromyography, the behavior of users can be detected and used as additional information for the evaluation of the user experience (Hussain et al., 2018). These physiological measures can be intrusive, which might make it more difficult to involve users in the user experience evaluation of developers.

UEQ:

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

UEQ-S:

obstructive	o o o o o o o	supportive	1
complicated	o o o o o o o	easy	2
inefficient	o o o o o o o	efficient	3
clear	o o o o o o o	confusing	4
boring	o o o o o o o	exiting	5
not interesting	o o o o o o o	interesting	6
conventional	o o o o o o o	inventive	7
usual	o o o o o o o	leading edge	8

UEQ items:

Scale	Original German items		English translation	
Attractiveness	unerfreulich	erfreulich	annoying	enjoyable
Perspicuity	unverständlich	verständlich	not under-standable	understandable
Novelty	kreativ	phantasielos	creative	dull
Perspicuity	leicht zu lernen	schwer zu lernen	easy to learn	difficult to learn
Stimulation	wertvoll	minderwertig	valuable	inferior
Stimulation	langweilig	spannend	boring	exiting
Stimulation	uninteressant	interessant	not interesting	interesting
Dependability	unberechenbar	voraussagbar	unpredictable	predictable
Efficiency	schnell	langsam	fast	slow
Novelty	originell	konventionell	inventive	conventional
Dependability	behindernd	unterstützend	obstructive	supportive
Attractiveness	gut	schlecht	good	bad
Perspicuity	kompliziert	einfach	complicated	easy
Attractiveness	abstoßend	anziehend	unlikable	pleasing
Novelty	herkömmlich	neuartig	usual	leading edge
Attractiveness	unangenehm	angenehm	unpleasant	pleasant
Dependability	sicher	unsicher	secure	not secure
Stimulation	aktivierend	einschläfernd	motivating	demotivating
Dependability	erwartungs-konform	nicht erwar-tungskonform	meets expecta-tions	does not meet expectations
Efficiency	ineffizient	effizient	inefficient	efficient
Perspicuity	übersichtlich	verwirrend	clear	confusing
Efficiency	unpragmatisch	pragmatisch	impractical	practical
Efficiency	aufgeräumt	überladen	organized	cluttered
Attractiveness	attraktiv	unattraktiv	attractive	unattractive
Attractiveness	sympathisch	unsympathisch	friendly	unfriendly
Novelty	konservativ	innovativ	conservative	innovative

Previous use of user experience evaluation tools in the medical context

Previous use of user experience tools in the context of medical systems

Usability questionnaires seem to be widely applied in the medical context but questionnaires for the whole user experience are less used in the medical context (Sousa & Lopez, 2017).

As discussed before, the most used standardized questionnaires for measuring the user experience are Attrakdiff, UEQ and meCUE, which are tools that have a wider scope than the **previously described** usability tools (Díaz-Oreiro et al.,2019). Some have argued that the pragmatic aspect of usability is the most important aspect in the context of medical systems, however there is no consensus on how much of the hedonic (non-pragmatic) aspect of the user experience evaluation is appropriate in the medical context (Bitkina et al., 2020). This section will discuss the appropriateness of the previously discussed user experience tools for the medical context.

Several researchers have used the Attrakdiff questionnaire as well as a short version of the Attrakdiff for evaluating the user experience of different systems (Fischer, Kuliga, Eisenberg & Amin, 2018**;** Stauder et al., 2012). A short version of the Attrakdiff questionnaire has previously been used in the context of assessing a medical application and also specifically for a medical imaging application (Shih & Zheng, 2020; Stauder et al., 2012). It seems therefore that the Attrakdiff questionnaire can be used for evaluating the user experience in the medical context.

With the construction of the UEQ, the developers found that the 'soft' (user experience) is equally important to 'hard' criteria (usability) (Laugwitz, Held & Schrepp, 2008). This would suggest that developers should take these soft criteria into account in their innovation process. However, it is not clear whether the soft and hard criteria are equally important in the context of medical systems (Bitkina et al., 2020).

The UEQ has been evaluated in many different contexts (Schrepp, Hinderks & Thomaschewski, 2014). Some researchers have advised to use a questionnaire like UEQ in the context of medical application development but have not tested it (Sauer, Hein, Muenzberg & Roesch, 2019). Others have used the UEQ to evaluate an E-learning platform for the user interface of medical diagnostic systems (Fajar, 2019). Like the Attrakdiff questionnaire, it seems that the UEQ is appropriate for the medical context, but more research is needed in the specific context of MRI and other medical imaging systems.

The meCUE and the Attrakdiff questionnaire have both been used in the context of measuring the user experience of different lower-limb orthoses (Doria, Minge, Riedel & Kraft, 2013). The importance of the user experience questionnaires was shown, as all three orthoses had an acceptable level of usability but there were significant differences between the products in terms of the hedonic product perceptions and emotions identified by the meCUE and Attrakdiff (Doria et al., 2013). Additionally, meCUE has been used in evaluating the user experience of different product concepts for navigating medical images from medical diagnostic systems (Hatscher, Luz & Hansen, 2017).

Appendix 10. Guides, PowerPoint slides and Miro board of focus group

Guide Final validation

1. ***Introduction***

Current process AD + PDLM

Four phases

Explain differences in usage of a questionnaire in each of the phases

> *1: Uncertain/prototypes and concepts (not possible to evaluate often in real context)*
>
> *2: integration of product, evaluate detailed designs. First components then full product evaluated*
>
> *3: more summative testing: often with clinical application specialists. Evaluating what has been accomplished*

4: First time actual evaluation in the real context, FOK → focus

2. ***Identifying labels***

Explain all labels

Should the usage mode be added as a label? This was something I found in literature that we have not discussed before

Are all labels useful? Are some labels not needed?

Are there any other labels that should be added?

What else needs to happen before the questionnaires can be send?

3. ***Conducting the evaluation***

Explaining step

Two different questionnaires (software or physical product)

Online tool → automatically sending questionnaire

For FOK → send in week 1, week 4 and week 12 (experience develops over time, expectation that first positive, after 4 weeks negative and after 12 week positive again)

Aim to keep context consistent to limit the influencing factors

Specify to users what they are evaluating

Is there something missing in this step that is important to consider?

4. Analyzing the results

Import data in prepared excel file,

Statistical results, → scales superfluous? Then remove

Would you add something to this?

What else is important for analyzing the results?

5. Communicating the results

Several different methods for communication

Which methods would you want to use?

Which are the most important

Something missing for the communication?

Simple enough for communication?

Linking to dashboard?

6. Overall process

What is missing from the steps described here?

What elements of these steps do you like and which do you not like?

Are these steps simple and clear enough?

Is it practical enough? Why/ why not>

What are the main boundaries for implementing this process?

Do you think these steps will help in evaluating the end-user experience?

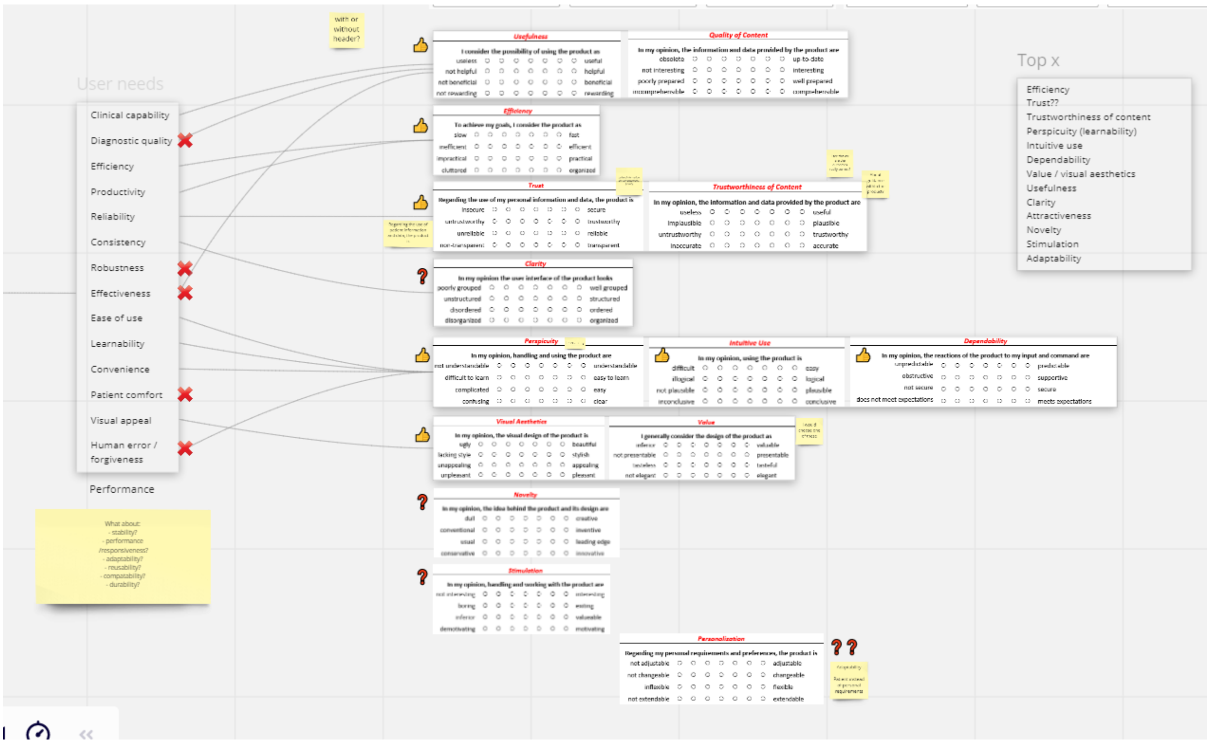
Do you think these steps will help in improving the experience of operators?

Focus group

	Employees involved

First focus group	Clinical application specialist 1 Clinical application specialist 4 Designer 1 Marketing 1
Second focus group	Clinical application specialist 1 Clinical application specialist 4 Clinical application specialist 6
Final validation	Clinical application specialist 1 Clinical application manager

Miro board second focus group:



Slides used in final validation

<Removed for confidentiality>

<Removed for confidentiality>

Four steps

- Step 1: Identifying labels
- Step 2: Conducting the evaluation (sending the questionnaire)
- Step 3: Analyzing the results
- Step 4: Communicating the results

Step 1 identifying labels

Label	Categories
Type of product	Software / Physical
Workflow steps	Workflow steps names
Phase in innovation process	Phase 1 / Phase 2 / Phase 3 / Phase 4
Market	USA / DACH / BENELUX / JAPAN
Previous experience	None / some / a lot
User type	Radiologist / lead technologist / technologist / application specialist
Solution	Solution name
Usage mode	Specific goal / freely using

Step 2 Conducting the evaluation (sending the questionnaire)

Physical product or **Software**

Online questionnaire (EFM Verint tools)

Sending in FOK (Week 1, week 4 and week 12)

Aim to keep context consistent (limit influencing factors)

Specify to users what they are evaluating (e.g., full experience or only solution / specific workflow steps)

Physical product:

Attractiveness, efficiency, learnability, dependability, fun-of-use, novelty, adaptability, usefulness, visual esthetics, intuitive use, clarity, *trust, haptics*

Software product:

Attractiveness, efficiency, learnability, dependability, fun-of-use, novelty, adaptability, usefulness, visual esthetics, intuitive use, clarity, *trust (privacy), trustworthiness of content*

Number of participants: Around 20 people for stable results
Depends on variance in the data
Reliability decreases with less participants

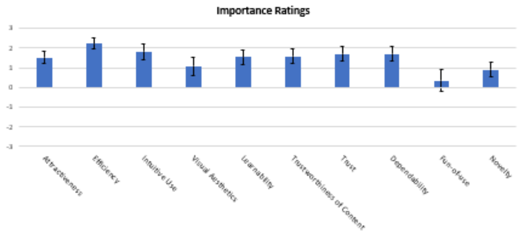
Usually, around 30 people involved in FOK

Step 3 Analyzing the results

- Import data in prepared Excel file
- Statistical results (e.g., t-test and Cronbach alpha > 0.7)
- Make visualizations

Over time:

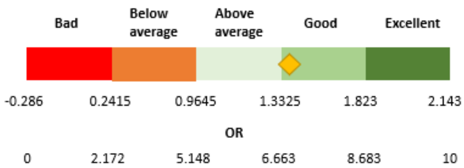
- Remove scales if not important for the experience
- Remove scales if time is too long for participants



Step 4 Communicating the results

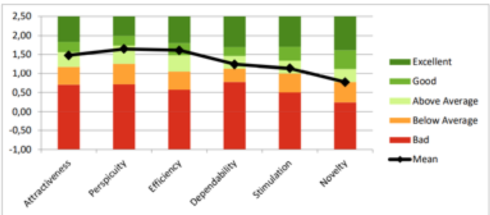
- Dashboard
  - External benchmark KPI score
  - Internal benchmark KPI score
  - Score for each scale
  - Importance ratings

1. External Benchmark (UEQ KPI)



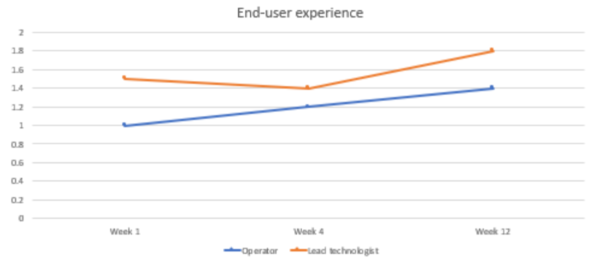
Score:
(UEQ KPI + 0.286) / 2.429 * 10
(UEQ KPI + 0.286) * 4.116920543

2. External Benchmark (six scales)



- *Excellent*: In the range of the 10% best results.
- *Good*: 10% of the results in the benchmark data set are better and 75% of the results are worse.
- *Above average*: 25% of the results in the benchmark are better than the result for the evaluated product, 50% of the results are worse.
- *Below average*: 50% of the results in the benchmark are better than the result for the evaluated product, 25% of the results are worse.
- *Bad*: In the range of the 25% worst results.

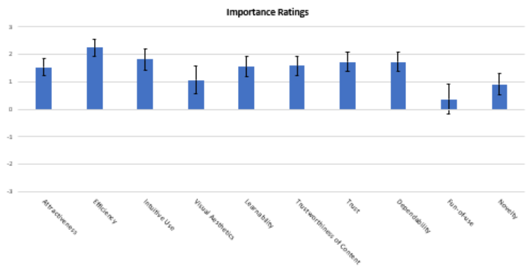
3. Internal Benchmark



4. Dashboard

	Baseline	New product	Difference
External UEQ KPI	1,55 (Good)	1,62 (Good)	+ 0,07
Internal UEQ KPI	1,62	1,80	+ 0,18

5. Importance ratings

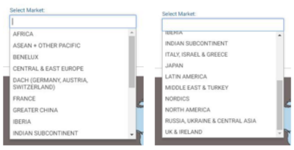


Label	Categories
Type of product	Software / Physical
Product fidelity	Idea / Concept / Prototype / Integrated product / Installed base
Innovation phase	AD / Develop user and business needs and concepts / Finalize high level requirements, designs and plans / Detail designs / Integrate product / Verify designs / Validate and release designs / Prepare launch / Monitor launch
Region	Africa / ASEAN + other pacific / BENELUX / Central + East Europe / DACH (Germany, Austria, Switzerland) / France / Greater China / Iberia / Indian subcontinent / Italy, Israel & Greece / Japan / Latin America / Middle east & Turkey / Nordics / North America / Russia, Ukraine & Central Asia / UK & Ireland
Professional experience	Less than 1 year / 1-3 years / 3-8 years / more than 8 years
User	Radiologist / lead-technologist / technologist
User proxy	User / Application specialist
Solution	Solution name
Usage mode	Guided / Non-guided use

Options

Label	Categories
Product fidelity	Idea / (Concept) / Prototype / Functional product / Installed base
Innovation phase	AD / Develop user and business needs and concepts / Finalize high level requirements, designs and plans / Detail designs / Integrate product / Verify designs / Validate and release designs / Prepare launch / Monitor launch

Markets



BENELUX
Central + East Europe
DACH (Germany, Austria, Switzerland)
France
Greater China
Iberia
Indian subcontinent
Italy, Israel & Greece,
Japan, Latin America,
Middle east & Turkey
Nordics,
North America
Russia, Ukraine & Central Asia
UK & Ireland

Appendix 11. Online Customized Questionnaire and Excel file

Online Questionnaire

Thank you for participating in this evaluation!

Please complete the following questionnaire to assess the product. The questionnaire contains pairs of opposing product properties. The grades between the opposites are indicated by circles. Check one of the circles to indicate your level of agreement with the individual terms.

Example:

1 2 3 4 5 6 7

unpleasant ○ ◉ ○ ○ ○ ○ ○ pleasant

With this assessment, you state that you consider the product rather unpleasant than pleasant.

For each group of four items, please fill in the importance of the respective aspect (e.g. attractiveness) for your overall impression of the product.

Example:

For me, attractiveness of the product is:

1 2 3 4 5 6 7

completely irrelevant ◉ ○ ○ ○ ○ ○ ○ very important

With this assessment, you state that the last four items (e.g. representing attractiveness) were completely irrelevant to your overall impression of the product.

Try to make a spontaneous decision! It is important not to think too long about the terms to reach a direct assessment. Please always select an answer, even if you are insecure about your assessment of one pair of terms or if you think that it does not fit the product. This survey will take approximately 5 minutes.

How many years of professional experience do you have with MRI?

- ○ Less than 1 years
- ○ 1-3 years
- ○ 3-8 years
- ○ More than 8 years

What region do you live in?

What type of user would you classify yourself as?

- ○ Technologist
- ○ Lead-technologist
- ○ Radiologist

Attractiveness

In my opinion, the product is generally:

	1	2	3	4	5	6	7	
annoying	○	○	○	○	○	○	○	enjoyable
bad	○	○	○	○	○	○	○	good
unpleasant	○	○	○	○	○	○	○	pleasant
unfriendly	○	○	○	○	○	○	○	friendly

For me, attractiveness of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Efficiency

To achieve my goals, I consider the product as:

	1	2	3	4	5	6	7	
slow	○	○	○	○	○	○	○	fast
inefficient	○	○	○	○	○	○	○	efficient
impractical	○	○	○	○	○	○	○	practical
cluttered	○	○	○	○	○	○	○	organized

For me, efficiency of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Learnability

In my opinion, handling and using the product are:

	1	2	3	4	5	6	7	
not understandable	○	○	○	○	○	○	○	understandable
difficult to learn	○	○	○	○	○	○	○	easy to learn
complicated	○	○	○	○	○	○	○	easy
confusing	○	○	○	○	○	○	○	clear

For me, learnability of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Dependability

In my opinion, the reactions of the product to my input and commands are:

	1	2	3	4	5	6	7	
unpredictable	○	○	○	○	○	○	○	predictable
obstructive	○	○	○	○	○	○	○	supportive
not secure	○	○	○	○	○	○	○	secure
not meeting expectations	○	○	○	○	○	○	○	meeting expectations

For me, the dependability of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Fun-of-use

In my opinion, handling and working with the product are:

	1	2	3	4	5	6	7	
not interesting	○	○	○	○	○	○	○	interesting
boring	○	○	○	○	○	○	○	exiting
inferior	○	○	○	○	○	○	○	valuable
demotivating	○	○	○	○	○	○	○	motivating

For me, fun-of-use of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Novelty

In my opinion, the idea behind the product and its design is:

	1	2	3	4	5	6	7	
dull	○	○	○	○	○	○	○	creative
conventional	○	○	○	○	○	○	○	inventive
usual	○	○	○	○	○	○	○	leading edge
conservative	○	○	○	○	○	○	○	innovative

For me, novelty of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Adaptability

Regarding patient characteristics and preferences, the product is:

	1	2	3	4	5	6	7	
not adjustable	○	○	○	○	○	○	○	adjustable
not changeable	○	○	○	○	○	○	○	changeable
inflexible	○	○	○	○	○	○	○	flexible
not extendable	○	○	○	○	○	○	○	extendable

For me, adaptability of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Usefulness

I consider the product as:

	1	2	3	4	5	6	7	
useless	○	○	○	○	○	○	○	useful
not helpful	○	○	○	○	○	○	○	helpful
not beneficial	○	○	○	○	○	○	○	beneficial
not rewarding	○	○	○	○	○	○	○	rewarding

For me, usefulness of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Visual esthetics

In my opinion, the visual design of the product is:

	1	2	3	4	5	6	7	
ugly	○	○	○	○	○	○	○	beautiful
lacking style	○	○	○	○	○	○	○	stylish
unappealing	○	○	○	○	○	○	○	appealing
unpleasant	○	○	○	○	○	○	○	pleasant

For me, visual esthetics of the product are:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Intuitive use

In my opinion, using the product is:

	1	2	3	4	5	6	7	
difficult	○	○	○	○	○	○	○	easy
illogical	○	○	○	○	○	○	○	logical
not plausible	○	○	○	○	○	○	○	plausible
inconclusive	○	○	○	○	○	○	○	conclusive

For me, intuitive use of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Trust

In my opinion, the product is:

	1	2	3	4	5	6	7	
useless	○	○	○	○	○	○	○	useful
implausible	○	○	○	○	○	○	○	plausible
untrustworthy	○	○	○	○	○	○	○	trustworthy
inaccurate	○	○	○	○	○	○	○	accurate

For me, trust in the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Haptics

In my opinion, the product feels:

	1	2	3	4	5	6	7	
unstable	○	○	○	○	○	○	○	stable
unpleasant to touch	○	○	○	○	○	○	○	pleasant to touch
rough	○	○	○	○	○	○	○	smooth
slippery	○	○	○	○	○	○	○	slip-resistant

For me, the haptics of the product are:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Clarity

In my opinion, the product looks:

	1	2	3	4	5	6	7	
poorly grouped	○	○	○	○	○	○	○	well grouped
unstructured	○	○	○	○	○	○	○	structured
disordered	○	○	○	○	○	○	○	ordered
disorganized	○	○	○	○	○	○	○	organized

For me, clarity of the product is:

	1	2	3	4	5	6	7	
completely irrelevant	○	○	○	○	○	○	○	very important

Was there anything you did not understand about this questionnaire? Or do you have any other remarks?

Excel file

Calculation of a KPI

For each scale the UEQ+ collects ratings for 4 items and one rating for the overall importance of the scale. These values are used to calculate a KPI that should represent the overall UX impression of the product.
For the calculation the relative importance of the scale and the scale mean per participant are calculated. The KPI is then simply the mean over all participants.

KPI 1.39 **Std.Dev. 0.88**

Attractiveness	Efficiency	Intuitive Use	Visual Aesthetics	Quality of Conten	Trustworthiness c	Trust	Choose Scale	Choose Scale	Choose Scale	KPI
0.61	0.85	0.99	0.41	0.81	0.74	0.49				0.90
0.77		0.90	0.77	1.03	0.90	1.26				1.62
0.88	1.24	0.62	0.79	0.53	0.59	0.71				1.35
1.16	0.74	0.71	0.48	1.05	1.05	0.39				1.58
0.77	0.94	0.77	0.70	0.76	1.02	0.62				1.59
1.00	1.00	1.00	1.00	1.00	1.00	1.00				3.00
0.67	0.73	0.48	0.81	0.85	1.00	0.54				1.08
0.99	0.99	1.03	0.34	0.66	0.82	0.79				1.61
0.59	0.86	0.63	0.41	0.70	0.66	0.31				0.16
0.73	1.15	1.20	0.59	0.70	0.46	0.51				1.34
0.88	1.11	0.70	0.63	0.80	0.84	1.15				2.12
0.94	1.14	0.98	0.98	1.06	0.84	0.42				2.35
0.76	0.75	0.47	0.44	0.73	0.79	0.67				0.62
0.64	0.21	0.29	0.29	0.29	0.50	0.46				-1.32
0.76	0.83	0.52	0.29	0.47	0.48	0.82				0.16
0.86	0.74	0.66	0.62	0.73	0.61	0.57				0.79
0.78	0.82	0.80	0.88	1.03	1.07	1.03				2.40
0.72	0.56	0.41	0.55	0.61	0.75	0.91				0.51
0.72	0.87	0.94	0.51	0.99	0.99	0.64				1.65
0.86	0.93	0.86	0.71	0.82	0.89	0.89				1.96
0.87	1.08	1.26	0.61	1.00	0.78	0.44				2.05
1.13	0.92	0.71	0.83	0.83	0.83	1.00				2.26
0.69	1.17	1.53		0.63	0.82	0.84				1.68
0.68	0.95	0.95	0.45	0.88	0.78	0.72				1.42
0.66	1.11	0.95	0.55	0.86	0.83	0.59				1.54
0.91	0.76	0.66	0.42	0.99	1.11	0.63				1.46
0.42	0.46	0.46	0.64	0.77	0.41	0.88				0.05
1.00	1.00	1.00	1.00	1.00	1.00	1.00				3.00

Enter the observed item ratings here

Attractiveness				Efficiency				Intuitive Use				Visual Aesthetics				Quality of Content				Trustworthiness of Conten				Trust				Choose Scale	Choose Scale	Choose Scale
5	4	5	4	6	6	4	5	3	6	6	6	6	5	4	5	6	5	7	6	6	6	6	4	3	4	3	2			
6	6	6	6	5	5	5	5	5	5	5	5	5	6	5	4	6	6	5	6	5	5	5	5	7	7	7	7			
5	6	7	6	5	7	6	6	6	5	5	5	4	4	5	5	4	5	4	5	6	5	5	4	6	6	7	5			
7	6	5	6	7	7	7	2	6	5	6	5	5	5	5	5	7	6	6	7	6	7	7	6	3	3	3	3			
5	5	6	5	5	6	6	5	6	5	6	4	7	5	6	5	7	7	5	6	7	7	7	7	4	5	5	3			
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7			
5	5	6	5	3	6	6	4	3	3	3	6	6	5	5	5	7	6	4	5	5	7	7	7	4	5	6	2			
6	6	5	6	6	6	6	5	6	6	6	6	4	4	4	4	6	6	5	6	6	6	5	6	6	6	6	4			
5	5	5	4	6	5	6	5	5	5	5	5	4	3	3	3	3	5	5	5	5	4	4	4	4	1	2	1			
6	6	6	6	7	7	7	6	7	7	7	7	5	3	4	4	7	5	5	6	4	4	4	3	4	4	4	2			
6	6	5	7	6	7	7	6	6	5	6	6	7	6	7	6	6	5	6	5	7	5	5	6	7	7	6	7			
7	7	7	6	7	7	7	7	7	7	7	7	6	6	6	6	7	7	6	6	7	6	6	5	5	5	5	3			
6	6	6	4	5	4	5	4	4	4	4	5	4	4	4	4	5	5	5	6	6	4	4	5	4	4	4	4			
6	4	6	2	2	2	1	1	2	2	2	2	1	3	2	2	3	2	1	2	2	4	4	4	4	3	4	2			
6	6	6	4	3	6	4	3	2	3	2	5	4	4	5	4	6	2	4	6	2	4	4	4	4	6	6	3			
6	5	6	5	4	6	5	4	5	3	5	4	5	3	4	4	7	4	5	6	5	6	6	5	4	5	5	3			
6	6	6	6	6	7	6	6	6	5	5	5	7	7	7	6	7	7	6	7	7	7	7	7	7	7	7	6			
6	5	5	5	3	4	3	4	3	3	3	3	4	4	4	4	6	6	3	3	4	6	6	6	6	6	6	5			
7	7	7	7	7	7	7	6	6	5	5	5	5	5	5	5	6	6	5	5	6	6	6	4	5	5	5	5			
6	6	6	6	7	7	7	5	7	6	6	5	5	5	5	5	7	6	4	6	6	6	7	6	6	7	6	6			
7	6	7	5	6	7	7	6	6	7	7	6	5	6	5	6	4	7	7	6	7	7	7	7	4	5	5	2			
7	7	7	6	6	6	5	5	6	6	6	6	5	5	5	5	7	7	7	7	7	7	7	7	7	7	7	7			
6	6	5	5	7	7	7	4	7	7	7	7	6	5	6	5	5	5	5	5	5	5	6	5	5	4	5	4			
7	6	6	5	6	6	6	6	6	6	6	6	4	4	4	4	6	6	5	5	6	5	6	6	4	6	5	3			
5	6	4	5	6	7	6	5	5	6	7	6	6	5	5	5	6	6	7	7	6	4	6	5	4	5	7	2			
7	6	5	5	7	6	5	5	4	5	5	6	4	4	4	4	7	6	6	6	7	6	5	6	5	4	5	5			
3	2	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5	5	5	4	4	4	4	6	6	6	5			
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7			
6	6	6	6	7	7	7	6	6	6	6	6	6	6	5	6	7	6	6	5	6	5	5	5	5	5	5	5			
7	6	6	4	6	7	7	3	6	6	6	6	5	5	6	7	4	5	4	6	6	4	4	4	3	4	3	1			
6	6	6	6	5	6	5	4	6	4	5	4	6	6	6	6	5	6	5	6	6	7	7	5	3	4	5	3			

Appendix 12. Workflow steps of MRI scanning