Assessment and Improvement of Initial Learnability in Complex Systems
A Qualitative Study to Promote Intuitive Software Development

Master's Thesis in Cognitive Science

JONATAN ANDERSSON

Department of Computer and Information Science
LINKÖPING UNIVERSITY
Linköping, Sweden 2016
Assessment and Improvement of Initial Learnability in Complex Systems

– A Qualitative Study to Promote Intuitive Software Development

Jonatan Andersson

Supervisor: Mathias Nordvall
Examinator: Arne Jönsson
Assessment and Improvement of Initial Learnability in Complex Systems
A Qualitative Study to Promote Intuitive Software Development
JONATAN ANDERSSON

© JONATAN ANDERSSON, 2016.

Supervisor: Mathias Nordvall, Department of Computer and Information Science
Examiner: Arne Jönsson, Department of Computer and Information Science
Department of Computer and Information Science
Linköping University
SE-581 83 Linköping
Telephone +013-28 10 00

Typeset in \TeX
Linköping, Sweden 2016

iv
Abstract

This Master’s Thesis aimed to assess and propose improvements for initial learnability in Sectra AB’s Picture Archiving Communication System (PACS) by integrating usability engineering and agile software development. Assessing initial learnability and re-designing complex systems is difficult as they have a high skill cap and take longer time to learn in comparison to simpler ones. Further, companies that use agile methodologies often focus on completing small items which might hide the overarching vision of a product that can lead to usability problems. While there are several methods for assessing usability, no research has specifically focused on assessing initial learnability in complex systems. This study however investigates how this may be achieved by combining current methodologies for measuring learnability with usability engineering and agile software development. Initial learnability issues and needs were assessed after evaluating Sectra PACS using 5 participants and analysed using impact mapping as well as conducting a focus group within the organisation that owns the product.
Acknowledgements

Without any specific order I’d like to thank Sanna Heurlén, Agatha Allen and the rest of the UX team at Sectra who has treated me as a team member straight from the start of this project and provided me with excellent insight and knowledge. Further, I’d also like to thank Magnus Ranlöf for believing in my work while continuously supporting me throughout this journey. I’d also like to thank my advisor Mathias Nordvall from Linköping University for his passionate discussions and suggestions for which I’ll always be grateful for. Finally, I want to thank my family for always believing in me throughout my academic education, and my love Emelie Arvidsson for her endless support.

Jonatan Andersson, Linköping, May 2016
# Contents

1 Background .............................................. 1

2 Introduction ........................................ 3
   2.1 Motivation ........................................ 3
   2.2 Purpose .......................................... 3
   2.3 Research question ................................ 3
   2.4 Limitations ....................................... 4

3 Theory .............................................. 5
   3.1 Learnability ...................................... 5
      3.1.1 Definition .................................... 5
      3.1.2 Learnability Evaluation Metrics .............. 8
         3.1.2.1 Task Metrics: Metrics based on task performance 8
         3.1.2.2 Command Metrics: Metrics based on command usage 9
         3.1.2.3 Mental Metrics: Metrics based on cognitive processes 9
         3.1.2.4 Documentation Metrics: Metrics based on documentation usage 9
         3.1.2.5 Usability Metrics: Metrics based on change in usability 10
         3.1.2.6 Rule Metrics: Metrics based on specific rules .... 10
         3.1.2.7 Subjective Metrics: Metrics based on user feedback 10
   3.2 Used Evaluation methodologies .................... 11
      3.2.1 Questing-suggestion Think-aloud Protocol .... 11
      3.2.2 Semi-Structured Interview .................... 12
      3.2.3 System Usability Scale ....................... 14
         3.2.3.1 Amount of users required ................. 15
         3.2.3.2 Using SUS ................................ 15
         3.2.3.3 Interpreting the results ................ 16
         3.2.3.4 All Positive SUS .......................... 17
         3.2.3.5 Additional Questionnaire Items .......... 19
         3.2.3.6 Conclusion ................................ 20
   3.3 Analyzing the data ................................ 20
      3.3.1 Persona ....................................... 20
         3.3.1.1 Motivation ................................ 21
         3.3.1.2 Creating a persona ....................... 21
      3.3.2 Focus groups ................................... 22
      3.3.3 Impact Mapping ............................... 23
3.3.3.1 Introduction ........................................... 23
3.3.3.2 Levels .................................................. 23
3.4 Conclusion .................................................. 25

4 Methods ...................................................... 27
4.1 Pre-study ..................................................... 27
4.1.1 Literature study ........................................... 27
4.1.2 Getting to know the system ................................. 27
4.1.3 Recruiting users ............................................ 28
4.1.4 Defining the tasks ........................................ 28
4.1.5 Pilot Study ................................................ 28
4.2 Design ....................................................... 29
4.2.1 Apparatus ................................................ 29
4.2.2 Procedure Overview ...................................... 30
4.2.3 Introduction .............................................. 30
4.2.4 Question-Suggestion Protocol ............................. 31
4.2.5 Motivation ............................................... 32
4.2.6 Tasks ...................................................... 32
4.2.7 Interview .................................................. 35
4.2.7.1 Motivation ........................................... 35
4.2.8 Questionnaire ............................................ 35
4.2.8.1 Motivation ........................................... 36
4.2.9 Users ...................................................... 36
4.3 Data analysis ............................................... 37
4.3.1 Question-asking protocol ................................ 37
4.3.2 Interview .................................................. 37
4.3.3 All Positive System Usability Scale ....................... 37
4.3.4 Clustering the data ....................................... 37
4.3.5 Persona ................................................... 40
4.3.6 Focus group meetings .................................... 40
4.3.6.1 Interview regarding learning ......................... 40
4.3.6.2 Focus groups ......................................... 40
4.3.6.3 Motivation ........................................... 41
4.3.7 Impact Mapping .......................................... 41
4.4 Conclusion .................................................. 42

5 Results ....................................................... 43
5.1 Persona: Ruby Radiographer ................................. 43
5.2 Tasks that were completed with ease ....................... 45
5.3 Clusters ...................................................... 45
5.3.1 Terminology ............................................. 46
5.3.2 Information overload ..................................... 49
5.3.3 Missing features ........................................ 51
5.3.4 Confusion, needs explanation ......................... 54
5.3.5 Flow problems .......................................... 55
5.3.6 Function placement ..................................... 57
5.3.7 System error management ............................... 61
# Contents

5.4 Learning ............................................. 62  
5.5 Questionnaire ................................. 65  
5.6 Impact Mapping ................................. 65  
  5.6.1 Goal ............................................. 66  
  5.6.2 Actors ............................................. 66  
  5.6.3 Impacts ............................................. 66  
  5.6.4 Deliverables ................................. 68  
5.7 Conclusion ............................................. 71  

6 Discussion ............................................. 73  
  6.1 Results ............................................. 73  
  6.2 Methods ............................................. 74  
  6.3 Ethical aspects ............................................. 76  

7 Conclusion ............................................. 77  
  7.1 Future Work ............................................. 78  

Bibliography ............................................. 85  

A Appendix 1 ............................................. I
1 Background

This thesis is done in collaboration with Sectra AB which is a company that specializes in medical IT and secure communications [1]. This study will focus on assessing, evaluating and improving the initial learnability in their so called Picture Archiving Communication System (PACS). A PACS is medical imaging technology which provides storage and access to images from different modalities (i.e. machines). Thus, as a PACS transmits both reports and images digitally it eliminates the need to manually file, retrieve, or transport film jackets. Simply put the PACS is used to both distribute and analyse images such as X-rays in order to diagnose and treat patients correctly. As the images and reports stored in the PACS are relevant to many kinds of users such as general practitioners, radiologists, radiographers, assistant nurses etc. the system will have a high level of functionality to fulfil the different users’ different needs. As a result of this, the system may be difficult to learn how to use as part of the users’ daily work. Today when a new customer buys their Picture Archiving Communication System (PACS) they’re offered a two week training session in order to learn how to use the system. The reason for this is as mentioned that the system is complex but also has a very high skill cap. Additionally it takes time to learn how to complete daily work tasks due to its complexity, which is why the two week training session exists. Sectra however believes that there’s room for improvement in their PACS and that it needs to be easier to learn how to use it. As such, they want to assess the PACS’ learnability in order to make it easier to use. In addition to this, they want current literature regarding learnability and methods to be evaluated with any results documented in a report summarizing findings, learnability guidelines and motivations of methods. Further, found learnability issues should be documented in a way that it’s possible for Sectra Staff to continue to develop/integrate them after the study. In addition to this, improvements that are suggested should be documented in a way that it is possible for Sectra staff to use it and continue to develop/integrate it after the project.
1. Background
2
Introduction

This chapter will motive my study, introduce its purpose and present my research questions.

2.1 Motivation

Sectra is a company that delivers advanced IT system within the medical sector. As their software treats confidential patient information it’s vital that the systems are secure, stable and easy to use. Designing for the medical sector is however not an easy task. Most programs involve a lot of different users all over the world. Thus, most medical IT products are highly complex systems as they must have functionality that allow different user groups to complete their goals and fulfil their needs. As a consequence, the system will have more options and therefore require more knowledge to manage effectively. This study focused on Sectra’s Picture Archiving and Communication System (PACS) which is a medical imaging technology that provides storage and access to images from different modalities. Further, Sectra has been considered to have best PACS available on the market by winning Best In KLAS 2015/16 which makes the product even more interesting. While it’s known that Sectra has a very good product, they want to make their PACS easier to use by assessing the PACS initial learnability and then improve it. In addition to this, initial learnability in complex systems seems to be an unexplored area within academia which I hope to help fill by proposing a process chain of methods that may be used for assessment, evaluation and improvement in any given context.

2.2 Purpose

The purpose of this study is to research, describe and understand the term learnability by evaluating and suggesting improvements for Sectra PACS. Further, ways to solve learnability issues will be mapped out by using agile methods combined with usability engineering.

2.3 Research question

As Sectra provided me with their system, knowledge and resources I was able to study their PACS from its core and gather information directly through their staff. This easy access to very accurate and rich data made it possible to dive deep into
their system as well as their organisation which gave me a good overview of their work as a whole. Thus, this unique environment let me form a set of methods which I wouldn’t have been able to study without their continuous support throughout my work. Further, initial learnability in complex systems hasn’t been studied before which makes this study an academical contribution. As a result of these premises I formed my research question which is listed below:

**How can learnability methods be used to assess and improve initial learnability in complex systems?**

Finally, I define a complex system as a system which requires a lot of knowledge to use as a novice and even more experience in order to become an expert while also consisting of many different user groups and segments.

### 2.4 Limitations

This study only evaluated a segment of Sectra PACS due to time being limited as well as the fact that one user group won’t use or experience the whole system. Further, only initial learnability will be assessed and learnability will therefore not be studied over time. Finally, this study only evaluated one complex system within a specific area of use - medical IT.
3 Theory

This chapter will introduce, motivate and explain the used theory in this study.

3.1 Learnability

This section will introduce, define and propose evaluation methods for learnability.

3.1.1 Definition

When we talk about interfaces and products in general, we often refer to usability which may be defined as 'quality of use' [2] or 'quality attribute' [3, 4] which assesses how easy user interfaces are to use. Further, Nielsen [3] described the terms in the following list in order to better understand what usability is:

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

In addition this he breaks usability down into 5 different usability components as shown in figure 3.1.

![Usability and its five quality components.](image)

**Figure 3.1:** Usability and its five quality components.

This study will however focus on 'learnability' which as mentioned is one of the five quality components which defines usability together with efficiency, memorability, errors and satisfaction [5, 6]. While many agree that learnability is fundamental and perhaps the most important quality component of the five (as all interface usage requires some learning) there is no agreed upon definition of the term [7, 8,
3. Theory

9, 10, 2, 11, 12, 13]. Some of the most commonly used definitions will include those listed below in table 3.1:

**Table 3.1: Learnability definitions**

<table>
<thead>
<tr>
<th>Number</th>
<th>Source</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Santos and Badre (1995) [15]</td>
<td>Measure of the effort required for a typical user to be able to perform a set of tasks using an interactive system with a predefined level of proficiency.</td>
</tr>
<tr>
<td>4</td>
<td>Hart and Staveland (1988) [16]</td>
<td>The speed and ease with which users feel that they have been able to use the product or as the ability to learn how to use new features when necessary.</td>
</tr>
<tr>
<td>5</td>
<td>Bevan and Macleod (1994) [17]</td>
<td>A measure of comparison the quality of use for users over time.</td>
</tr>
<tr>
<td>6</td>
<td>Butler (1985) [18]</td>
<td>Initial user performance based on self instruction” and “[allowing] experienced users to select an alternate model that involved fewer screens or keystrokes.</td>
</tr>
<tr>
<td>7</td>
<td>Kirakowski and Claridge (1998) [19]</td>
<td>Degree to which users feel they can get to use the site if they come into it for the first time, and the degree to which they feel they can learn to use other facilities or access other information once they have started using it.</td>
</tr>
<tr>
<td>8</td>
<td>ISO 9126-1 (2001) [20]</td>
<td>The capability of the software product to enable the user to learn its application.</td>
</tr>
<tr>
<td>9</td>
<td>ISO 25010 (2011) [21]</td>
<td>Degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use.</td>
</tr>
</tbody>
</table>

In addition to this, Grossman et al. [22] examined 88 papers dating from 1982 to 2008 and divided found definitions of learnability into 8 categories:

- Used without a definition: (45)
• Generic learnability (i.e. “easy to learn”): (7)
• Generic usability (i.e. “easy to use”): (3)
• First time performance: (17)
• First time performance after instructions: (4)
• Change in performance over time: (8)
• Ability to master system: (4)
• Ability to remember skills over time: (2)

Their work confirm that there is indeed a lack of an agreed definition of learnability. This means that studies that do not use a definition are difficult to replicate and therefore have a lower validity and reliability due to its lack of describing what they actually want to research. In order to prevent this, Grossman et al. propose a taxonomy which will be listed below:

Figure 3.2: Grossman et al.’s learnability taxonomy [22]

The purpose of this taxonomy is to create a framework that may be used by researchers and practitioners to isolate specific areas of learnability and convey their intentions. Additionally, rather than declaring a set definition of learnability they instead divide it into two main categories:

• Initial Learnability: Initial performance with the system.
• Extended Learnability: Change in performance over time.

‘Performance’ itself may have different definitions such as time on task, task success, error rates or subjective measurements. These metrics will be presented later in the theory chapter. With Grossman’s framework in mind, I define learnability in this study as the user’s ability to eventually achieve specific performance. This performance would be translated into completing tasks that are related to their work and will be explained further when describing the methodology. Thus, my
3. Theory

definition falls into initial learnability as once the specific performance is achieved the user would still be considered a novice as they don’t have to complete the tasks optimally. To clarify, an expert would find the optimal solution to completing the tasks while I only want the user to simply complete them. By having a set definition of what learnability refers to it is easier to understand what a 'learnability issue' would imply, which in this case is a problem that makes it more difficult to achieve specific performance.

3.1.2 Learnability Evaluation Metrics

The metrics that will be presented in this section are mainly found by Grossman et al. during their comprehensive learnability study. While they have done a great job finding and shortly introducing metrics that aim to evaluate learnability, I’ll explain why they aren’t fit for my specific purpose of this study but might be of interest when evaluating learnability in general. Further, by understanding what metrics are available I hope that it’ll become clearer why I chose the metrics that I did. The purpose of this section is therefore to introduce possible evaluation metrics and explain why they aren’t fit for evaluating initial learnability specifically in my study. The metrics which I have decided to use will then be discussed in next section and further clarified.

3.1.2.1 Task Metrics: Metrics based on task performance

Task metrics aim to evaluate the participant’s performance and may include:

1. Percentage of users who complete a task optimally. [23]
2. Percentage of users who complete a task without any help. [23]
3. Ability to complete task optimally after certain time frame. [18]
4. Decrease in task errors made over certain time interval. [24]
5. Time until user completes a certain task successfully. [5]
6. Time until user completes a set of tasks within a time frame. [5]
7. Quality of work performed during a task, as scored by judges. [25]

While task performance is a great way to gather quantitative data, they require a large set of participants to be statistically valid. Nielsen [26] suggests that you should have at least 20 participants in a study with quantitative metrics to get statistically significant numbers, and even more to have tight confidence intervals. Thus, using these metrics would require a lot of resources and finding participants might be difficult. For instance, recruiting 20 users for my study wouldn’t be possible as there aren’t enough participants available who fulfil my requirements which will be discussed later. These metrics could however be extremely powerful for a company like Sectra. By using logging data it’s possible to automatically collect data over time with minimal effort. The interpretation of the data may however not be easy. For instance, if we’d like to know how many users complete a task optimally we would have to use logical rules to determine how the path would look.
3. Theory

3.1.2.2 Command Metrics: Metrics based on command usage

Research has shown that command usage may be used to evaluate learnability and include metrics such as:

1. Success rate of commands after being trained. [27]
2. Increase in commands used over certain time interval. [24]
3. Increase in complexity of commands over time interval. [24]
4. Percent of commands known to user. [28]
5. Percent of commands used by user. [28]

The metrics mentioned above take commands as an attribute for how well a user understands and uses the system. Command usage could however be considered as something that will evolve over time and therefore fits better as a tool to evaluate extended learnability. Thus, as this study aims to assess and evaluate initial learnability command usage will not be a part. With this being said, logging user commands could be very beneficial for Sectra to see how their users gain experience over time.

3.1.2.3 Mental Metrics: Metrics based on cognitive processes

Mental metrics evaluate the participant’s cognitive processes in order to assess learnability and include:

1. Decrease in average think times over certain time interval. [24]
2. Alpha vs. beta waves in EEG patterns during usage. [29]
3. Change in chunk size over time. [15]
4. Mental Model questionnaire pretest and post test results. [30]

Metrics based on cognitive processes may be interesting when evaluating learnability on a micro-level, but when evaluating a large and complex system the data analysis would be very complex to interpret and understand. Further, as this study aims to provide methods which may be used with ease and at a low cost the metrics above aren’t suitable for my specific purpose nor something that Sectra would want to invest in at this time.

3.1.2.4 Documentation Metrics: Metrics based on documentation usage

Documentation usage is another metric cluster that may be used to evaluate learnability and include:

1. Decrease in help commands used over certain time interval. [24]
2. Time taken to review documentation until starting a task. [24]
3. Time to complete a task after reviewing documentation. [24]

These documentation metrics are also quantitative metrics that would require extended use of the system. They may however be of interest for Sectra as they could...
potentially find areas for improvement without actually having to setup meetings with participants. Further, it could also be analysed which commands are used and some that perhaps aren’t used at all. This could help Sectra find "hidden" functions, also known as function that users maybe want to use but can’t find due to their placement.

3.1.2.5 Usability Metrics: Metrics based on change in usability

Metrics based on change in usability include:

1. Comparing “quality of use” over time. [2]
2. Comparing “usability” for novice and expert users. [2]

These metrics require intense data collection over extended time with a large user group. As quality of use is most often based on subjective data you’d have to find a group of participant whom be willing to provide continuous feedback over a large time span. Further, comparing usability between novice and expert users is very interesting as it might show what issues disappear with experience and which may not.

3.1.2.6 Rule Metrics: Metrics based on specific rules

Rule metrics may also be used to evaluate learnability and are based on:

1. Number of rules required to describe the system. [31, 32]

These metrics aim to find potential learnability issues in an early design process by creating a prototype, specifying the meaning of the prototype labels and then specify the actions for manipulating the device which results in a model. The model then learns consistent from guidance rather than being programmed. As a result of this the model generalizes from the examples through which the designer guides it. By doing this it’s possible to test how consistent an interface is, and as such evaluating how good its learnability is.

3.1.2.7 Subjective Metrics: Metrics based on user feedback

Metrics based on user feedback produce subjective data and will may include:

1. Number of learnability related user comments. [24]
2. Learnability questionnaire responses. [33, 34]
3. Twenty six Likert statements. [33]

The metrics aim to collect subjective data from the users. While subjective data is vulnerable to the fact that people are different, it evaluates whether or not the user actually enjoys using the product. As I’m going to use metrics based on user feedback in my study, the methodologies I’ve chosen will be more carefully explained in the next section.
3.2 Used Evaluation methodologies

With the previous section in mind, quantitative measurements may not be the optimal choice for evaluating initial learnability in general. The reason for this is that time measurement for instance will not tell us why it is difficult to find a specific function while task succession will not make us understand why some tasks are more difficult to complete than others. Instead, I propose the use of qualitative methods in order to assess the subjective information from the user. Some advantages include:

- It gives more in-depth data
- It requires fewer test participants to produce relevant data
- Issues and needs can be discussed on spot

The methods chosen for the study will be introduced and motivated below in the next segment

3.2.1 Questing-suggestion Think-aloud Protocol

The traditional think-aloud (TA) protocol is one of the most commonly used tools when conducting usability tests. When using a think aloud protocol the participants are asked to use a system while simultaneously thinking out loud, or in other words verbalizing their thoughts. Usability issues are found by encouraging the participant to verbally articulate what s/he is thinking or feeling when encountering a problem and how said problem can be solved [35, 36].

Many variations of the original TA protocol have been proposed with the two most common ones being the concurrent TA and the retrospective TA [37, 38, 35, 39, 40]. When using the concurrent TA a test administrator will ask the participant to voice aloud thoughts, feelings and reasoning while completing one or more tasks using the system that is being evaluated. In contrast, the retrospective TA is used at the end of a test session in order to collect the participant’s thinking and reasoning processes while they are still in the short-term memory of using a system. While both variations have their advantages, it is important to choose method depending on what kind of system is being evaluated. For instance, research has shown that our gaze will be slightly disrupted when talking. If we then use the concurrent TA while also using eye-tracking equipment our data may be corrupted [41]. Instead, the retrospective TA allows us to use the eye-tracking equipment without a disrupted gaze while still getting access to the user’s thoughts.

A variant of the traditional concurrent TA protocol named "coaching" think-aloud protocol was introduced by Kato in 1986 [42]. This condition uses active intervention by having a test administrator acting as a coach who asks direct questions about different areas of the system and gives help or assists when a participant is struggling. Research has shown that users are more successful and satisfied with the product when using the coaching TA in comparison to the traditional one. This is an important aspect as it is desired to encourage participants to continue to use a complex system like PACS, especially since they have to in their daily work. Further, as many prefer learning in a social context the coaching TA can be considered
3. Theory

A natural way to get to know a complex system as informal collaborative learning in teams may be used in instances such as hospitals [43, 44, 45]. Finally, Mack and Robinson [46] summarize Kato’s word by stating that “question asking . . . creates opportunities for users to learn something about the system they are using”. This makes it a proper methods to use for identifying learnability issues when recruiting new customer as it’ll benefit both parties. Grossman et al. [22] considers the methods to be interesting for initial learnability as it highlightens areas of the system where learnability challenges are present. However, they also emphasize the fact that a disadvantage with the coaching TA-protocol is that the coach will only give input when the test participant is actually stuck.

Instead they propose a methods that they call the question-suggestion (QS) protocol, in which the expert can also freely provide advice to the user. In other words, in contrast to the coaching TA protocol the QS protocol gives suggestions when there’s room for improvement and not only when the user is stuck. Thus, the purpose of the QS protocol is to replicate a scenario where a user is performing the task next to a colleague and the colleague notices a usage behavior which could be improved upon. This term is called informal or "over the shoulder" learning and is a common way for users to learn in general [47, 43] and hospitals in particular [45, 44]. Further, by including suggestions into the protocol system evaluators are allowed to identify causes for suboptimal performance which then would indicate barriers for extended learnability [22]. Finally, in a study conducted by Grossman et al. [22] it was shown that the QS protocol found more learnability issues in comparison to the traditional TA protocol.

3.2.2 Semi-Structured Interview

As a complement to the TA protocol, a semi-structured interview will also be held. In contrast to methodologies such as TA protocol that are considered as direct usability tests, interviews are labeled as indirect usability tests as they shouldn’t be held without a direct usability test to refer back to [48]. Thus, the semi-structured interview will in my study be a methods that I hope to catch learnability issue that weren’t found when the user interacted with the system.

When interviewing people with very limited time, it’s important to get as much valuable data as possible without overextending the time limit. Regardless of whether or not the participant is considered an elite member of society or just an ordinary person, their time as your test participant is extremely valuable and should therefore be treated as such [49, 50, 51, 52, 53].

Guidelines

Goodwin [54] has summarized some general guidelines for interviews which I will briefly address below:

1. Make it a conversation, not an interrogation: Goodwin suggests that the interview should be loose, casual and laid back in order to promote an open and revealing discussion. As a result of this, any prepared question might
3. Theory

be inadequate after the initial testing.

2. Be sympathetic and non-judgmental: It should always be assumed that your interviewee is a good and capable person. If s/he has a negative attitude regarding work, it’s most likely due to a problem that I as a designer may solve.

3. Be the learner, not the expert: It’s important to establish a rapport with the interviewee before bringing up topics that may be touchy. The reason why this is important is because having someone look over ones shoulder can feel threatening, which is why adopting this mindset of being a learner will send reassuring signals that the participant is the respected expert. Spradley [55] describes this as encouraging elaboration by 'expressing ignorance and interest' more often that you would do in a typical conversation.

4. Ask naïve questions: A question that Goodwin [54] has often received during her work as a consultant working in complex domain such as healthcare is how a design team can possibly be effective in a field that they don’t know. The response to this is that ignorance is indeed bliss as interviewers who believe they know the industry or topic well often make assumptions about it that could be wrong. Further, as the design team doesn’t know what terms mean or how some processes are supposed to work they can ask 'dumb' questions which often reveal critical design insights.

5. Ask people to show you: As I’m going to conduct the interview in the same room as the testing, I may ask the test participants to show me rather than only verbalizing their thoughts about the system. As people self-reporting about their own behavior often generalize they tend to obscure or omit crucial details. Instead, if we see people in action we’ll be able to observe numerous things they’re unlikely to mention.

6. Take opportunities when they’re offered: Sometimes an interviewee refers to a particular person, process or thing that could be relevant to the design problem. When this happens it’s good to follow up on it by asking for more detail. However, if interrupting the interviewee would disturb the flow it could be better to make a note (mentally or in your notebook) and then bring it up after the current train of thought is completed.

7. Go beyond the product, but not beyond the design problem: When designing a complex system such as Sectra PACS it’s important not only to evaluate the program itself but also other activities around it. To exemplify, when designing an e-mail system it might be interesting to not only see how people create, view and organize their messages but also to see how the users’ communication also works outside of the system as those activities could be implemented into the design too.

8. Pay attention to nonverbal cues: Even though nonverbal cues might be difficult to analyse, it should give a hint about whether or not an answer actually is true or not. In an article published by Mehrabian [56] it was found that feelings and attitudes is only seven percent verbal, and that tone of voice and body language account for the remainder of the meaning.

9. Think ahead a little (but not too much): As mentioned in a previous section it’s important to be prepared and to treat your test participant’s time
carefully. Further, Goodwin states that if you’re the inquisitive sort it’s easy to get immersed in learning about the interviewee’s world. It’s however more important to focus on the information that you’ll need the most for the design later. As such it’s essential to understand processes, priorities, what type of information is used when etc.

**What not to do in user interviews**

This section will aim to briefly describe what not to do in user interviews according to some of the guidelines presented in Goodwin’s book [54].

1. **Don’t ask leading questions:** Asking leading questions is one of the worst mistakes you can do in an interview as it implies the answer you’re looking for. If I for instance would ask the question "Would you like to be able to chat with support within Sectra PACS?" A typical interviewee who’s trying to be polite and helpful might say, “Sure, that could be useful.” Even if the answer could be truthful you have no idea where it falls into the participant’s list of priorities as she might have 30 other things she’d rather do. Sometimes it’s difficult to avoid having leading questions, but a good guideline is to save those until the very end of the test session as you might get answers to during the interview.

2. **Avoid asking the interviewee for solutions:** As solving the problem is the designer’s job and not the informant’s the primary object is to gather information and not brainstorm ideas. Thus, it’s better to ask the interviewees something like: "If you had a magic tool, what would it help you accomplish?" The question doesn’t ask how the tool works but rather what problem it would solve and is therefore focused on goal rather than solutions.

### 3.2.3 System Usability Scale

As large and complex systems often have a lot of users and different user groups, it’s good to also collect quantitative subjective data as it can show differences between these groups. The most common way to do this is by having a questionnaire. The advantage of using a survey is that they’re time efficient to fill out, easy to analyse and may be used electronically. One of the most common questionnaires is the System Usability Scale (SUS). SUS was originally created by John Brooke in 1986 to give usability practitioners a "quick and dirty" methodological tool that would easily measure and compare usability in different contexts at low cost [57, 58]. Brooke defines SUS as "a simple, ten-item scale giving a global view of subjective assessments of usability". The 10 items are listed below with odd-numbered items worded positively and even-numbered items worded negatively.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use
3. Theory

this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Since its birth the questionnaire has undergone some minor changes. For instance, the word 'cumbersome' has been replaced with 'awkward' by most due to concerns regarding whether or not the majority of the population know what the word meant [59]. In addition to this it is also recommended to change the word "system" to a word that is suitable for the given context, such as product or website [60, 61, 59].

3.2.3.1 Amount of users required

As mentioned in a previous section, Nielsen suggests that you need at least 20 users for a quantitative data analysis [26]. While you need 2 users to actually use a questionnaire to measure variability (the standard deviation) Sauro claims that SUS will provide useful with as low as 5 users [62]. To prove this, Sauro [63] used several computer simulations which showed that when using a sample size of 5 the mean is within six points of a very large sample SUS score 50% of the time. Thus, if the actual SUS score was 74, average SUS scores using a sample of 5 users will fall in between 66 and 88 half of the time. Further, 75% of the time the difference between scores was 10 and 95% of the time by about 17 points. As such, Sauro argues that you get into the ballpark of the actual SUS score in more than half of the cases with small samples.

3.2.3.2 Using SUS

When using the traditional SUS, present the items on a 5 point-scales numbered from 1 (labeled with 'Strongly disagree') to 5 (labeled with 'Strongly agree'). If a participant fails to respond to an item it will be assigned a 3 (the center of the rating scale). After completion each item will generate a score between 0 and 4. Positively-worded items’ (1, 3, 5, 7 and 9) score are calculated by reducing their scale position by 1. Negatively-worded items (2, 4, 6, 8 and 10) are determined by reducing 5 with their scale position. The overall SUS score is measured by multiplying the sum of all scores by 2.5 which will produce a number between 0 and 100 in 2.5-point increments [60]. Sauro et al. explains the scoring of SUS by proposing two stages:

1. Subtract one from the odd numbered items and subtract the even numbered responses from 5. This scales all values from 0 to 4 (with four being the positive response).
2. Add up the scaled items and multiply by 2.5 (to convert the range of possible values from 0 to 100 instead of from 0 to 40).
3. Theory

While this is the general idea of the traditional SUS, other versions of the questionnaire has been proposed such as an all positive version which will be introduced in the next section.

3.2.3.3 Interpreting the results

SUS produces a number between 0 and 100 to give a hint of how good the usability of a given product or system is. As the questionnaire is not limited to a specific context it is possible to compare the score to the SUS’ validated database which consists of 5 grades A to F [64]. Anything below 51 would be considered an F or 'fail'. At this rank the system would most likely struggle too much in practice and should therefore not be launched before a bunch of its errors are fixed. Further, 68 is considered to be the median level of all SUS-scores. As such, anything under 68 is below average and therefore a score greater than 68 is above average. Sauro also states that a SUS of 74 has higher perceived usability than 74 percent of all products tested. This score would also fall into the B-grade interval. To get the highest grade however, you would need a score of at least 80.3. This would not only land your system into the grade A-interval, but is also believed to be at the level where the user will recommend the system to a friend.

There has been debated whether or not SUS should use a 5, 7 or even 10-point scale. In a study by Dawes [65] it was found that the 5- and 7-point scale scored higher in comparison to a 10-scale point. Further, in a study conducted by Matell et al. [66] the research team found that both reliability and validity are independent of the number of scale points used for Likert-type items. In contrast, Finstad [67] found that 7-point Likert items lead to a more accurate measure of a participant’s true evaluation and are more appropriate for both electronically-distributed and otherwise unsupervised usability questionnaires such as SUS. Finstad didn’t however present which condition scored the highest mean of SUS, meaning that using a 7-point scale might cause significantly lower or higher scores. Even though a truer evaluation of course is wanted, one of SUS’ key selling points is that it’s context free and has a validated database that practitioners may compare their scores to. As the database is built with traditional SUS-results, we can assume that most used a 5-point scale. Thus, if research later showed that using a 7-point scale will generate significantly lower results it’s unfair for the system to be compared to the database as they’ll score lower than a equivalent system. As a result of this I’ll be using a 5-point scale for SUS in my study.

Brooke refers to the methods as 'quick and dirty' because it is 'fairly quick and dirty to administer' [57, 58]. Sauro [64, 68] however agrees that SUS is by all means quick, but definitely not dirty. He emphasizes that SUS has data from over 5000 users and has been used in 500 different studies which proves its appreciation. Further, he suggests that its versatility, brevity and wide-usage means that despite inevitable changes in technology it is still being used. Finally he believes that SUS will still be around in 25 years due to its ability to be adapted to different areas of use.

While Brooke claims that SUS is dirty as it is easy to administer and Sauro says that it is not dirty because it produces good data, I believe that there could be a third definition of the term. Even though SUS gives us a good hint of whether or not the system we evaluate has good usability, it does not give us many clues as
to how we could improve it. The data SUS produces could therefore be considered shallow rather than dirty as might tell us there is a problem but not where nor how to fix it. In other words the data itself is not dirty but the story it tells could be considered as such. Further, recent research has however showed that while SUS does not explain the problem, it might point to what might be wrong. Ever since its origin in 1986 SUS has been assumed to be uni-dimensional [60, 68]. In 2009 Lewis and Sauro [68] however showed that SUS has two factors - usability (8 items) and learnability (2 items). Thus, if one would only want to focus on usability it would be possible to drop Items 4 and 10 to save time. This time benefit could however be traded into an even deeper analyse of the SUS data by using the two items to evaluate perceived learnability. To clarify further, practitioners can decompose the overall SUS score into its usability and learnability components with little additional effort. However, in a follow-up study by Borsci et al. [69] it was found that the uni-dimensional model and the two-factor model with uncorrelated factors proposed by Lewis and Sauro [68] had a unsatisfactory fit to the data. As a result of this they propose the hypothesis that usability and learnability are independent components of SUS ratings which they found to be significantly more appropriate to represent the structure of SUS ratings. Further, they propose that future usability studies should evaluate SUS according to the rule suggested by Lewis and Sauro as it’s the best fitting model which is shown in their work. Finally they call for future studies that research the circumstance where usability is dissociated from learnability as they found relative correlation between the two’s factors (i.e. systems with high learnability but low usability) which is an analysis that I’ll show in the results chapter.

An issue regarding learnability however is that Lewis and Sauro don’t define the term in their work. Items 4 and 10 are the ones that’ll evaluate learnability and are listed below:

4: I think that I would need the support of a technical person to be able to use this system.
10: I needed to learn a lot of things before I could get going with this system.

Their items however state that they seem to be interested in the user’s initial performance with the system. As such the items and therefore the learnability component of SUS seems appropriate for this study as I wish to evaluate the initial learnability of Sectra PACS.

3.2.3.4 All Positive SUS

As mention in the previous section, an all positive version of the traditional system usability scale has been proposed by Sauro et al. [61]. In this condition, the negative items has been changed to also be positive:

1. I think that I would like to use the website frequently.
2. I found the website to be simple.
3. I thought the website was easy to use.
4. I think that I could use the website without the support of a technical person.
5. I found the various functions in the website were well integrated.
6. I thought there was a lot of consistency in the website.
7. I would imagine that most people would learn to use the website very quickly.
8. I found the website very intuitive.
9. I felt very confident using the website.
10. I could use the website without having to learn anything new.

In a personal statement with Sauro et al., Brooke states that the advantages for alternating scales is to control acquiescence response bias (tendency to agree or disagree with survey items regardless of the items’ actual content) [61]. In addition to this it also provides protection against serial extreme responders whom may provide all high or all low ratings (when alternating the items it doesn’t make sense to respond all 1’s or 5’s). In contrast, if the items wouldn’t alternate a response of all 1’s could represent a legitimate set of responses.

Alternating the items may however have negative consequences. Bangor et al. [60] analysed a large database of SUS questionnaires with over 2300 items and found that participants tended to give slightly higher than average ratings to most of the negatively phrased items and in contrast slightly lower than average ratings to most of the positively phrased items. Thus, this suggests that participants tended to agree slightly more with negatively worded items and to disagree slightly more with positively worded items. In order words regardless of what the items state, there seems to be reason to believe that the item will be graded depending on its positive/negative label. Sauro et al. [61] argue that the traditional SUS has 3 major disadvantages:

1. Misinterpret: Reversing items between negative and positive may not account for the different responses that users give. Sauro et al. emphasizes that misinterpreting negative items include creating an artificial two-factor structure which would lower internal reliability.
2. Mistake: Users may forget to reserve their score and accidentally agree with a negative statement when they meant to disagree.
3. Miscode: Researchers might forget to reverse the scales when scoring, leading to incorrectly reported data. While many use software to calculate SUS, forgetting to reverse the scales is not an obvious error.

While miscoding should be easily avoided, research has in fact shown that it is a real problem. Sauro et al. found two sources of miscoding errors - the first from the Comparative Usability Evaluation-8 (CUE-8) workshop at the Usability Professionals Association annual conference [70] and the second from a study done by themselves where they examined 19 contributed SUS data sets [68].

With these two sources in mind, they state that three out of 27 SUS data sets (11.1 %) had negative items that weren’t reversed. With the assumption that the data sets are reasonable representatives of the larger SUS population, Sauro et al. assume that we can be 95% confident that miscoding affects between 3% and 28% of SUS data sets. They do however emphasize that they hope future research will shed light on whether their assumption is correct.
Additionally, there doesn’t seem to be any published concerns apart from Sauro et al. [61] about acquiesces bias in SUS. Further, they state that they couldn’t find any research documenting the magnitude of acquiescence bias in general nor whether it specifically affects the measurement of attitudes toward usability. In an attempt to find articles that may have been published since Sauro et al.’s article which came out 2009, I only found studies that focus on assessing or calculating acquiescence bias - not studying what impact it actually does have [71, 72, 73].

3.2.3.5 Additional Questionnaire Items

In addition to the items of the SUS questionnaire, Grossman et al. [22] suggests that a distinction of learnability needs to be made based on the assumption of the user skill. Indeed, as I agree with their view that learnability should be divided into an initial and an extended segment, it is important to gain knowledge of the user’s experience with the system. The authors propose that four relevant dimensions that should be used:

- Level of experience with computers
- Level of experience with interface
- Quality of domain knowledge
- Experience with similar software

The authors argue that the first three dimensions correspond to Nielsen’s categorization of user experience [5] while the fourth aims to account for designers interested in 'subsequent learning’. In subsequent learning, the user has perceptions about the capabilities of a system and how it may be used [25]. As this study will evaluate a group from a clinic that’s moving from one PACS to another, subsequent learning is an important aspect. The reason for this is that prior knowledge of the system’s possibility and use may aid in a learning subsequent package which will affect the ease at which a user may achieve specific performance. Further, as the first three items indeed correspond to Nielsen’s categorization of user experience one could draw the conclusion that if the user would rate them low, the user’s ability to achieve specific performance with the system would be reduced. In other words, a user with low computer experience would naturally have more difficulty learning a digital system in comparison to another user with high computer experience.

An issue is however that it’s difficult for the user to interpret how they should evaluate their own level of experience with computers for instance. On a scale 1 to 5 for instance, how good with computer must they be in order to feel comfortable with rating 5? One user could for instance rate themselves lower because they believe that a person that rates their computer experience a 5 is a very tech-savvy person that can code at a very high level. In contrast, another user might rate him- or herself a 5 simply because s/he feels comfortable with computers. Thus, it’s important to reflect on the data that’s being collected as the scales most likely will be rated differently.
3. Theory

3.2.3.6 Conclusion

While SUS isn’t context based and evaluates usability and learnability in any given area, Sauro [62] mentions that SUS isn’t always the best questionnaire as different jobs do require different tools. For instance, Sauro suggests the following:

1. Measuring website usability: Use the 13 item SUPR-Q questionnaire [74]. This questionnaire aims to measure perceptions of usability, trust & credibility, appearance, and loyalty for websites.
2. Measuring task-level usability: Use Single Ease Question [75]. In this condition, the users will rate how difficult or easy a task was to complete on a scale 1-7. Depending on the rating, it’s also encouraged to ask the user why they rated like they did.

Regardless, in a study conducted by Tullis et al. [76] SUS proved to be the most reliable tool for measuring website usability in comparison to the questionnaires listed below:

1. QUIS (Questionnaire for User Interface Satisfaction) [77]
2. CSUQ (Computer System Usability Questionnaire) [78]
3. Words (Adapted from Microsoft’s Product Reaction Cards) [79]

Another questionnaire has also been proposed by Zaharias et al. [80] which aims to evaluate e-learning usability by focusing on a more holistic learning experience as an alternative to SUS. In their article that was published in 2009 they however ended their work by calling for future studies that may validate their work, and in 2016 I couldn’t find an updated article from them that state that they consider their method as complete. Thus, as their methods didn’t come with a validated database it cannot compete with SUS yet.

For my specific purpose I want to use the all positive SUS as it both evaluates usability and learnability which is the goal for this study. As SUS is validated, proved to be reliable in comparison to other questionnaires and context-free it’s the best choice for my specific purpose in this study. Later on in the theory chapter I’ll also discuss how the SUS-grading may be used as a quantitative goal when using impact mapping.

3.3 Analyzing the data

This section will describe how the collected data will be analysed in order to produce results that may be used by Sectra to better understand their users and to increase their PACS’ learnability.

3.3.1 Persona

This section will introduce and motivate the usage of so called personas in usability studies.
3. Theory

3.3.1.1 Motivation

When conducting usability tests with a specific user group it’s good to gather knowledge about the specific user and create a so called persona. Goodwin [54] defines personas as archetypes that describe the various goals and observed behavior patterns among potential users and customers. Further, she explains that "personas are helpful in creating and iterating a design, building consensus, marketing the product, and even prioritizing bug fixes". By using personas it’s easier to encourage people to relate to users in uniquely human ways. Noessel [81] further suggests to avoid treating personas as creation (even though we all know that they are in fact creations) and instead treat them as real human beings.

The reason for this is that how we think about something changes depending on which type of agency we think is affecting it which is often referred to as Dennett’s theory of intentionality [82]. Noessel suggests that we when we design software we want to focus on our users and how they accomplish their goal which would translate into keeping an intentional stance in Dennett’s terms. When treating personas as persons rather than creations the design team might then find more interesting. As an example Noessel [81] talks about a scenario where he and his team decided to give Tracy, a persona, two kids rather than one. By discussing how many kids Tracy should have the conversation lead to them also talking about what tweaks they could do in order to make Tracy a better fit for the client’s expectations.

Thus, the persona should be presented in a lively story with as much relevant information as possible.

As Sectra PACS has many different user groups I want to be able to tell a story of what kind of user group I met during my user study, regardless of how many that might be. Further, in order to explain why my designs and possible solution to problems may be good I need to effectively show how it would help my persona fulfil her different goals. Thus, my findings will be used to create a persona to accomplish this and the process of doing so is explained in the next section.

3.3.1.2 Creating a persona

In order to create a persona data may be gathered from an interview. Goodwin uses 9 steps in order to create a persona:

1. Divide interviewees by role if necessary
2. Identify behavioral and demographic variables for each role
3. Map interviewees to variables
4. Identify and explain potential patterns
5. Capture patterns and define goals
6. Clarify distinctions and add detail
7. Fill in other persona types as needed
8. Group and prioritize personas
9. Develop narrative and other communication tools

These steps will be more carefully explained in the methods chapter where I use most of them in order to create my persona.
3.3.2 Focus groups

During this study I continuously presented my findings to Sectra’s User Experience (UX) group in order to get their feedback. The UX group worked as a focus group which may be defined as facilitated 60-90 minute meeting with anywhere from five to a dozen members of a target market (in this case Sectra products). The purpose of a focus group was to promote collaborative brainstorming [83] and may be used to identify needs, expectations and problems in a specific context [84, 85]. Rather than letting each participant respond to a question individually, focus groups are encouraged to speak with one another in order to generate data [86, 87, 84]. Thus, focus groups aim to collect qualitative data by interviewing people as a group rather than a group of individuals. While focus groups often are stimulating and fun for the participants, its primary purpose is to obtain data and it’s therefore important to stay on topic and avoid drifting away [88]. Even though the session should feel free-flowing and somewhat unstructured, it’s important for the moderator to follow a preplanned script of specific issues in order to collect the right kind of information. As I used a presentation during these focus group meetings I was be able to make it easier to stay on topic by using so called trigger pictures. A trigger picture is a picture which aims to help the group gain common ground and reach a shared understanding [89]. They may for instance be used in problem-based learning cases within medicine as a way to enhance learning by promoting visual cues for students [90]. Further, research has shown that even subtle exposure of certain picture may affect us cognitively, suggesting that just having a picture to refer to during a focus group may help the group discuss things [91].

When conducting focus groups, the rule of thumbs favors using strangers due to their ability to rely on the kind of taken-for-granted assumptions that the researcher is trying to investigate [92]. However, as my purpose of this study was to assess and improve Sectra PACS learnability, I wanted to bring together Sectra staff as their knowledge is the most valuable I possibly could find. The result of this is that there won’t be any strangers and instead colleagues. While strangers might be the rule of thumbs it’s certainly not a requirement to form a focus group [93]. Social scientists often conduct focus groups in organizations and other naturally occurring groups where acquaintanceship is unavoidable [94]. Instead, it’s suggested that working with prior acquaintances may help the researcher deal with issues of self-disclosure [95]. Thus, the underline is that stranger and acquaintances can generate different group dynamics which is why researchers have to make choices depending on the nature of the research goals. As I evaluated a Sectra product from a developer’s view even though I’m not an employee it would seem natural to form a focus group with actual developers as I wanted their feedback on my work and findings. From a usability perspective, using focus group may be a poor methods for evaluating interfaces as no amount of subjective preferences will make a product viable if users can’t use it [96]. Instead, a focus group session focusing on an interface should rather discuss why issues are issues and how they could be solved. As my study focused on initial learnability, the natural subject for my focus group sessions was how said learnability issues could be solved. The exact subject of each focus group session will be explained and presented in the methods and result chapters. By conducting a focus
group within the organisation in which the evaluated system has been created I get a unique window into the thought process behind said product. This makes it possible for me to collect data which wouldn’t been possible to gather without that specific focus group as I get access to years of extremely valuable information straight from the systems creators. Thus, conducting a focus group within the organisation that owns the product that’s being evaluated makes it possible to get unique information with little effort.

Finally, the analysis of a focus group may be very simple and its results can be treated as any other qualitative data such as summarizing notes taken from the session [96, 54, 86]. In an upcoming theory section I’ll explain how qualitative data can be modeled and in the methods chapter I’ll show how I did it.

### 3.3.3 Impact Mapping

This section will shortly introduce and motivate the use of impact mapping in usability studies.

#### 3.3.3.1 Introduction

Impact mapping is a synthesized (agile) methods that has been stitched together by Adzic [97, 98, 99, 100, 101]. Adzic [97] defines impact mapping as a strategic planning technique which prevents organisations from getting lost while building products and delivering projects by clearly communicating assumptions, helping teams align their activities with overall business objectives and make better roadmap decisions. He lists three unique advantages that impact mapping has over similar methods:

1. It’s based on a method invented by an interaction design agency and is similar to a team-building method and therefore facilitates collaboration and interaction. It’s less bureaucratic and easier to apply in comparison to many alternatives. Further, it facilitates the participation of groups that are built with people from different backgrounds such as technical delivery experts, business users while helping organisations use the wisdom of crowds.
2. It visualises assumptions. Adzic emphasizes that alternative models usually doesn’t communicate assumptions clearly. Further, the visual nature of impact mapping promotes effective meetings and supports big-picture thinking.
3. It’s fast and should therefore a nice fit with iterative delivery models that are becomming more common in software industries.

Impact mapping should therefore be easily integrated into organisations that use agile methodologies. Sectra for instance applies SCRUM today but isn’t using impact mapping as a go to tool.

#### 3.3.3.2 Levels

Impact mapping is built around four keystones which are supposed to be filled out by the user or team and may look like figure 3.3.
1. **Goal (The why?):** The goal in impact mapping is the goal we want to achieve and could be phrased as "Why are we doing this?". As such, goals are supposed to be "S.M.A.R.T." - Specific, Measurable, Action-oriented, Realistic and Timely. Further, goals shouldn’t be about building products or delivering project scopes but instead explain it might be useful. In addition to this goals should present the problem that we want to solve rather than the solution. Avoid design constraints might for instance be a good goal definition.

2. **Actors (The who?):** The actor is the first level of the impact map which should provide the answer to the following questions: Whose behaviour do we want to impact? Who can produce the desired effect? Who can obstruct it? Who are the consumers or users of our product? Who will be impacted by it? As such, Adzic refers to these as the actors who can influence the outcome. Further, there are three kinds of actors that are listed below as mentioned by Adzic:

   (a) Primary actors, whose goals are fulfilled, for example players of a gaming system
   (b) Secondary actors, who provide services, for example the fraud prevention team
   (c) Off-stage actors, who have an interest in the behaviours, but are not directly benefiting or providing a service, for example regulators or senior decision-makers

Finally, the actors should be as well defined as possible and it’s recommended to avoid generic terms such as "users" as different categories of users might
3. Theory

have different needs.

3. Impacts (The how?) The second level of an impact map sets the actors in the perspective of the chosen business goal. It should answer the following questions: How should our actors’ behaviour change? How can they help us to achieve the goal? How can they obstruct or prevent us from succeeding? In addition to this it’s important to avoid listing everything an actor might want to achieve. Instead only the impacts that really help the organisation move towards the central goal should be listed. Further, impacts are not product features. Therefore listing software ideas should be avoided and the focus should instead lay on business activities. If possible it’s also encouraged to show how a change in actor behaviour rather than just the behaviour. The reason for this is to show how the activity is different from what’s currently possible. Finally it’s good to both consider negative and hindering impacts as well as positive ones as well as to think about what else an actor could do after discovering their first impact.

4. Deliverables (The what?): The third and final level of an impact map answers the following question: What can we do, as an organisation or a delivery team, to support the required impacts? These are the deliverables, software features and organisational activities. Adzic states that this is the least important level of an impact map. The reason for this is that this should be an iterative process and may be refined during the process. Further, the deliverables should be treated as options and not something must be delivered.

3.4 Conclusion

The theory presented in this chapter has shown that learnability may be divided into initial- and extended learnability. In addition to this it’s important to define learnability when presenting research as it will promote the study’s validity and reliability. Further, depending on which kind of learnability that it’s desired to assess, appropriate learnability metrics should be used. However, as mentioned in the introduction I couldn’t find any research studying initial learnability in complex systems during my literature study which didn’t give me much previous work on this specific area to explore. Instead, methodologies such as the question-suggestion (QS) protocol has been thoroughly presented along with the all positive system usability scale questionnaire which both may be used to assess initial learnability in general and could be applicable within complex systems too. Finally, the theory that has been presented were of interested to Sectra as they’re also interested in alternative learnability metrics that I haven’t used in this particular study too.
3. Theory
Methods

This chapter will introduce and argue for the process chain of methods used in this study which assessed and proposed improvements for initial learnability in Sectra PACS.

4.1 Pre-study

This section will describe the process before collecting data during the test-session.

4.1.1 Literature study

The literature used in this study was found by first using the search term "learnability" in Google scholar. As a student at Linköping University I had access to almost all articles found on Google scholar which made it easy for me to get the publications that I wanted. After finding a good general learnability article I searched on the used references in it in order to find more articles. Further, I also searched for the most recent articles published within the main theoretical themes in my study in order to see if for instance any major learnability finding has been made recently. I kept this process going until I was satisfied with the theory that I could write. In order to confirm that there hadn’t been any study focusing on initial learnability in complex medical systems I tried using a lot of different combinations of search words, and stopped when further combinations didn’t give any new results.

4.1.2 Getting to know the system

The first part of the study aimed to get to know the system on a general level personally. As I’m going to be the coach during the question-asking think-aloud (TA) protocol session it’s important that I’m confident with what I’m showing. In order to help me get good knowledge I was handed training material for application specialist tailored for their PACS which consisted of an introduction, tasks and general information of the system.

During this introduction with the system I also had discussion with Sectra’s user experience (UX)-team about limiting the evaluation of PACS as the system is too large to study as a whole.

They suggested that I should focus on the worklist- and demonstration segment of PACS as it’s used by the most user groups. The reason why it’s good to have a large target group to recruit is because even though Sectra is a well-known and
4. Methods

appreciated company it’s difficult to set up meetings with their users as hospital staff often are busy. Thus, having a larger target group makes it easier to find possible test participants.
In addition to the training material I was personally introduced to the system by a UX designer that had also worked as an application specialist at Sectra’s 2nd line support whom also acted as my supervisor during this work. Further, a second personal presentation was held together with my supervisor and an application specialist that worked solely with Sectra PACS. These meetings gave me a good insight of how PACS works on a general level as well as how the worklist and demonstration segment works on a specific level.

4.1.3 Recruiting users

As this study aimed to assess and evaluate initial learnability it was important to recruit users that had little to no experience with Sectra PACS. This is not an easy task as hospitals do not change their IT-systems often, and when they do it is usually a long process. Luckily a new Sectra customer had recently bought their PACS and were interested in participating in this study as a way to both learn how to use the system and contribute to its improvements. The group was recruited when they visited Sectra for a crash course in how to use PACS. With the help of the application specialists who held the training I got the opportunity to shortly introduce myself and my study. The group was recruited on the premise that they would get an opportunity to have an additional training session while also contributing to hopefully making Sectra PACS easier to use by improving its learnability.

4.1.4 Defining the tasks

As mentioned in the theory chapter, the question-suggestion TA-protocol requires the test participant to think aloud while completing a set of tasks. As I was going to be the coach during the QS protocol I needed to learn the specific segment that I was going to evaluate. Thus, I got a personal introduction to it by an application specialist whom also helped me to from the tasks which would be relevant for someone working as a radiographer or assistant nurse. As a result of this 22 tasks were defined which were limited to the worklists and demonstrations. The ambition was to find as many relevant tasks that will be used in their daily job as possible. In order to confirm that my tasks were relevant I once again talked to my supervisor and the application specialist who held the PACS presentation for me. After this, some tasks were removed while others were added. This iterative process was finished when both my supervisor and the application specialist found my list of tasks sufficient and relevant.

4.1.5 Pilot Study

To test and potentially adjust the design of this study before the actual session, a pilot study was constructed. As the main focus for this study was to assess
learnability, two participants that had never worked with PACS were recruited. Further, it was desired that the participant would have some medicine knowledge in order to easier grasp the concept of a PACS. As such, a biomedical engineering student and an assistant nurse with 7 years of experience in different areas that is currently studying medicine were recruited for the pilot. Some tasks were adjusted after the pilot to make the flow as good as possible, such as creating a static list rather than finding an empty work list in the test environment. Finally, the pilot confirmed that the design was efficient, within the time-limit and gave appropriate data for the purpose of this study.

4.2 Design

This section will describe the design used in this study which took approximately an hour per participant.

4.2.1 Apparatus

A test environment of an English version of Sectra PACS 18.1 was setup using a MacBook Pro Retina 13” running Windows 10 with an external Windows-based keyboard and mouse. The test sessions were video recorded using an iPhone 6 on a tripod while the interviews were only recorded by audio. The iPhone 6 captured the screen as seen in figure 4.2.

Figure 4.1: Picture of the apparatus setup.
4. Methods

Figure 4.2: Still picture from the iPhone 6 recording.

4.2.2 Procedure Overview

During the test session, three methodologies were used:

1. Question-Suggestion Protocol
2. Semi-structured Interview
3. All Positive System Usability Scale

The usage of these methods will be more carefully explained and motivated later in this chapter. Further, each test session took approximately an hour to complete. In order to be as effective and efficient during each session conducting a semi-structured interview was considered a good choice as it’s flexible and doesn’t need a specific amount of time. To clarify, if the user would need 40 minutes to complete the tasks during the QS protocol, the interview would last for approximately 10 minutes in order to save 10 minutes for the questionnaire which will be presented in the next section. In contrast, if the user would complete the tasks in 25 minutes the interview would be able to last for 25 minutes which would still save 10 minutes for the questionnaire. The procedure will be explained in whole in the following sections.

4.2.3 Introduction

The participants were greeted when they entered the testing room and asked to carefully read a consent form which described the purpose of the study, what data would be collected and that all material would be treated completely anonymously and can be found in the appendix. In addition to the consent form, the participants were also given verbal information of the study’s purpose while it was also stated that any mistakes or issues that could happen during testing wasn’t their fault but mine as a tester.
4.2.4 Question-Suggestion Protocol

As mentioned in the theory chapter, the question-suggestion protocol seems to be the best option available when aiming to assess initial (and extended) learnability when using subjective metrics according to Grossman et al. [22]. In order to help the user understand how to think aloud in general, it has been suggested to show a short demo-video before the session begins [36]. By showing rather than only explaining how the TA protocol works, users understand the concept of thinking aloud better and will generate better data. I used the sample video made by Nielsen [36] as it didn’t evaluate a system which was similar to Sectra PACS as it wouldn’t bias the user in some way (link to the video can be found via the reference). As I wanted to use the Question-Suggestion protocol I explained that we’d to the tasks together as a team, where I’d be the coach that the test participant could ask for assistance whenever s/he felt that it was needed. In addition to this I gave the test participants suggestions when they struggled rather than giving them the solution. During the test I used the same instructions for both the participant and the coach (myself) as Grossman et al. [22] presented in their article:

**Question-Suggestion protocol instructions to participant:**

1. Ask relatively specific, procedural questions.
2. Try to answer your own questions first, but do not engage in extensive problem solving.
3. Focus on getting the task done, as you would in the real world.

**Question-Suggestion protocol instructions to coach:**

1. Reply with specific procedural answers, to the underlying form of the question.
2. Do not tutor, or explain at length.
3. Maintain focus on the task, even when providing suggestions.
4. Provide suggestions if you notice inefficient usage behaviors, and if the suggestions would likely be welcomed and beneficial to the user.

After the introduction the participants were presented with a task verbally with a short scenario which aimed to describe how the task would be relevant for their daily work. In addition to this the task was also printed on paper which was laid in front of the participant so that they could look at it during the task as shown in figure 4.3.

![Image of task 5 printed on paper](image)

**Figure 4.3:** Picture of task 5 printed on paper.
4. Methods

4.2.5 Motivation

While the traditional TA protocol is one of the most successfully used usability tools, research has shown that the question suggestion (QS) protocol is even better suited for identifying areas where learnability is low and learnability issues in general [22]. Thus, using the QS protocol together with its tasks seems to be the best available options when using subjective usability metrics for evaluating initial learnability. The chosen tasks will be presented in the next section.

4.2.6 Tasks

The chosen tasks will be introduced and motivated in this section. Further, the tasks were chosen together with a UX designer with previous experience working as part of Sectra’s 2nd line support and an application specialist in order to everything that’s relevant for a test administrator to do within worklists and demonstrations in the PACS.

**Task 1:** Locate the area in which it is possible to create worklists and create a static list.

Motivation: One of the primary jobs as a radiographer/nurse is to create worklists. Thus, finding where to do so is an important part of the job.

**Task 2:** Find a worklist with more than 5 patients.

Motivation: This task was chosen as the participant will browse different worklist as they work. In addition to this we wanted to see how the patients would navigate between lists.

**Task 3:** Find an empty worklist.

Motivation: In addition to wanting to see the participants explore the system as mentioned in task 2, I wanted to see whether or not the patients would try to find the list that they had created in task 1 or if they would continue to scan other lists.

**Task 4:** Add two patients to the worklist in two different ways (one way per patient).

Motivation: Adding patients to different lists is another feature that radiographers/nurses will do often while working with worklists. There are two different ways to add a patient to a worklist and this task will see which one works the best.

**Task 5:** Make the list one of your favourites.

Motivation: By using favourite groups the user may save worklists or demonstration to make them easier to fetch.
Task 6: Create a demonstration.

Motivation: Creating demonstrations is another task that radiographers/nurses will do often.

Task 7: Make the demonstration one of your favourites.

Motivation: In addition to easier being able to fetch worklists, this question is added again to see if the users will complete the task with more ease than last time.

Task 8: Locate the place where your favourites are saved.

Motivation: This task was chosen as the users might know how to add lists to favourite groups but can’t find where they are located.

Task 9: Add the two patients from your worklist to the demonstration in two different ways (one way per patient).

Motivation: This task was chosen as the participants have to add patients to their demonstration too.

Task 10: Change day and time for the demonstration.

Motivation: As a demonstration is a scheduled event it’s important to know how to change its day and time if need be.

Task 11: Show in two ways how to switch between patients in a worklist.

Solution: Sectra PACS allows the user to browse between patients (and examinations) in different ways and is part of a radiographer/nurse daily work.

Task 12: State that both your patients have been demoed on the demonstration.

Motivation: After a patient has been demoed during a demonstration the radiographer/nurse has to tell the system so.

Task 13: Search for <name shown on screen>.

Motivation: Being able to search for patients is a good way to quickly find someone rather than clicking through all worklists. The name that the patient would search for depended on which patients that was present on the screen at the moment.

Task 14: Change the status of the patient to "prioritized".

Motivation: Changing status of a patient is something that a radiographer/nurse do in their daily work.
Task 15: Merge the patient with the other patient in the list.

Motivation: Sometime during emergencies a patient is given a temporary ID in the system. The temporary ID then as to be merged with the patients real ID in order to keep his/her examination list up to date.

Task 16: Change the status of the patient to "finished".

Motivation: This task was picked in order to see if the patients remembered how they completed task 14.

Task 17: Go to a list with more than 10 patients.

Motivation: This task was chosen to give the test participant a large worklist to complete the next tasks in.

Task 18: Sort the list by date of birth.

Motivation: Clicking columns in Sectra PACS is a fast way to sort patients by different parameter's, such as date of birth.

Task 19: Add "priority" to columns".

Motivation: If wanted, the user may add new columns in order to both get new information in their overview while also being able sort by the same information

Task 20: Reset the list settings.

Motivation: Being able to reset the list settings is a quick way to make the list settings go back to default.

Task 21: Use the list settings for all worklists.

Motivation: If the user for instance has added columns that s/he always want to see regardless of which worklist is being active there’s a feature that applies the current list settings to all worklist.

Task 22: Filter the list after a body part.

Motivation: Being able to filter lists after different keywords is a great way to quickly find a specific examination or a cluster of similar examinations.
4. Methods

4.2.7 Interview

The primary purpose of the interview was to find needs and catch issues that might not have been found during the think-aloud session. In addition to this, the interview also aimed to get feedback on the training sessions that had been held for four out of five participants at Sectra in Linköping.

In order to direct the interviews toward the same subjects, a short interview guide was formed which consisted of the following questions:

1. How do you think the test went?
2. What was easy during testing?
3. What was difficult during testing?
4. How do you normally learn a new system?
5. What do you think about the training sessions at Sectra?
6. If you could change one thing with Sectra PACS, what would you change?

These questions were chosen as they would for instance generate general feedback about the testing, but mainly to catch issues and needs outside of the PACS which the QS protocol itself wouldn’t be able to generate.

During the interview, notes were also taken accordingly to the suggestions made by Goodwin [54] presented in the theory chapter in order to remember follow-up questions and better keep track of the session in general. As mentioned earlier, the interviews were audio recorded as I wanted to relisten to them later on during the data analysis.

4.2.7.1 Motivation

By having a semi-structured interview after the QA protocol, the user would be able to further explain specific problems that they encountered during testing while also being able to speak about general learning at a meta level.

4.2.8 Questionnaire

The original all positive SUS-questionnaire by Sauro was translated to Swedish for this study. In order to make the terminology as pleasant as possible, the translation aimed to produce natural sounding questions without losing its original purpose while doing so. Three additional questions were added in order to get the subjective measurement of the participant’s computer skills, experience with Sectra PACS and similar systems according to the suggestions by Grossman et al. [22]. As mentioned in the theory chapter, it was decided to keep the 5-point scale as the consequences of using a 7-point scale were unclear. During the time that the test participants filled out the questionnaire, I was available in order to explain a question if they didn’t understand it completely.
4. Methods

4.2.8.1 Motivation

The SUS questionnaire is a quick way to get an overview of a system’s usability. Further, as it isn’t limited to a specific context and has a validated database it’s easy to get a good overview of how good a system is in comparison to others. In addition to this SUS was the only questionnaire with a validated database which also calculates a system’s learnability which of course is important for my study. Further, using the all positive version of the questionnaire seems to be an obvious choice as research hasn’t showed any reason not to do so. Thus, by adding the three additional questions by Grossman et al. [22] an even deeper learnability evaluation may be quickly conducted by using a 13 item questionnaire. Finally, the SUS score will also be used as the goal to achieve in the impact map that is explained in an upcoming section.

4.2.9 Users

The group of participants consisted of 4 test administrators (2 assistant nurses and 2 radiographers) and one end user all working at a mammography clinic at a hospital in Sweden. In addition to administrating the PACS, the test administrators were also responsible for educating their clinic how to use Sectra PACS. As such, these users need to know the system well enough to be able to teach their colleagues. The additional end user was recruited in order to make it a total of 5 users accordingly to the suggestion by Nielsen [3, 26]. As previously mentioned the test participants had been at Sectra for their crash course in how to use the system. The course was divided into two weeks: During the first week, only the four test administrators attended as the session aimed to describe how the PACS environment should be set up from a back-end perspective. During the second week the test administrators came back with so called 'super users’ whom are responsible together with test administrators to teach their peers how to use the system. As such, 4 of the 5 participants had some limited experience with PACS as they had been attending the two weeks of training while the 5th had never seen the system before. It’s however important to once again emphasize that the PACS is a very large system and that even though the participants have seen the system before they haven’t used or even seen all its functionality in my particular case.

Finally, users rated their different computer experience on a scale 1-5 as following:

1. Experience with computers: 3.6
2. Familiarity with Sectra PACS: 2.4
3. Familiarity with similar systems: 3.4

The data seemingly shows that the recruited users consider their experience with computers decent overall. What’s interesting however is that they rated their familiarity with Sectra PACS fairly high considering that they haven’t used the system as part of their actual work yet. Further, the familiarity difference between Sectra PACS and similar systems was fairly low. Thus, one can draw the conclusion that the data collected is difficult to interpret as no labels have been used which is a setback as it’ll be difficult to any deeper data analysis. However, what the data does
reveal is the users’ perceived experience and familiarity with Sectra PACS and similar systems which tells us that they don’t consider themselves to be extremely bad nor good with computers, and that they’re not completely unfamiliar with PACS but definitely not experts. I’ll discuss these findings more thoroughly in the result chapter and also discuss how they may be used to evaluate subsequent learning.

4.3 Data analysis

This section will describe how the collected data was analysed during this project.

4.3.1 Question-asking protocol

The Question-asking protocol data was analysed by watching and listening to each video recording while taking notes and extracting valuable quotes until no new data was found. As mentioned earlier a learnability issue is an issue that would make it difficult for the user to in this case complete the tasks. Thus, what was documented was how and why the user was struggling. Further, potential needs were also written down as well as uncertainties.

4.3.2 Interview

The interview data was analysed by listening to each session and writing down essential findings. In contrast to the question-asking protocol, the interview focused on catching issues and potential needs that weren’t found when completing the tasks. Further, as the interview guide also aimed to bring up how the learning sessions at Sectra had worked the purpose wasn’t only to find learnability issues.

4.3.3 All Positive System Usability Scale

The questionnaire was calculated accordingly to the steps described in the theory section by using IBM SPSS 24 and re-calculating the score in order to make sure they’re correct. Further, the learnability items were calculated by adding item 4 and 10 together and then multiplying their score by 10.

4.3.4 Clustering the data

After all data had been analysed from each source individually, I wrote down issues and needs on post-its where each color represented an identified, broad theme which was used to find larger groups of problems or needs that may be solved as a whole and is shown in figure 4.4. The colors represented the following:

- Blue = learning
- Orange = interaction issues
- Green = visual issues
- Pink = other
In order to get a second pair of eyes on my findings a Sectra UX designer assisted me during the final clustering session. First, we used the colored post-its to create groups and wrote down themes next to them as shown on figure 4.5.

Figure 4.4: Initial clustering by color.

Figure 4.5: First notes added next to the post-its together with a UX designer.
Secondly, we placed all post-its into a group labeled with a description of what they aimed at which is showed in figure 4.6 after discussing the themes labeled to the right.

<table>
<thead>
<tr>
<th>Solved Function</th>
<th>Terminology</th>
<th>Drag &amp; Drop</th>
<th>Planning/Beaver Region</th>
<th>Planarizing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.6:** First clustering of post-its.

Finally, we segmented the clusters once again to find as precise groups as possible in figure 4.7. The clusters will be presented as a whole in the result section.

<table>
<thead>
<tr>
<th>Solved Function</th>
<th>Terminology</th>
<th>Drag &amp; Drop</th>
<th>Planning/Beaver Region</th>
<th>Planarizing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.7:** Final clustering.
4. Methods

4.3.5 Persona

A persona was created based mainly on the collected data from the semi-structured and named Ruby Radiographer. She was made with Goodwin’s [54] suggestions in mind by mapping out themes in all 5 participants data. This was done by first re-listening to all the interviews until no new information was found, and then mapping the data into larger text files where it was then possible to find common themes, needs and goals. Further, the purpose of the persona is to make it easier to describe what the users that I met actually want to achieve. Therefore, having a persona is an effective way to present user data as a lively story.

4.3.6 Focus group meetings

After collecting my data, I conducted three focus group meetings with the UX team in order to get feedback on my results as well as my possible solutions. This section will describe this process.

4.3.6.1 Interview regarding learning

As the UX team doesn’t have anything to do with the training sessions that Sectra provide for new customers, it was wanted to get feedback on their thoughts regarding these courses. As a result of this, the application specialist that helped me with the recruitment of the test group was interviewed in order to get feedback regarding how they form their learning sessions as well as get the opportunity to give them feedback too.

4.3.6.2 Focus groups

Finally, three focus group meetings were held. The purpose of these focus group meetings was to get feedback on my work as well as discuss problems and solutions. All three sessions used a presentation made with Keynote which included pictures and videos in order to get the discussion going as well as properly showing what I’ve found. The videos that were showed were interactive solutions made using Adobe Muse and recorded with QuickTime. During the two first sessions, all participants were writing down notes while the third session was sound-recorded with additional note taking from me as the moderator. The sessions are listed below:

Session 1, findings: This session presented the study as a whole before the data analysis.

Session 2, analysis: The second session gave a deeper description of how the data was analysed as well as potential solutions to the found problems.

Session 3, learning: The final session only focused on learning in general and both discussed of learnability can be improved within the PACS but also outside.
4.3.6.3 Motivation

The focus group meetings made it possible for me to get access to extremely valuable information by letting experts discuss problems and solutions together.

4.3.7 Impact Mapping

This section will describe how impact mapping was used in this study as the agile software development tool which I visualized using the program *astah professional*.

**Defining the goal**

The step of the process was to define the impact mapping goal. As the goal must be measureable, I chose to achieve the next grade of my calculated SUS as my impact map’s goal - for instance, if the PACS achieved the B-grade of SUS, my goal would be to reach the A-grade.

**Defining the actors**

As SUS measures the perceived usability from the user, my primary actor is the test administrators as they are my target group for this study. Further, the secondary actors are the actors who provide service which in this case mainly are Sectra’s training personnel. In contrast, the off-stage actors are defined as the actors who have an interest in the behaviours but aren’t directly benefiting or providing a service and were in this case the Sectra developers.

**Defining the impacts**

The impacts, or what behaviour change we’d like to achieve, was found by analyzing the data previously collected after the clustering had been done. The impacts were chosen in order to solve the found issues and needs which had been presented during the first focus group meeting and then discussed during the second one. To clarify, I searched in my data to find ways that the defined actors’ behaviour potentially could change in order for test participants issues and needs to be solved which then would push the SUS rating up to the C-grade which was the goal.

**Defining the deliverables**

The deliverables were find by both analyzing the data previously gathered and by discussing it during the focus group meetings previously mentioned. The deliverables would aim to help the actors achieve their impacts. As not all deliverables in the impact map must be solved or even possible at the current time, all solutions that seems reasonable were put into the map. Thus, the deliverables were to be treated as possible solutions that together could support the actor’s to change their behaviour which would make it possible to reach the goal of the map.
4. Methods

Motivation

Impact mapping is a great way to visualize what we want to achieve, whose behaviour we need to change in order to achieve our goal, and how we as an organisation may aid our actors’ change of behaviour.

4.4 Conclusion

In this chapter I’ve explained how I’ve gathered and analysed my data as well as showed how the results may be implemented into an impact map. Further, the methods used to gather the data were all designed to assess initial learnability both within Sectra PACS and during its training sessions. In addition to this, the methods used to analyse the data were chosen in order to be able to produce deliverables which could be implemented into an organisation’s agile methods. Additionally, the used methods didn’t require a lot of resources and were confirmed to be something that Sectra would consider to use in their organisation. Furthermore, the methods used for data analysis worked well in order to find possible ways to improve the initial learnability in Sectra PACS which was confirmed during the focus group meetings as Sectra found the possible solutions to be of interest. Thus, the set of methods used in this chapter should be applicable when evaluating any complex systems as they aren’t tailored specifically for a PACS within medical IT. In addition to this, the methods in my process chain aren’t context based at all nor require a lot of resources to conduct. I do however think that in order to produce good results, it’s important to have direct contact with the organisation that owns the product in order to produce good tasks that are relevant for the target user group as well as being able to teach someone to act as a coach during the QS protocol. Finally, the process chain is an academical contribution as it has produced a lot of rich data from a complex system without showing any indication that it’s limited to this specific context nor require a great lot of resources to use.
5

Results

In this chapter the results will be presented. First, the persona will be introduced in order to give a good view of the users that I met. Secondly, the segments consisting of problems and needs will be presented in clusters. Finally, impact mapping will be used to show how found issues may be solved.

5.1 Persona: Ruby Radiographer

Ruby Radiographer is mainly based on the data gathered from the semi-structured interview with some additional data that was collected during the QS protocol.

Forty-six-year-old Ruby really enjoys her job as a radiographer at a mammography clinic, but sometimes she feels limited in her work environment. There’s a lot of patients that needs her assistance and she wishes that she could spend more time helping them rather than struggling with their IT-system. Even though she’s been using the same system for quite some time, it simply doesn’t allow her to be as efficient as she wishes to be. She’s been told that the system has been created by
5. Results

A pretty large company and that her clinic has gotten a customized version of it specifically made for mammography. The creators of this software has however quit which means that there’s no support at all. Ruby’s not an IT-guru but she does consider herself to be pretty decent with computers. Together with her colleagues she has for instance created ways to avoid issues and help each other out, such as hand-written manuals to remember how to complete certain tasks. This hasn’t however worked out too well which is why the clinic is now changing IT-systems. Ruby’s excited by this as she really doesn’t think that it may get worse than their current situation.

Even though she’s thrilled about the new system she’s also concerned. Her time is limited and she remembers how she struggled to understand their current software, and how much time she and her colleagues had to spend helping each other out. In order to make the transition between systems as good as possible the clinic has therefore decided to pay extra for two weeks of training at the new provider’s facilities. As Ruby will be one of the few persons who’re responsible for educating the rest of the clinic she’s calmed by the fact that she’ll get to spend time getting to know the system thoroughly - especially since she’ll be administrating it for the clinic too. After the first week of training she’s however not feeling confident at all. She’s been learning how to set up environments and getting to know which parameters does what. While she understands that what she’s doing is important she has trouble visualising it. Instead, she thinks that the system feels like a jungle. It’s not until the second week where she gets to know the system’s flow that she truly understand the impact of the first week. Ruby doesn’t however know if she’d like to do it the other way around, and instead believes that the application specialists are doing this for a reason.

What she does know is that she’s getting somewhat annoyed at the inaccurate information that was in the work material which they used during the courses. It might be minor things, but it’s distracting and doesn’t feel like real-life examples. In addition to this Ruby likes to complete things that she’s started (she’d for instance never stop reading a book in the middle of a chapter) because it gives her a sense of satisfaction. In a perfect world she’d like to get to know this new system like she reads a book - chapter by chapter. She was therefore surprised when she was instructed to quote on quote "just practice a little". Practice what? Instead she wants instructions and templates to work with. Overall she however enjoyed the course and feels really good about the system and looking forward to getting access to the test environment that they’ll soon have.

Goals:

1. Learn the essentials of the system as quick as possible
2. Learn little by little
3. Be able to use a resettable test environment
5.2 Tasks that were completed with ease

This section will list the tasks that the users completed with ease before we move along to the found issues and needs.

- **Task 16**: Change the status of the patient to 'finished'.
- **Task 18**: Sort the list by date of birth.
- **Task 20**: Reset the list settings.
- **Task 21**: Use the list settings for all worklists.

Task 16 was a variant of another task which also was completed by changing the status of a patient. Further, some tasks were re-used as mentioned in the methods chapter as a way to see how well the patients would remember how to complete them. Thus, when the task was completely the same (such as task 16) the users completed it with ease, but if a task was almost but not exactly the same they got confused which will be elaborated in the next section. Task 18, 20 and 21 were all solved by using columns which the users were confident doing. In addition to this, task 18, 20 and 21 were all some of the last tasks which means that the users had completed 17 other ones and therefore knew more about the system. A quote that might give more insight is provided below:

“If you’re familiar with the Office package you know that most things are usually hidden behind a right click”

5.3 Clusters

This section will describe the clusters that were formed and presented in a previous section in the methods chapter together with a UX designer from Sectra. The clusters were based on both the question-asking protocol- and interview data. First, the cluster will be described. Secondly, the issues and needs will be explained. The solutions that will be presented has been confirmed as viable options by the Sectra UX team during the focus group sessions. The cluster segment that will be presented include:

1. Terminology
2. Information overload
3. Missing features
4. Confusion, needs explanation
5. Flow problems
6. Feature placement
7. System error management
8. Learning

Finally, the term 'learnability issue' will be defined as an issue that will obstruct the users ability to eventually achieve specific performance - in other words, complete
5. Results

their daily work tasks.

5.3.1 Terminology

This section will describe the terminological issues and needs found during testing.

Description

Terminology is an important aspect of an interface as it helps the user understand its content. However, during testing some terminological issues were found which will be explained below.

Findings

Some interesting discoveries were found when the participants were working on the task listed below:

Task 6: Create a demonstration.

During the task the participant were supposed to create a demonstration. They had previously created a static worklist and thought that they would create the demonstration via the same option - create new worklist. Instead, they must click on "create a new demonstration" which is listed below as the picture below will show in Figure 5.2:

![Figure 5.2: Drop-down menu for creating worklist & demonstrations.]

The data revealed that the reason behind the mistake of clicking create new worklist rather than new demonstration is that there’s a terminological issue with the differentiation between demonstrations and worklists. The reason for this is that all kinds
of lists (dynamic worklist, static worklist and demonstrations) are found under the
tree node system/role/user worklists. Thus, the user believe that demonstrations
are also treated as worklists rather than demonstrations. The consequence of this is
that they want to create a new demonstration by either clicking "worklist" or 'static
worklist' which in this case is wrong.

Solution: This is a difficult issue as its impact is unclear and solving it might disrupt other work flows. However, being able to search for "how do I create a demonstration?" would solve this and many other problems.

Task 10: Change the date and time for the demonstration.

During this task the participants had issues finding the properties menu of the
demonstration. While they understood that they had to right-click the demonstra-
tion in order to change its date and time, some test participants thought that they
were in the wrong place when they couldn’t find the word 'option' or such. As such,
while the word properties might be common for native English speakers it seems to be more difficult for someone from Sweden. However, as the user group I met will be using the system in Swedish this specific problem will be a non-factor. Even so, it shows that someone that doesn’t have English as their mother tongue might have issues if they’re for some reason forced to use an English version of Sectra PACS (i.e. Swedish radiographer working in England).

Solution: Consider changing the word 'properties' to 'options' or 'edit'.

Task 12: State that both your patients have been demoed on the demonstration.

In Sectra PACS there’s a function that quickly lets the user tell the system that a
patient has been demoed on during a demonstration. However, for someone new
to the system this function might be difficult to find. Thus, the users wanted to
right-click the patient in order to change its status.

Solution: This issue will most likely disappear after the user has realized that they may confirm that a patient has been demoed on a demonstration. However, it’s once again worth to mention that a search for feature-function would be helpful during the initial learning curve for a novice user as the system is very large. A more simple fix might however be to add 'demoed' to the status list.

Additional finding: Another issue that was found by myself and confirmed to be a terminological issue by the UX team is that if a user would want to search for help, they would find the menu that is shown in figure 5.4:

When clicking the second help, the user will automatically launch a 400 page PDF-file which could be a crucial error for someone working on a slow computer. Further, if someone is actually searching for the manual they might not find it as there isn’t anything labeled "manual" within Sectra PACS which is shown in figure 5.5.
5. Results

**Figure 5.3:** Sectra PACS’ option to state that a patient has been demoed.

**Figure 5.4:** The current state of the help-dropdown.
5. Results

Solution: The 'help' within the drop-down menu should be renamed to 'manual'.

![Figure 5.5: 'Help' changed to 'Manual'](image)

5.3.2 Information overload

This section will describe the information overload issues and needs found during testing.

Description

As Sectra PACS is a large system with many different user groups, the user might sometimes experience that there’s too much information being presented. Thus, the information overload problems and needs that I’ve found during testing will be presented in the next section.

Findings

The first task will be presented below:

**Task 13:** Search for the patient <name shown on screen>.

This task was difficult due to several reasons. First, the users struggled to find the search function even though it’s always shown in the test environment that they used. One reason they couldn’t find it was because the search icon is designed as
binoculars which wasn’t the archetype of what the users thought the icon would look like. Instead, they searched for a magnifying glass or some sort of input field. Further, the icons’ colors are in general similar to each other which has resulted in a lack of contrast between them. However, when they couldn’t find what they were looking for they instead went looking for a search function under 'tools' as they considered search to be a tool that they’d use to find information. Instead, the search tool is placed under 'edit'.

![Figure 5.6: Current location of the search tool.](image)

During a focus group session I learned that the reason why 'search' is placed under 'edit' is because Sectra PACS is built on a Microsoft platform and that they try to keep similar to their Office Package. In Microsoft Word for instance, the search tool is under edit. The reason for this is that when the user uses the search tool in Microsoft Word, they often want to search and replace a word. In this case putting the search tool under 'edit' makes sense as the user wants to edit his or her file. In Sectra PACS however this doesn’t make sense as the user doesn’t want to search something and edit it right away.

**Solution 1:** Change the icon to a magnifying glass and consider changing the icons colors to create better contrast between them. Also consider adding an input field for quicker access to search.

**Solution 2:** Move the search tool to the 'Tool' drop-down rather than having it under 'Edit'.

The second found problem is that the search tool itself is somewhat overwhelming due to its many options which can be seen below:
Figure 5.7: Search tool in Sectra PACS.

The users found that the search function had too many options and that it was difficult to scan all available attributes that the system wanted as input.

**Solution:** Rather than forcing the user to use a lot of search values before actually searching, they’d prefer a quick search and filtering afterwards instead.

**General findings:**

All users emphasized the fact that there’s always a lot to chose from, so there’s sometimes difficult to find what you’re looking for. Further, some users stated that they’d like a visually limited system that only consisted of functions that they use as seen in this quote:

### 5.3.3 Missing features

This section will describe missing features that they user group think that they’d like to have in their PACS.

**Description**

The missing features presented in this section are functions that the user group would like to have while completing their daily work tasks and were found during the test session.
5. Results

Findings

The first missing feature was found during the second and third task:

**Task 2:** Find a worklist with more than 5 patients.
**Task 3:** Find an empty worklist.

During these two tasks, the patients had to click between the lists in order to find a worklist that would fulfill the criteria presented in each task. The solution to the task is simply clicking in worklists until they’ve found a list matching the task. However, even though the task itself is easy to complete the users didn’t enjoy blindly clicking around. The reason for this seems to be that they’re scared of interacting with the system as they think they might make some error that’s difficult to fix. As a result of this they looked for visual indications to see how many patients were in each list rather than exploring them.

**Solution 1:** Add visual indications to easily see how many patients a list has. This could look something like this:

![Figure 5.8: Possible solution for showing how many patients a list contains.](image)

In a focus group meeting that was held afterwards it came to my knowledge that this feature does exist if the user works with favourite lists: The feature isn’t set to default however as it eats up performance which may have crucial consequences on a users workflow.

**Task 4:** Add two patients to the worklist in two different ways (one way per patient).
During this task it was found that dragging a patient onto a worklist worked rather well. An issue that most patients however had was that there weren’t sure whether or not they actually had succeeded dropping the patient into the list. In contrast to drag and drop, the users had trouble with adding a patient to a worklist by right clicking the patient and adding it to a specific worklist. Further, one user said that she wanted to add a patient to a worklist via the worklist rather than via the patient. To clarify; she wanted to right click on a worklist and get a list of patients that she could add to the worklist too. Finally, there seems to be a general need for a search function as stated by the quote below:

*I want to be able to search for words and get help sections, that’s really important to me.*

**Solution 1:** Investigate whether or not adding patient to worklist via worklist is wanted by the community.

**Solution 2:** A reworked search tool.

**Task 15:** Merge the patient with the other patient in the list.

During this task, the user was supposed to select two patients, right-click and chose merge patients. After they had struggled for a while they were encouraged to try to select both patients at once. At first, the users tried to right click one patient to find a merge function. When that didn’t work they tried to select two patients by drag-selecting which isn’t supported in the PACS. This task is one of the reason why a user expressed the following:
5. Results

“The most difficult thing is not knowing what I can do with patients.”

Finally, all users completed the task by control-clicking on the second patient in order to complete the task.

**Solution:** Allow drag and select to select several patients at once. This would also make it possible to drag several patients into a worklist at the same time too, making the interaction faster.

5.3.4 Confusion, needs explanation

This section will describe situation that caused the test participant confusion.

**Description**

When getting familiar with a new system there’s often confusion which may disturb the user’s learning process. The findings in this section will account some general issues that were found during testing.

**Findings**

The task below will exemplify some confusing moments that were found when the user’s completed their tasks:

**Task 5:** Make the list one of your favourites.

During task 5, some users were struggling with using right click, but once they tried there were no issues with creating a new favourite group.

**General findings**

- In general, most users were unsure when to use left- or right-click in the beginning. After a while they were however comfortable with the interaction and did well.
- The list-icons were difficult for the users to differentiate. It’s however important to emphasize that we did use a fairly small screen (13”) but during one of the focus group meetings it was also said that this problem has been seen before.
- Lists within favourite groups are limited. For instance, one user tried to edit a demonstration within a favourite group. That’s however not possible, and instead the user only gets the option to either go to the specific worklist, remove it from favourites or simply go to all favourite groups.
5.3.5 Flow problems

This section will describe the flow problems that the users experienced during testing.

Description

A flow problem is an issue that disrupts the user’s workflow in some way. These issues will be exemplified in the next section.

Findings

The relevant tasks will be presented below:

Task 9: Add the two patients from your worklist to the demonstration in two different ways (one way per patient).

This task is a variant of task 4 (add two patients to the worklist in two different ways (one way per patient). As mentioned earlier there’s a terminological issue with demonstration which in this task lead the users to believe that they could right-click a patient and use the "add to worklist" feature to add him/her to a demonstration. However, in order to solve the task they have to right click the patient, hover over "demonstration" and then find the option "add to demonstration". The quote below will exemplify what happened:

“What was difficult was when I encountered the problem where I couldn’t copy my patient into my worklist which I just had done, I got completely baffled.”

Further, the same user addressed the fact that she didn’t feel that she got enough feedback from the system:

“Exactly, there’s a lack of feedback. Why did it work before but not now?”

She wants to understand what she did wrong, but there’s no way for her to get this information. In addition to this, another participant said:

“I don’t think it’s difficult to find the features, but how to actually use them.”

Solution: The solution to the problem could be to change the default menu as shown below:
5. Results

Figure 5.10: Default look of the patient pop-up menu.

Figure 5.11: Default menu but with "Add to demonstration".

However, during a focus group session it was said that Sectra is avoiding having too long drop-down lists which is good to have in mind if considering this change. Low feedback is however a problem that appears in different areas of the system.
5.3.6 Function placement

This section will describe the function placement that the users experienced during testing.

Description

The issues and needs that are presented in this section will focus on features that are difficult for the users to find.

Findings

The first task is presented below:

**Task 1**: Locate the area in which it is possible to create worklists and create a static worklist.

When the users are presented with the initial view of PACS, they first didn’t know how to activate the initial window where are the worklists are. They are presented with the view shown in Figure 5.12:

![Figure 5.12: Default starting view of Sectra PACS. Orange box shows where the worklist tree will be expanded.](image)

Rather than clicking on the marked area on the picture, some users found the option to show worklist via the view menu bar as well which is another way to activate it. However, the users managed to activate the field fairly quickly. When the users then were supposed to find the place where they could create a new worklist however, they got completely lost as they didn’t know that they could right-click "system worklist" for instance.
Three possible solutions are listed below:

- **Solution 1:** Make the system trees grow on hover and also show a plus icon to create new lists.
- **Solution 2:** Make the manual searchable. By having a searchable manual, users could type in "how do I create a new worklist", and the info would show. At the moment, it’s possible to search the PACS manual as in any PDF, but there are no keywords which makes it difficult to find what you’re looking for unless you phrase it correctly. By adding keywords to the existing manual, users could type in something in the lines of 'keyword create worklist' and hopefully they’d find what they’re looking for.
- **Solution 3:** As previously mentioned, a search for function-feature would be a great way to solve this problem too.

**Task 8:** Locate where your favourites are saved.

In this task, the test participant had recently added their static worklist and their demonstration to a favourite group. This task aimed to see whether or not they could find the area in which their favourites were saved. It was however found that the users struggled a lot to find the favourites, and the reason for this seems to be an issue with the visual hierarchy.

**Solution:** My solution to this problem was to re-use the star that’s located next to the 'Make list favourite’ shown in figure 5.13. In addition to this I also added some shadow to the marked areas below to make them "pop" in contrast to their surrounding elements shown in Figure 5.14.

**Task 22:** Filter the list after a body part.

The participants all had major issues with this task as they simply couldn’t find the input area which is shown under the text in Figure 5.15

**Solution 1:** Once again, being able to search for feature could help the user find the input field as shown in figure 5.16.

**Solution 2:** Some minor visual changes could also make it easier to see the input field.

**Solution 3:** In addition to the visual changes above, some interaction could also be added in order to make the filter field easier to locate.

**Task 14:** Change the status of the patient to "prioritized".

The user found the function with ease, but when she saw the list she got unconfident. The list is showed in figure 5.17:
5. Results

Figure 5.13: The "add to favourites"-star.

Figure 5.14: Shadows added to the label texts.
5. Results

Figure 5.15: Default view of the filter input field.

Figure 5.16: Possible "search for feature"-view.
During a focus group meeting I however realized that the list isn’t unsorted, but a chronological path of a patient’s status from new to final. However, in order to make the list look sorted, numbers could be added to the list as well. This would make the list easier to scan too, as the users could for instance learn that step 19 equals 'transcribed' and so forth.

5.3.7 System error management

This section will describe the function placement that the users experienced during testing.

Description

This section will describe how Sectra PACS manages errors from the user.

Findings

One task caused a serious error for a user during testing:

Task 19: Add "priority" to columns".

To complete this task, the test participant was supposed to right-click on a column and press 'add a column'. However, as previously mentioned the filter input box was hard for the users to locate. As the users completed this task before the filter for body part task (which was the last task, 22), they didn’t know that the filter
field was actually there (they hadn’t done the task previously at Sectra). Thus, one user managed to miss right-clicking on the column box and instead accidentally hit right-click in the filter field which results in this menu showing up shown in figure 5.18:

![Figure 5.18](image)

**Figure 5.18:** Picture of the options showed when right-clicking in the filter field

Further, another user managed to get the same error when she by mistake double left-clicked on a worklist’s name which allows the user to change its name, and then right-click in the input field shown in figure 5.19:

5.4 Learning

This section will describe the findings related to how the users thought about their training session at Sectra as well as how they want to learn in general.

Description

Learning in general is something that differs from person to person. Thus, understanding how different users learn is an important aspect as it helps us understand how a system may aid the user during his or hers learning process. The findings presented in the next session were all found during the interview sessions.

Findings

During the interview it was found that the training sessions at Sectra had been good in general but with a few rooms for improvement. These findings will be presented
5. Results

**Figure 5.19:** Picture the same error as in figure 4.18 but this time when changing a worklist’s name

below with a possible solution afterwards:

**The order of the sessions:** The users were unsure of whether the order for the training sessions were the best. During the first week, the users got to learn how to setup their PACS environments from a back-end perspective. Further, in the second week they got to learn how to use the workflow of the PACS. All users said that what they had learned during the first week fell into place during the second week when they could interact with the things that they had setup. Further, one user said that the first week was a complete jungle as they didn’t know exactly what they were working with. However, when questioned about whether they would like to switch the weeks around no user explicitly said that they thought that it would be better. Instead, they thought that there was a reason for why the sessions were in this particular order. However, when I asked an application specialist who were responsible for the training sessions if there were any thoughts behind the order she stated that there wasn’t.

**Solution:** Investigate whether it might be better to swap the order of the training sessions.

**The users were instructed to work around in the system without any given goal or task:** As a novice it’s really difficult to know what you’re able to do in a completely new system. The test participant expressed frustration when they were talking about how they were instructed to play around in the system and train without any given goal. Instead, they would like to always have a set of tasks that they can do rather than clicking around in the system without a purpose.
5. Results

**Solution:** Create enough tasks so that the user never have to click around in the system without any given goal.

Some of the information during the training sessions at Sectra was inaccurate: The test participants mentioned that it was very distracting to work with some of the tasks during the training sessions at Sectra is the medical information was inaccurate or false which disrupted their flow (i.e. something in the lines of a task stating that there is a lunge on the picture but it’s actually a breast).

**Solution:** Let a team of individuals with great medical knowledge go through all the tasks and additional Sectra material in order to keep all information valid and correct.

**Learning needs:** When questioned about how they would actually like to learn a system in general, the following was found:

1. The users want to be independent when learning.
2. They want to learn little by little.
3. They want a comprehensive manual.
4. They want tutorials and tasks to work with.
5. They want a resettable test environment.

**Solutions:**

The main issue that the users said they struggled with is the fact that their time is very limited. Further, when some of them actually do have time they still might be unable to learn as they get stuck without any nearby help. Therefore, the users would like to have short task-based tutorials that doesn’t take a lot of time to complete in an environment that is resettable. The issue right now is that in order to complete a set of tasks, the users must create and manage patients as the patients are unusable after they for instance have been checked out from the system. Thus, a big part of the upcoming test environment that the customers will have in their clinic soon is to create patients. In other words, the users will not be able to simply work on tasks that they feel insecure doing as they will also have to create patients all the time. A solution to this problem would be to simply create a resettable test environment which just brings back all the patients to their original state with a simple click, which is something that doesn’t exist today. Further, a user mentioned that she usually learns a new system by interacting with it while also having a manual in her knees. Even though there is a manual available today, it might be difficult to find as it’s called 'help' which was mentioned in a previous section. Further, tutorials are something that the users would gladly use too. This is however something that Sectra has available on their platforms.
5.5 Questionnaire

The system usability scale (SUS) questionnaire scored a mean of 59.5 (SD = 13.85) which translates into a D-grade, the second lowest (the median SUS score is 68). Further, one participant rated the system highly in comparison to the others which is shown in figure 5.20 and therefore increased the standard deviation.

![System Usability Scale](image)

**Figure 5.20:** Scatter plot of the SUS results.

Learnability was calculated by only using item 4 and 10 in the questionnaire:

4: I think that I would need the support of a technical person to be able to use this system.
10: I needed to learn a lot of things before I could get going with this system.

The learnability reached a mean of 60 (SD = 20.0) which shows that the participants rated the questionnaire with great variation. Further, as the general SUS score was 59.5 there was almost no difference between the level of general usability and learnability. As learnability hasn’t been used as a factor before 2009 there still isn’t a validated learnability database. The reason for this is that even though it’s based on filtering the original SUS questionnaire’s data, there aren’t any grading segments yet for learnability within SUS. Instead, what could’ve been interesting is if for instance the learnability was significantly higher than the general perceived usability or the other way around. This wasn’t however the case in my study.

5.6 Impact Mapping

This section will showcase the use of the impact map that was made as a final product of this study. The data which the impact map is based on is collected both the QS protocol and the semi-structured interview and analyzed during the focus group meetings.
5. Results

5.6.1 Goal
The previous SUS score placed the PACS in the D-grade segment. In order to improve the PACS general usability (and as a result of this also its learnability), I chose reaching C-grade as a goal for my impact map which would require the PACS to reach the median score of SUS, 68.

![SUS C-grade](image)

**Figure 5.21:** Scatter plot of the SUS results.

5.6.2 Actors
After defining the goal, I identified the primary- and secondary actors as well as the off-stage actors:

1. **Primary actors (whose goals are fulfilled):** The primary actors in my study are the test administrators that I met during my study.
2. **Secondary actors (who provide services):** The secondary actors are the Sectra training personnel as they are the ones supporting the users during their initial learning curve by providing training sessions, manuals, tutorials etc.
3. **Off-stage actors (who have an interest in the behaviours but are not directly benefiting or providing service):** The offstage actors are the Sectra developers who design the PACS in order for it to be as intuitive and easy to use as possible.

Thus, our initial node is expanded by the actors as seen in figure 5.22

![Primary actors](image)

**Figure 5.22:** Primary actors.

5.6.3 Impacts
After defining whose behaviour we want to impact, we need to answer how our actors’ behaviour should change. The impacts are listed under each actors:
5. Results

Primary actors (whose goals are fulfilled)

The impacts of the primary actors are shown in figure 5.23

![Figure 5.23: Primary actors’ impacts](image)

**Motivation:** In order to know how to improve the users' needs, it’s important to have a good way to report issues. Right now, if the user wants to report an error in the PACS he or she has to recreate the error. However, if the problem is due to the user’s lack of knowledge it’s something that can’t be recreated. Further, if the users are requesting help in a good and meaningful way, it’s a lot easier to help them. Additionally, it’s also important that the users are able to suggest improvements in general in order to make their experience with the system as pleasant as possible. Thus, the point of the chosen impact is to help the users help us help them. To exemplify; if we give the users good ways to report errors it’ll be easier for them to report errors, and if we give them a quick way to suggest improvements we’ll easier understand their needs and so forth.

Secondary actors’ (who provide services) impacts

The impacts of the secondary actors are shown in figure 5.24

![Figure 5.24: Secondary actors’ impacts](image)

**Motivation:** As mentioned in the learning section, the impacts chosen aims to find changes that will improve Sectra’s training personnel to improve their ability to assist users during their learning curve.
5. Results

Off-stage actors’ (who have interest in behaviours but aren’t directly benifitting or providing a service) impacts

The impacts of the off-stage actors are shown in figure 5.23

![Off-stage actors' impacts diagram](image)

**Figure 5.25:** Off-stage actors’ impacts

**Motivation:** The impacts mentioned above are segments that if solved will help us reach our goal. Thus, these are most of the issues found during the QS protocol and are limited to functionality within the PACS in contrast to the ones mentioned for the secondary actors.

5.6.4 Deliverables

This section will introduce the deliverables that were found during this study. The deliverables aim to answer what the organisation can do to support the required impacts. Further, while all deliverables should help the organisation reach its goals, not all deliverables have or should be implemented. Instead, this should be seen as an iterative process. All of the deliverables have either be confirmed by the users or Sectra employees to be of interest. Each actors deliverables will first be presented and then motivated individually below.

**Primary actors’ deliverables**

The suggested deliverables are listed in figure 5.26:
5. Results

Figure 5.26: Primary actors’ deliverables.

**Report issues:** In order to report issues, a ticket tool-feature could be implemented into the PACS which would make it easier to both report issues and file for help.

**Report assistance:** Requesting assistance could be made easier by for instance allowing the user to chat with support directly via a chat in the PACS. Further, being able to ping a colleague directly through the PACS or see what they’re currently up to could also be an interesting feature in the future.

**Suggest improvements:** An interesting function that was discussed during the focus group meetings was a feedback tool that could be implemented into the clinics test environment. In this test environment, the users would complete a task and afterwards rate how difficult the task was, if something could be clarified or suggest improvements. This would allow an automated way for Sectra to collect quantitative data from novice users which could be extremely valuable if treated and analyzed right. One way to do this would for instance be to implement the SUS questionnaire into the training session at Sectra and let the users answer it in order to assess both general usability and initial learnability in larger quantities with little effort.

**Secondary actors’ deliverables**

The suggested deliverables are listed in figure 5.27:

Figure 5.27: Secondary actors’ deliverables.
5. Results

**Improve test environments:** Real-life tasks with accurate and valid information should be implemented into a resettable test-environment that comes with already made patients that are made specifically for each task.

**Create tools for improved learnability:** Improve the manual by making it searchable for keywords. An easy way to do this could be to simply print "keywords: create, worklist" in the PDF which would make it possible to find the correct segments by simply searching the document for "keyword create worklist". Further, more task-based tutorials should be made in order to keep the participants busy during the training sessions at Sectra.

**Improve training sessions:** The training sessions may be improved by creating more tasks with accurate and valid information. Further, it might be better to swap the training sessions and instead start with the workflow and then show how everything is setup. This deliverable however only suggest that this should be investigated before any decisions are made.

**Off-stage actors’ deliverables**

The suggested deliverables are listed in figure 5.28:

![Figure 5.28: Off-stage actors’ deliverables.](image)

**Solve information overload:** First, the search tool should be reworked in order to avoid presenting the user with too much information. Secondly, it might be good to completely rework the PACS skin in order to allow for a better visual hierarchy which would make the whole interface easier to scan.

**Solve terminological issues:** It’s always important to keep lists easy to scan by having them sorted. Further, the worklist versus demonstration-issue that was presented earlier could be solved by simply presenting more options in order to avoid disambiguation.

**Consider adding missing features:** It’s always good to consider adding missing features such as weighted lists even outside favourite lists or being able to add
patients to a worklist via the worklist. However, which features that should be considered first is above my knowledge as I don’t know exact what limitations are present or what customers want the most.

**Improve error management**: Right-click in text input fields should be disabled.

**Increase user’s trust**: A feedback field could be an interesting way to solve the lack of feedback that the users experienced during my test. For instance, when adding a patient to a list a small box could appear stating for instance that "patient a has been added to worklist b", with the option to undo by either clicking the box or using the ctrl + z command.

### 5.7 Conclusion

This chapter has presented the results which may be used to improve the initial learnability both within and outside Sectra PACS. The findings correlated well with what was expected after the literature study. The research question was answered well as initial learnability was first assessed using the QS protocol, semi-structured interview and all positive SUS. Afterwards, possible improvements were suggested after the data had been analysed mainly by using focus groups, clustering and impact mapping. Further, the QS protocol found issues within the PACS by focusing on the interaction between the user and the PACS while completing tasks that were relevant to their work. In contrast, the semi-structured interview found issues and needs outside of the PACS by mainly focusing on how the test participants learn in general as well as their thoughts about the training sessions at Sectra which was presented in the persona. In addition to this, the SUS aimed to assess general usability and initial learnability at a general level to give a quick (and dirty) overview of the system. While it’s difficult to say how good Sectra’s learnability is in comparison to competitors for instance, the results of my study have showed where there’s room for improvement. My study could therefore be used as a baseline which future improvements could be measured against. Further, the results have also been exported into an impact map that produced deliverables which may be implemented into agile methods. This impact map could also act as a baseline if Sectra decides to do a second iteration of my work. If this would be the case, my study has also provided a process chain of methodology which Sectra should be able to replicate with ease. Thus, the results of this study showed that the used methods in this study assessed initial learnability by finding needs and issues both within the system and the organisation that provides it while also being able to find possible solutions to them. Further, the results as a whole shows that the chosen metrics worked well. As mentioned in the methods chapter, they aren’t limited to a specific context and should therefore be applicable to any given complex system. Thus, the result chapter also shows that the process chain of methods contributes to academia as it gives both researchers and practitioners a way to assess, evaluate and improve initial learnability in complex systems. Finally, it was confirmed that Sectra were satisfied with the results and found them to be relevant as well as something they’d like to investigate further.
5. Results
6

Discussion

This chapter will discuss the results and the methods as well as ethical aspects for this study.

6.1 Results

The results for this study showed that there’s room for improvement both within Sectra PACS and during its training sessions. The main issue seems to be that the system is very large due to its many different user groups which makes it difficult to quickly grasp the whole concept of the PACS. Further, the issue during the training sessions seems to be that there’s mainly a lack of tasks for the participants to complete, that the information is sometimes medically inaccurate and that the training weeks potentially should be swapped around. However, during my literature study I didn’t find any work that examined initial learnability in complex system which makes it difficult to evaluate how good Sectra PACS’ initial learnability actually is. Instead, this study’s main result is ways to find areas for improvements rather than evaluating how good or bad the system’s initial learnability is perceived in contrast to others. Further, my used definition of initial learnability was the user’s ability to eventually achieve specific performance which in this case translates into "what issues and needs exists for a test administrator trying to learn how to complete her daily work tasks?". With this definition in mind, I’ve been able to show how it’s possible to combine a set of methods in order to produce results which will assess and propose improvements for initial learnability which answers my initial research question.

Something that could’ve been interesting however would be if the two factors of the SUS questionnaire had been significantly different. For instance, if the perceived learnability had been a lot lower than the perceived usability one could draw the conclusion that even though the system is difficult to learn, it’s pleasant to use. In contrast, if the learnability had been a lot higher than the perceived learnability the system would be easy to learn how to master but awkward to actually use. But, as the learnability and usability scored almost exactly the same, no such conclusion could be drawn.

After my study I’ve also been able to answer that question using previous literature and applying it in a new context with good results. Thus, while I wasn’t surprised that I got the results I did, I’m glad that the proposed methodologies worked well in a very specific and complex environment. Additionally, my study shows that the proposed methods for evaluating initial learnability seems to be applicable to other
6. Discussion

complex environments too. I do however believe that in order to be able to do these kinds of evaluations the system needs to be limited to smaller areas to get good and relevant data, especially since time is often limited. Further, I wouldn’t have been able to complete this study without support from Sectra as their insight has been a invaluable throughout the whole process. For instance, the results would have been difficult to interpret without their help as they already knew some solutions to the problems that I’d found. In addition to this, I wouldn’t have been able to create the tasks for the QS protocol without the help of the UX designer and application specialist. It was also confirmed that Sectra [1] were satisfied with the results which was very important in this project. A problem is however that it’s difficult to know which issues, needs and problems should be solved first. The reason for this is that there are many different stakeholders with different needs and desires. Further, even though the user sample was fairly low there was a lot of findings which could be discussed in different focus group meetings which generated interesting results. Finally, this study showed that the chosen methods produced findings that were satisfactory and answered the research question by showing how said methods could be used to produce valuable results.

6.2 Methods

The methods used in this study produces valid data with little overlap as the question-suggestion protocol found issues and needs within the PACS while the interview mainly described areas for improvements outside of it. This divided the collected data into two segments - the interaction between the user and the system as well as the interaction between the user and Sectra training personnel. Thus, the process chain of methods used managed to assess factors that were relevant for initial learnability improvement within and outside of the given product. Even though there was a small user sample in the initial testing they generated a lot of data and found issues both within and outside of the PACS which is good as the recruitment was really difficult for this study. Further, as suggested by Nielsen [26] it’s however most of the time better to have 5 users and do 4 tests between iterations rather than have 20 users and do 1 tests as you’ll probably find about 80% if the issues with those 5 users anyway.

As mentioned earlier, the test version of PACS used in this study was in English rather in Swedish. The reason for this is simply that Sectra is an international company and that anyone reading this thesis shall be able to understand it. With this being said, some of the tasks presented would have been a lot easier for the user if the system had been in Swedish (such as understanding the word properties). The QS protocol [22] proved to be very good for finding both initial learnability issues and also possible improvements while doing so. In addition to this the participants seemed to enjoy the fact that they also got to learn the system while evaluating it. The semi-structured interview proved to be very useful as well as it caught issues outside of the PACS too and made it possible to create a persona [54] with goals and needs. Further, as the persona was created by mapping out all the participants goals and needs with data mainly acquired from the semi-structured interview it’s seemingly a quick and effective way to give a good overview of the target user.
The SUS questionnaire [60, 57] proved to be an interesting tool both for evaluating usability and learnability, but also for being used as a goal in an impact map. I used the all positive version [61] which I prefer to use over the traditional one as there doesn’t seem to be any significant differences between them and therefore no reason to alternate the questions. The three additional questions that aimed to assess the users’ perceived experience with computers and familiarity with Sectra PACS and similar systems should’ve been designed with labels that the participants could relate to. To give an example, rather than just showing them a scale 1-5 it should’ve looked something like "1 - I can’t use a computer without help’ and "5 - I can code myself out of any computer problem. As a result of this mistake, it was difficult to interpret the data as each participant had their own subjective interpretation of the scale which made it impossible to draw any conclusion about subsequent learning.

Furthermore, impact mapping [97] was a good tool for sharing knowledge with the rest of the UX group as a way to get their feedback. I do however consider myself very lucky to have had the UX team supporting me throughout this study as I do believe that the impact map should be made in collaboration with others from the organisation in order to be as effective as it can be. Furthermore, the focus group meetings worked really well and it was a treat to see the UX team discuss issues and possible solutions with each other and I found the trigger pictures [90] used to be an effective tool to start discussions.

The study’s validity is good as it found what it was looking for - ways to improve initial learnability in a complex system. Further, if I’d do the same study on another test group consisting if the same user type I’m confident that I’d get similar results. The reason for this is that the test instructions mentioned in the methods chapter are sufficient to get the same results that I did in order to find the needs and issues in the initial testing. Even though this study was done in a very specific environment there’s no reason to believe that the chosen methods wouldn’t be applicable in another system. Further, I’m confident that if someone else replicated my study by they’d find very similar results even though I’m sure new findings will be made too as I’m only counting on having found approximately 80% of the issues. The reason for this is that the process chain as mentioned earlier only consists of context free methods which may be used in any given area without using a great lot of resources.

With this being said, one potential problem that could arise is the recruitment of a coach for the QS protocol. As I limited my study of the PACS to a small segment of it I was able to learn that specific environment well with the assistance of Sectra. However, if I were to learn a completely other segment within the PACS I’d have to rely on getting the support required to do so. Thus, it’s important to be able to have a coach that’s confident in actually supporting the test participant that’s completing the tasks as it’s a requirement to get good data. This is especially important as one of the advantages of using the coaching or QS think-aloud protocol is that it generates more satisfied test participants in comparison to the traditional TA protocol, which makes them suitable for a company meeting its users and/or stakeholders.

Finally I’m satisfied with the literature study that was done in the beginning of my work. All sources have been carefully picked, and most of them have at least 50-100 citations. Further, I put a lot of effort into actually confirming that there
was a gap within academia when it comes to initial learnability in complex systems which is why I believe this work’s purpose has been argued for. In addition to this, I’ve searched for possible new findings within each theory segment to make sure I’ve research that’s always up to date.

6.3 Ethical aspects

As Sectra PACS is a medical system with patient data it was very important to be careful with its data even though I used a test environment. As a result of this all patient data was blurred in the pictures presented in this study. Further, all user data has been treated anonymously and no patient data has been revealed in the pictures from the PACS. Finally, all testing that has been done has been under controlled circumstances and the test participants have been fully informed of the study’s purpose both verbally and in a consent form.
Conclusion

How can learnability methods be used to assess and improve initial learnability in complex systems?

This study showed how current learnability evaluation methodologies may be combined in order to assess and improve the initial learnability in a complex system such as Sectra PACS. The term learnability was studied and defined after a thorough literature study in which available metrics, evaluation methodologies and learnability guidelines were found and discussed. The most fitting metrics and methods were chosen after the literature study which generated data both within and outside the given system that made it possible to propose possible solutions that would improve the initial learnability in Sectra PACS. The data was collected by using a combination of the question-suggestion (QS) protocol, a semi-structured interview and an all positive system usability scale (SUS). The findings were analyzed by re-visiting recorded footage and conducting a focus group within the organisation which made it possible to create an impact map. By creating the impact map, findings were converted into deliverables which made be integrated into agile software development that many organisations such as Sectra use today. Thus, this study has showed how combining usability engineering and agile software development may be an effective and efficient way of assessing and improving initial learnability in complex systems. The approach of this study should be applicable to any complex system in order to assess and propose improvements to its initial learnability by using the process chain of methods presented. The reason for this is that the methods in the process chain aren’t context based or require a lot of resources to use. In order to replicate this study it’s however crucial to do a similar study in cooperation with a company that owns the given system that’s being evaluated. The reason for this is that the company’s insight is vital to form proper tasks for the QS protocol, recruit a coach with good experience of the given segment that’s being evaluated as well as getting good feedback and general insight during the data analysis. Before conducting this study initial learnability in complex systems has been a unexplored topic with little insight regarding the matter. Furthermore I hope that my work has showed that methods generally used in lesser complex systems may be applicable to complex environments too, and that improvements can be suggested using impact mapping with ease. Finally, this study has produced a process chain of methods that may be used both by researchers and practitioners to assess, evaluate and improve initial learnability in complex systems and has helped filling out an unexplored gap within academia.
7. Conclusion

7.1 Future Work

Finally, some future work will be presented in order to contribute with possible solutions that may improve Sectra PACS. Right now, there seems to be a lack of data collection. For instance, during the training sessions data could be collected by the responsible test administrators verbally, or by automation from the system itself by for instance integrating SUS into the system which users could fill out once in a while. Further, by making the training sessions more task-based it would be possible to collect data by letting the user fill out a survey after each task which would give direct feedback for that specific task. The survey could include questions such as:

- How difficult was the task?
- How may it be improved?
- Is there anything that you would like to change in the task itself?

Additionally, data could also be collected when the user has begun using the PACS as part of their work in order to assess his or her initial learnability with the system as a whole by using some of the metrics mentioned in the theory chapter such as time on task, task completion, number of commands used etc. Essentially, by collecting more data Sectra could actually get paid by their customers to collect valuable data with little effort. Additionally, if I had more time I’d like to study other possible evaluation methodologies. For instance, I was tempted to try a method named Instant Data Analysis (IDA) which requires at least two persons but produces good results in the end of the day. I did however decide against this method simply because I did this work alone. Thus, as my study aimed to find methodologies that could be used in order to assess and improve initial learnability, the natural next step would be to do the same for extended learnability. Furthermore, I’d find it very interesting to also study extended learnability within Sectra PACS by collecting quantitative data through data logging which would make it possible to study learnability over time and see the users progress for novice to expert. Finally, it’d also be interesting if someone would use my process chain of methods on another system in order to see how their case would be different from mine.
Bibliography


Bibliography


[65] John G Dawes. Do data characteristics change according to the number of scale points used? an experiment using 5 point, 7 point and 10 point scales. International journal of market research, 51(1), 2008.


[85] Jenny Kitzinger. The methodology of focus groups: the importance of interaction between research participants. *Sociology of health and illness*, 16(1):103–121, 1994.


Appendix 1
1. Jag är datakunnig.

2. Jag är bekant med Sectra PACS.

3. Jag är bekant med liknande program.

4. Jag tror att jag skulle kunna använda systemet utan teknisk support.

5. Jag tycker att systemet är enkelt.


7. Jag tror att jag skulle kunna använda det här systemet utan teknisk support.

8. Jag tycker de olika funktionerna fungerar väl tillsammans.


Stort tack till dig för att du ställer upp i min studie! Som ett resultat av din medverkan kommer jag kunna skriva min masteruppsats samt förbättra Sectra PACS. Studien genomförs som ett avslutande moment i det kognitionsvetenskapliga masterprogrammet vid Linköpings universitet. Innan sessionen börjar kommer jag muntligt att informera dig om vad informationen kommer att användas till samt hur studiens upplägg ser ut. Läs noga igenom de punkter som finns listade nedan och skriv under om du känner dig bekväm med vad som nämns. Om du är intresserad av att få ta del av studien efter att den är färdig får du gärna fylla i din e-post så skickar jag uppsatsen när den är klar!

- All data kommer att behandlas anonymt och kommer att användas till att utvärdera en prototyp till en applikation.

- Video- och ljudupptagning kommer att ske under studien för att underlätta senare dataanalys. Inspelningarna kommer att raderas efter projekets slutförande.

- Testet kommer endast att utvärdera hur väl Sectra PACS fungerar och alltså inte hur duktig du är som användare. De eventuella fel som uppstår är således pga. systemet och inte dig som testdeltagare.

---

Namn

Datum

E-post

---

Tack för din medverkan!