Lappeenranta University of Technology
LUT School of Business and Management
Degree Program in Computer Science

Hanna Salopaasi

USABILITY METRIC DESIGN FOR PRENATAL HEALTHCARE INFORMATION SYSTEM

Examiners: Professor D.Sc. (Tech.) Jari Porras
Associate Professor D.Sc. (Tech.) Kari Heikkinen

Supervisors: Professor D.Sc. (Tech.) Jari Porras
Associate Professor D.Sc. (Tech.) Kari Heikkinen

ABSTRACT

Lappeenranta University of Technology
LUT School of Business and Management
Degree Program in Computer Science

Hanna Salopaasi

Usability metric design for prenatal health care information system

Master's Thesis

2015

51 pages, 9 figures, 12 tables

Examiners: Professor D.Sc. (Tech.) Jari Porras
Associate Professor D.Sc. (Tech.) Kari Heikkinen

Keywords: prenatal healthcare, usability metrics, design

This work goes through the concept of usability in general and healthcare, especially prenatal healthcare, context. Different frameworks and guidelines used to measure it are considered. A collection of metrics is suggested to be used at a prenatal unit of one Finnish healthcare district. The metrics consist of a set of 12 general measures and a supplementary System Usability Scale questionnaire including a Fun Toolkit Smileyometer. The metrics are tested in real life work situations by observing meetings with patients and presenting the questionnaire for the focus group personnel. A total of 6 focus group patient meetings were observed. This work suggests that in order to get more conclusive data from the metrics the focus groups need to be more involved and observation situations need to be more controlled. Revised metrics consist of the 12 general measures.

TIIVISTELMÄ

Lappeenrannan teknillinen yliopisto
LUT Kauppatieteet ja tuotantotalous
Tietotekniikan koulutusohjelma

Hanna Salopaasi

Äitiyshuollon tietojärjestelmään soveltuvien käytettävyysmittareiden kehitys

Diplomityö

2015

51 sivua, 9 kuvaa, 12 taulukkoa

Työn tarkastaja: Professori Jari Porras
Tutkijatohtori Kari Heikkinen

Hakusanat: äitiyshuolto, käytettävyysmittarit, suunnittelu
Keywords: prenatal healthcare, usability metrics, design

Tämä työ käsittelee käytettävyyttä yleisesti ja terveydenhuollon, erityisesti äitiyshuollon, näkökannalta ja kertoo mitä viitekehyksiä ja suosituksia sen mittaamiseen on. Tässä työssä erään suomalaisen sairaanhoitopiirin äitiyshuollon yksikölle suositellaan käytettävyysmittaristoa. Mittaristo muodostuu 12 yleisestä mittarista ja täydentävästä Fun Toolkitin Smimeyometerin sisältävästä System Usability Scale-kyselystä. Mittaristoa testataan oikeissa työtilanteissa tarkkailemalla potilastapaamisia ja esittämällä määritelty kysely hoitohenkilökunnasta muodostuvalle kohderyhmälle. Kokonaisuudessaan 6 kohderyhmän potilastapaamista tarkkailtiin. Tämä työ ehdottaa, että kohderyhmien osallistumista pitää lisätä ja tarkkailutilanteita kontrolloida enemmän, että mittariston toimivuudesta saadaan tarkempaa näyttöä. Päivitetty versio mittaristosta sisältää ehdotetut 12 yleistä mittaria.

ACKNOWLEDGEMENTS

This Master’s Thesis is done at the Lappeenranta University of Technology. I would like to thank Antti Herala for the support he’s given during this process. I would also like to thank the customers of prenatal healthcare who let me observe their meetings with the personnel.

The process of writing a Master’s Thesis was a painfully long one as my topic changed multiple times along the last two years of my studies. Now at the final moments of this process I feel grateful that it is at the end. This topic was not without downsides but it gave me a great insight on usability and the things slowing ICT advancement in healthcare.

I would like to thank my examiners for giving me this topic and for helping me to this point.

And finally with love for Juuli, perhaps now mommy can spend some time with you and get a job to support your growing needs.

Table of Contents

..........i
1 INTRODUCTION..........4
1.1 Background..........5
1.2 Goals and limitations..........5
1.3 Structure..........6
2 General usability..........8
3 Measuring usability..........12
3.1 Methods of collecting usability data..........13
3.1.1 Think Aloud..........14
3.1.2 Questionnaires..........15
3.2 Choosing the right metrics for a specific case..........16
3.3 Usability data..........17
3.4 General metrics and frameworks..........18
3.4.1 System Usability Scale..........19
3.4.2 Fun Toolkit..........21
4 Usability in healthcare information systems..........22
4.1 Prenatal HCIS..........22
4.2 Usability aspect in HCIS..........25
4.2.1 General measures in healthcare context..........26
4.2.2 General frameworks in healthcare context..........28
4.2.3 Healthcare specific usability frameworks..........29
4.3 Summary of healthcare usability metrics..........31
5 Usability metrics for prenatal HCIS of one Finnish healthcare district..........33
5.1 Case analysis..........33
5.2 Proposed composition of metrics..........35
5.2.1 General metrics..........36
5.2.2 Usability questionnaire..........38
5.3 Testing set up..........38

6 Results..40
6.1 Metric testing..40
6.2 Revised metrics..40
7 Discussion..42
8 Conclusion..45
REFERENCES..46

LIST OF SYMBOLS AND ABBREVIATIONS

AHRQ	Agency for Healthcare Research and Quality
ASQ	After Scenario Questionnaire
CSUQ	Computer System Usability Questionnaire
D	Doctors
EHR	Electronic Health Record
EUCS	End-User Computing Satisfaction Questionnaire
HCIS	Healthcare Information Systems
HIMSS	Healthcare Information and Management Systems Society
ICT	Information and Communications Technology
IS	Information Systems
MW	Midwife
NIST	National Institute of Standards and Technology
PHN	Public Health Nurse
P&P	Pen and Paper
PUTQ	Perdue Usability Testing Questionnaire
RQ	Research Question
SUS	System Usability Scale

1 INTRODUCTION

Usability, in the context of information systems, tells about the ease-of-use of the interface of the system. It is a concept dependent on the system and the user it is applied to. As the ISO 9241-11 standard defines usability:

"*extent to which a system, product or service can be used by* ***specified users*** *to achieve* ***specified goals*** *with effectiveness, efficiency and satisfaction in a* ***specified context of use***"[1].

A classic example of usability in general are the English water taps with hot water tap and cold water tap on the opposite sides of the sink. They both serve perfectly the purpose of their respective taps but getting lukewarm water for washing your hands is very difficult. Therefore they suffer from poor usability in the context of washing ones hands. [2]

To define how usable something is metrics are needed. The "e*ffectiveness, efficiency and satisfaction*" mentioned in the standard are used as the base for different metrics[3]. Nielsen adds *learnability, memorability and errors* to this list[4]. Metrics measure for example time needed to perform some defined task or how confusing a user finds the navigation options of a system.

Usability is a big part of ICT(Information and Communications Technology). The growth of ICT has increased the need for usability and a way to measure it. As ICT is taking ground everywhere healthcare is not "spared" from it. Computers are part of the work of healthcare personnel for example in documenting the patient record. ICT is used to support existing processes but not all processes can be directly transformed to an electronic form without the system suffering from usability issues. Some claim that the healthcare systems seem to be designed for clinical transactions rather than clinical care and they require extensive training and lack of standard user interfaces[5].

In healthcare environment the matter of usability gets more critical as lives might depend on it. It is even suggested that the difficult of use of electronic nursing record systems may cause healthcare professionals to leave some patient information undocumented. Which then may cause errors in patient care and thus reduce patient safety. Better usability of information systems can reduce those errors and even increase the time spent actually treating the patient. [6], [7] Usability needs to be measured in order for it to be increased. This needs metrics.

1.1 Background

This Master's Thesis is done in co-operation with one Finnish healthcare district at Lappeenranta University of Technology. It's purpose is to define metrics for measuring the usability of the information systems (IS) at use at their prenatal healthcare unit.

At the start of this Thesis the prenatal unit was in a transitional phase with their IS. They were changing from a collection of old information systems to a completely new one in order to reduce the amount of concurrent systems and concurrent data. For this rose a need for a way to measure if any improvements were made. This Thesis presents metrics that can be used to measure the usability of prenatal healthcare IS.

Previous work has been made of the usability of a single program in the healthcare environment but no work has been made in pure prenatal setup and consisting of the combined usability of multiple different systems[8]. Related research on usability in Finnish healthcare environment have been questionnaire studies [8], [9].

1.2 Goals and limitations

The goal of this Thesis is to **design usability metrics** which provide data of perceived satisfaction from the end users and quantifiable data of e.g. task times and the number of different systems in use at a specific task for the prenatal healthcare IS that the healthcare district uses. An overview on how the usability data can be collected using the metrics is also included.

The main research question(RQ) is “How can usability be measured (and what affects it) in the context of healthcare noting the special needs involved?” It is divided and answered in the following sub questions:

1. What kind of guidelines and frameworks are used to measure usability?
2. What usability metrics have been used in healthcare domain?
3. What usability metrics can be used in this particular healthcare instance?
    3.1. What different subparts does the suggested usability metrics consist of?
    3.2. In what scope can the measurements be made?
4. What metrics were successful in this study?

There are few limiting factors:

- Metrics must be designed so that the data can be collected in real life patient situations.
- Finnish law dictates strict rules for patient privacy protection so the way in which the usability data is captured is important as the healthcare district wants to be extra careful. So no video nor audio recordings can be made.
- Every observation needs a written permission from the patient(s) involved.
- Time is limited because of the IS change schedule.

1.3 Structure

Section 2 presents more of the concept of usability. Section 3 continues to broader explain how usability can be measured and gives an introduction to some tools of the trade. Section 4 gives an overview of information systems at use in healthcare, how they differ from other ISs and how usability can be measured in the context they create. Section 5 presents metrics that could be used to measure the usability of the healthcare district's prenatal healthcare unit's IS. Section 6 presents the results of metrics testing and proposes revised metrics. Section 7 answers the research question 4 and goes through how the metrics were suited for the given task and suggest some future work needed. Section 8 concludes the whole process. Figure 1 presents the structure.

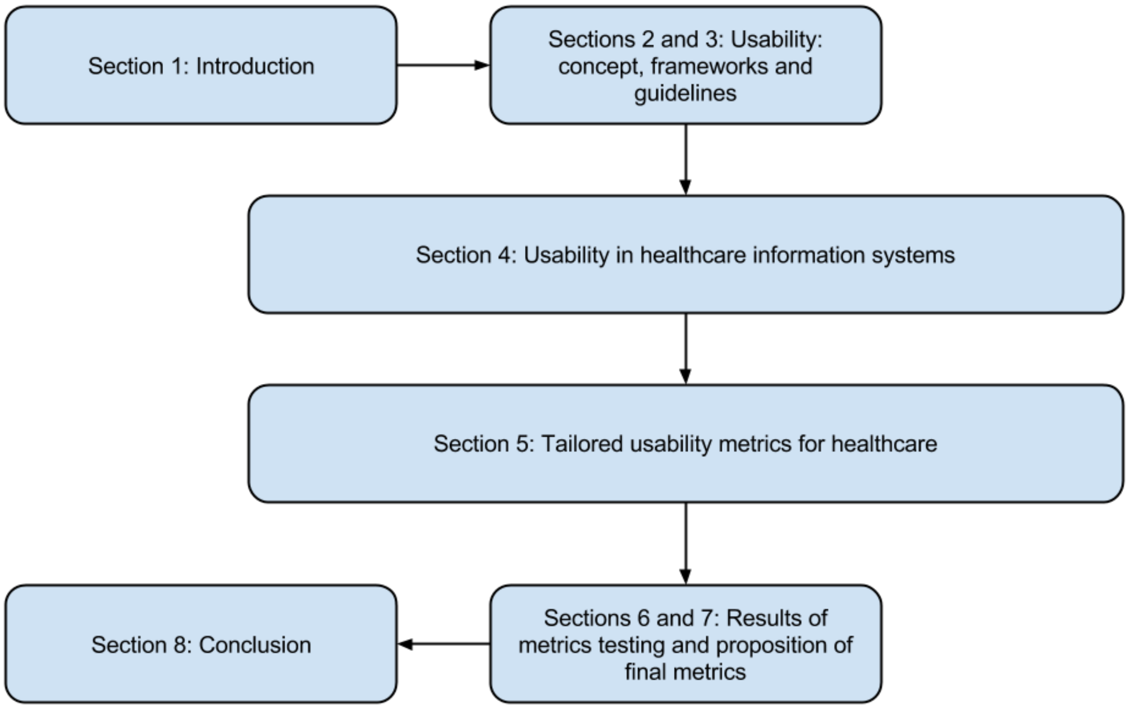

Figure 1. Structure of the work

2 General usability

In this section the concept of usability is defined. The meaning of usability needs to be known in order to design metrics to measure it. It has many definitions, one could argue that everyone has their own. As usability is such a wide concept it is important to pin down what part of it is relevant to this work. This work focuses on the ISO-standards definition and adds to it with Nielsen's components.

The ISO definition of usability gives us three variables by which one can explain and measure usability:

- **Effectiveness**: "*accuracy and completeness with which users achieve specified goals*".
- **Efficiency**: "*resources expended in relation to the accuracy and completeness with which users achieve goals*" and
- **Satisfaction**: "*freedom from discomfort and positive attitudes towards the use of the product*".

Figure 2 presents a view of usability based on the ISO. Usability is based on the users, tasks & goals and environment and the three variables, *effectiveness*, *efficiency and satisfaction*, are the "outcome" of usability. They are the parts of usability that are seen from the outside and can be used to define metrics and eventually measure the usability of a system. All of these are inside the wanted context of use.

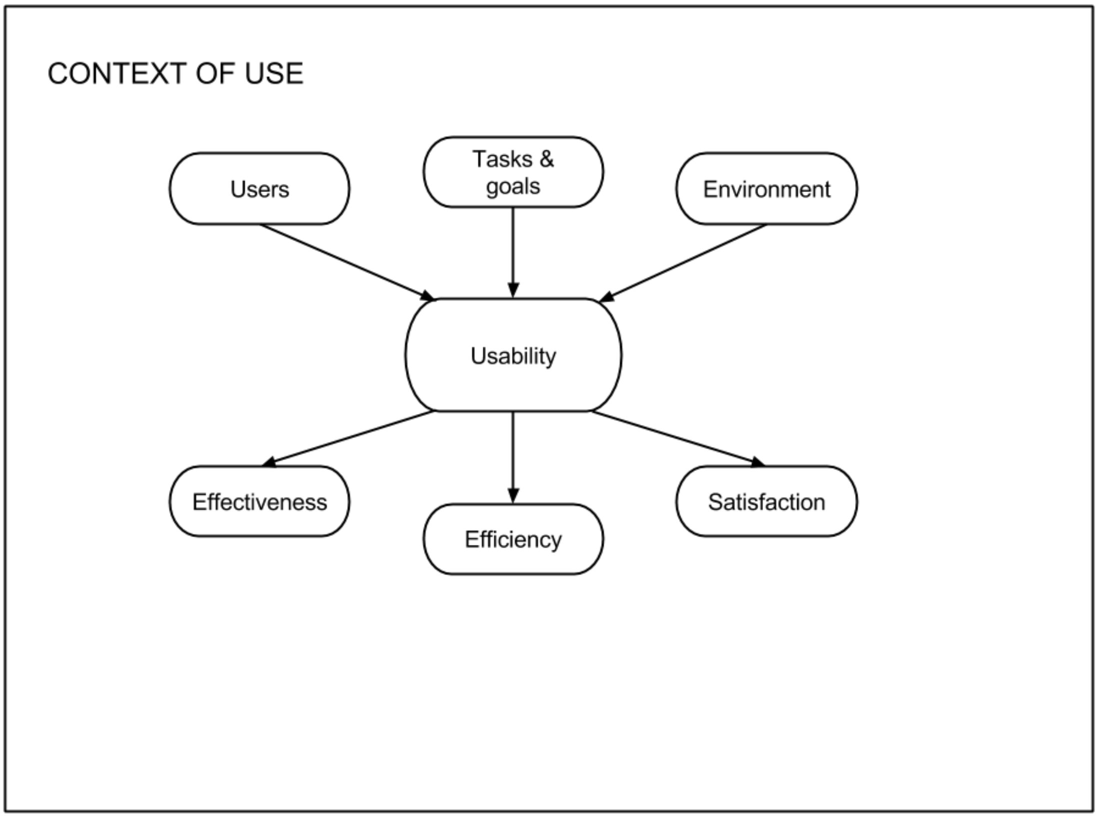

Figure 2. A view of usability based on the ISO-standard definition.

But these are not the only variables to usability. Nielsen defines more in his book Usability Engineering [4]:

- **Learnability: "***how easy is it for users to accomplish basic tasks the first time they encounter the design"*
- **Memorability: "***when users return to the design after a period of not using it, how easily can they reestablish proficiency"*
- **Errors: "***how many errors do users make, how severe are these errors, and how easily can they recover from the errors"*.

Figure 3 shows how these "Nielsen variables" add to the concept of usability. As it can be seen, the concept is quite a jungle even without considering the interrelation between the different variables; Errors, for example, can affect the perceived satisfaction.

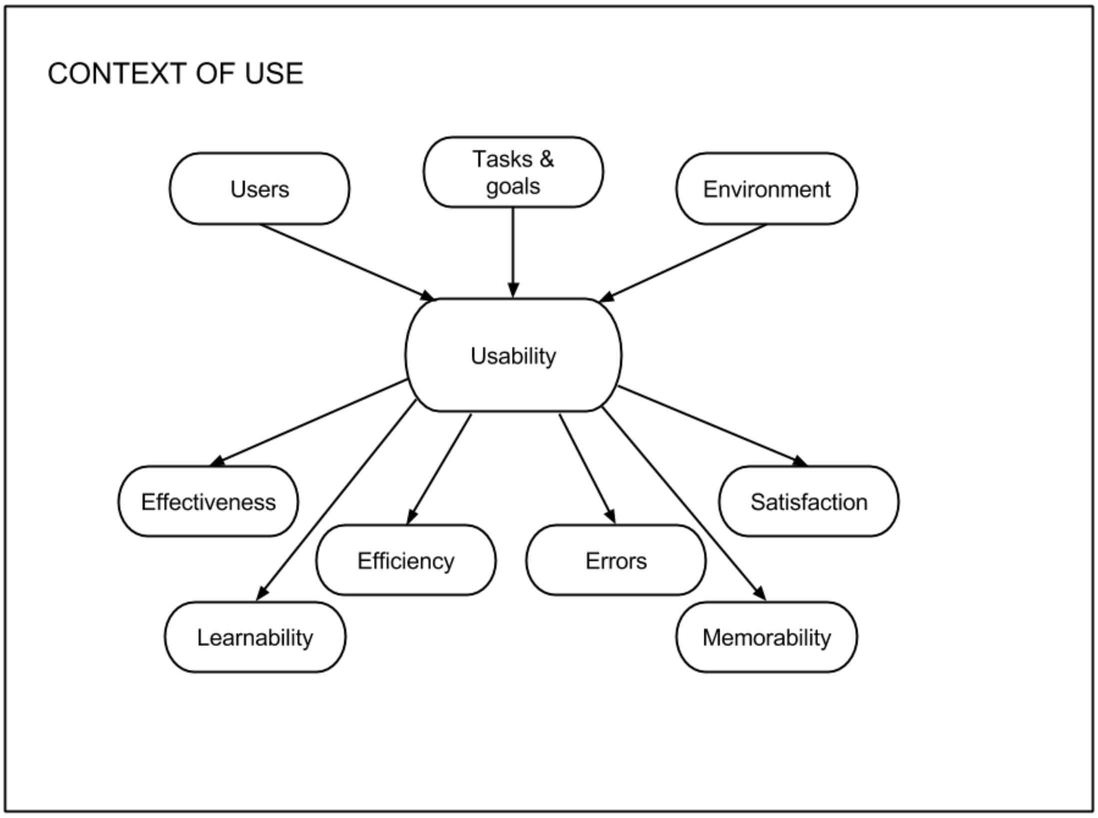

Figure 3. The view of usability completed with Nielsen variables.

These variables are more relevant when measuring usability but they give a picture of how interrelated the concept of usability is. It's complexity might be used to discard the importance of usability design and testing [3].

However, there are many reasons to take usability seriously. The ISO-standard 9241-210 rationalises the importance of usability by giving examples on how it improves quality[10]:

"Systems designed using human-centred methods improve quality, for example, by:

a) increasing the productivity of users and the operational efficiency of organizations;
b) being easier to understand and use, thus reducing training and support costs;
c) increasing usability for people with a wider range of capabilities and thus increasing accessibility;
d) improving user experience;
e) reducing discomfort and stress;
f) providing a competitive advantage, for example by improving brand image;
g) contributing towards sustainability objectives."

So usability gives benefits of cost reduction and more satisfied and more productive employees. Still usability is sometimes viewed as irrelevant and too subjective for it to matter.

3 Measuring usability

Now that the concept of usability is defined it raises a question of how can an immaterial attribute such as usability be measured. In this section is presented a short pretext to what are metrics. In 3.1 are listed methods in which metrics can be collected and 3.2 continues to explain what needs to be taken into account in choosing what metrics to use in a specific case. Section 3.3 gives examples of what kind of data one can expect from usability testing and what one can do with it. Finally section 3.4 explains plainly what usability metrics and frameworks are.

To know if a system is usable or not one needs a set of metrics like volume has litres or imperial units depending where the volume is measured. Metrics depend on what one wants to measure; task success, user satisfaction, errors are examples of usability metrics. [3]

The metrics need to be chosen based on what part of usability is wanted to measure. Figure 4 presents an example of a measure for each part of usability that was defined in the previous section. There are multiple measures for each part and used together they make a metrics system.

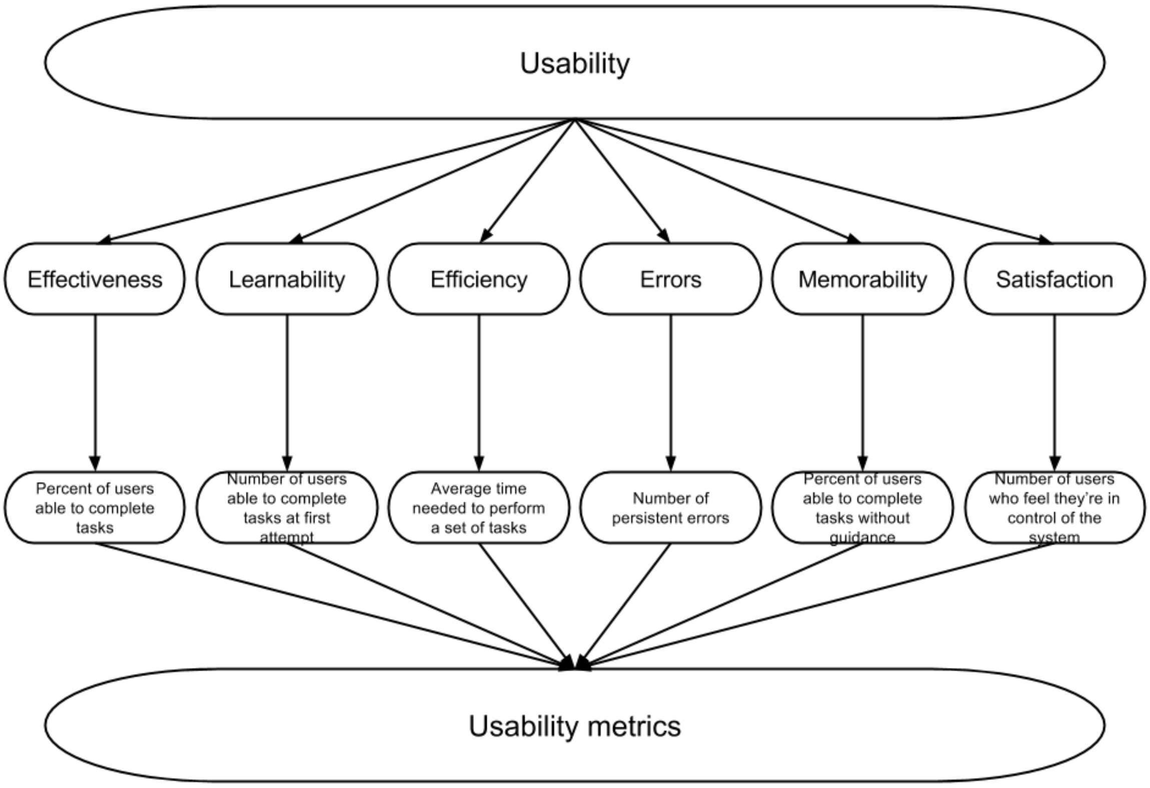

Figure 4. Example of a measure for each part of usability forming usability metrics.

3.1 Methods of collecting usability data

There are different methods in which the usability data can be collected using wanted metrics. But also the metrics that can be used depend on the methods in which the data can be collected. There is no point to use task time, for example, if the time used to do the given task cannot be measured reliably, especially if the measurer wants hard data instead of directional data.

The methods in which the usability of a system can be measured can be separated into three categories [11]:

1. Testing approach
2. Inspection approach
3. Inquiry methods

In the testing approach users are given typical tasks of their work to perform under inspection of evaluators, usually in a testing environment[11]. Example of such a situation could be that an intended user of the system is given a task to fill an information form(name, date of birth, etc.) of a person while being observed by a testing software, testing personnel and cameras in a testing lab. In the inspection approach the system is evaluated by usability experts[11]. Like the system containing the information form from previous example would be evaluated by the personnel who specialise in knowing the specifics of intuitive and easy to use interfaces. The inquiry methods include inspecting real work and presenting the users some questions about their likes and dislikes about the system [11]. In this approach the testing personnel from the example are situated in the workplace of the intended users observing the users doing the given task of filling the form as well as presenting them questions like “did you feel in control while you used the system”.

These three methods are the theoretical background for various tools which can be used as an aide in data collection. There are different logging software like Morae[12] which record data like keystrokes and time between certain clicks, the software might even track eye movement. These are more likely used in testing and inspection approach. Then there are tools to help inquiry methods like think aloud, described in 3.1.1, which can as well be used with testing approach.

3.1.1 Think Aloud

Think aloud is a user based testing method/tool. It was developed to gather information on the cognitive behaviour of humans performing tasks. In think aloud the user “thinks aloud”, that is to say narrates, what they do when they perform a task on a system. For example if the user is trying to find a submit-button they voice it somewhere along the lines “I’m looking for the submit-button. Where is it... Can it be at the top? Ah, there. I click the button.”[13] Think aloud allows the observer not only to know what the user does but also provides *why* the user does it [4]. Think aloud might gather a large amount of negative comments so it is good for finding out what features are dissatisfying for the user [14].

Think aloud can reveal usability flaws in a system and it is frequently used in usability testings with end users. It is commonly used during recorded usability sessions so all the data can be gathered without a loss. In these the user interacts with the system doing preset tasks and verbalising their thoughts.[13]

Due to the rich data think aloud provides a small sample of approximately 8 subjects suffices to gain a thorough understanding of task behaviour and to understand the usability problems of the system. If there are multiple types of end users all of them need to be represented in sufficient numbers in the think aloud test sessions.[13]

As the narrative is gathered on a full audio tape, and possibly on a video record, different coding schemes can be used to gain data from it. A code could be “guessing” and it could be assigned to “Can it be at the top?” from the previous voicing example. With this scheme one can find out how much the user needs to guess while performing the task, perhaps indicating about the intuitiveness of the system. As the think aloud method produces various types of responses due to different people thinking differently, a virtually unlimited number of different coding schemes can be developed to analyse the same dataset. [13]

3.1.2 Questionnaires

Another end user based testing method for the inquiry method approach is the use of questionnaires. Questionnaires can identify areas within a system that need improvement and can provide a measurement of the overall usability of the system [15]. They are best suited to generate feedback on general topics concerning the system [14]. Compared to think aloud of previous section (3.1.1) questionnaires may gather more comments on controllability, unobtrusiveness and privacy [14].

There are multiple prefabricated questionnaires suiting different needs. For example, IBMs After Scenario Questionnaire (ASQ) has only 3 questions while Purdue Usability Testing Questionnaire (PUTQ) has 100 [16], [17]. This is to say even considering the length only they are suited for quire different needs. ASQ is purely used to assess participant's satisfaction after completing a scenario [16]. PUTQ considers a total of eight different human factors that are relevant to software usability [17]. Section 4.2.2 shows more questionnaires available.

3.2 Choosing the right metrics for a specific case

The reasons listed in section 2 were clear about the importance of usability. In this light it is important to know what metrics to use as usability testing can go wrong if usability data is collected with “wrong” metrics. If the problem is long tasks times and only error rate is considered, the result will be that nothing is wrong with the systems usability. It is important to remember the context of use and the wanted goals. Even frameworks made for specific setting might not be 100% suitable for other cases than the one it was based on[10].

In order to know what metrics are suitable for the specific system in mind, a task analysis needs to be performed. The context of use, the user's goals, their information needs and their approach reveal important data on the problems of usability and how to get to the bottom of them. For example, if the users so called natural flow of tasks works differently to what the the system supports there might be a design fault. The metrics need to chosen in the way that they reveal information rich data about the tasks and point to the direction of better usability. [4]

Firstly the users need to be observed and surveyed while they perform their tasks on a prototype or a “real” system depending on if the development is still on going or if the usability testing is done in post-developmental stage. Then the task descriptions and observations are examined in the context of usability (presented in Figure 3, section 2). As in *what can be measured* quantifiably in the context of efficiency, effectiveness, satisfaction, errors, memorability and learnability.

After the examination in the process needs to be considered in which way the data can be collected and tools which can be used. Some measures are better suited for one collecting method than for other. It is more difficult to get objective data if one person is collecting task time data by hand versus if the data is gathered by a software designed for it.

After the metrics, measurement method and the tools used are chosen, a listing like Table 1 presents can be formed. More examples of typical quantifiable usability measurements can be seen in Nielsen's Usability Engineering [4].

Table 1. A model of usability measurement [4]

Component	**Quantification**	**Measurement Method**	**Data-Collection Technique**
Efficiency	Average time needed to perform five specified tasks	User brought to lab, given list of the tasks, and performs them without help	Stopwatch (with rules for when to start and stop the watch)

3.3 Usability data

Usability data is the concrete result of measuring. Table 2 has the measurement model supplemented with a data field. In the measure in question the usability data is the actual average time the user needs to perform the five specified tasks, in this example 30 minutes.

Table 2. A model of usability measurement, data perspective

Component	**Quantification**	**Measurement Method**	**Data-Collection Technique**	**Data**
Efficiency	Average time needed to perform five specified tasks	User brought to lab, given list of the tasks, and performs them without help	Stopwatch (with rules for when to start and stop the watch)	30min

Now that the time taken is known one can consider if it is too high and if some actions need to be taken in order to reduce it. This measure in itself does not tell if it is too high or that if there are usability improvements to be made. So it is important not to take the numbers at their face value but rather consider the whole metrics set of data as a whole.

3.4 General metrics and frameworks

Frameworks are predefined sets of metrics to measure the wanted parts of usability, usually the “whole set”. Some frameworks are more general and can be applied to any system in any environment, like the System Usability Scale (SUS) described in 3.4.1 and others are made for a specific setting in mind, for example TURF which is more closely described in section 5.

There are general metrics which don’t necessarily belong to any specific framework or are considered in multiple frameworks. These are metrics that can be used in different contexts but which are not generalised. For example, task time can be measured in an accounting software and access control system, but they are not comparable to each other as the tasks differ radically. [18]

3.4.1 System Usability Scale

The System Usability Scale is “a quick and dirty usability scale” according to its creator John Brooke. It was formed based on the need to create a generalised usability tool for cross-system comparison. Brooke argues that it is a nearly impossible task like “comparing apples and oranges” but that similarities can be found from the view of user satisfaction. [18]

SUS was constructed by collecting answers to 50 potential questionnaire items and then choosing the ones with the most extreme responses on a 5 point Likert scale depicted in Figure 5. [18]

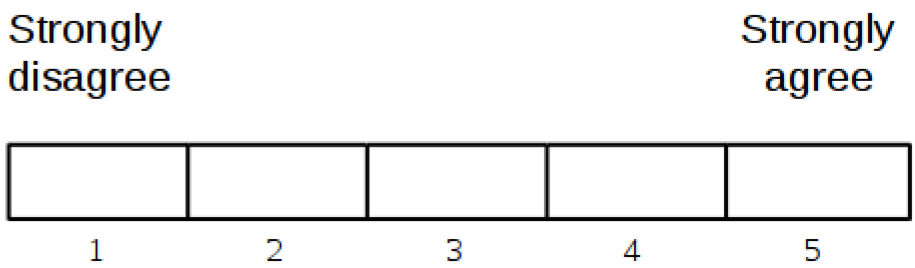

Figure 5. 5 point Likert scale [18]

The 10 questions that were chosen are[18]:

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 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.

SUS is scored from 0 to 100, 100 being the best score. For odd questions the question score is the scale position minus 1 and for even items the contribution is 5 minus the scale position. So for example:

- the question 1 gets a scale position 5 (strongly agree) the score is 5-1=**4**
- question 2 gets a 4 position so the contribution is 5-4=**1**.

None of the question scores can be considered individually. The calculated scores are added together and then multiplied by 2.5. Based on research a SUS score of 68 is the average. Jeff Sauro argues that as the SUS score is not a percentile rank it should be normalised to such for better understanding. A SUS score of 70 could be easily be wrongly interpreted as more usable than average when in truth it is very close to the average score so it would be better to say the application is 50% usable. Figure 6 presents Sauro's view of how the SUS scores correspond to percentile rank and letter grading.[19]

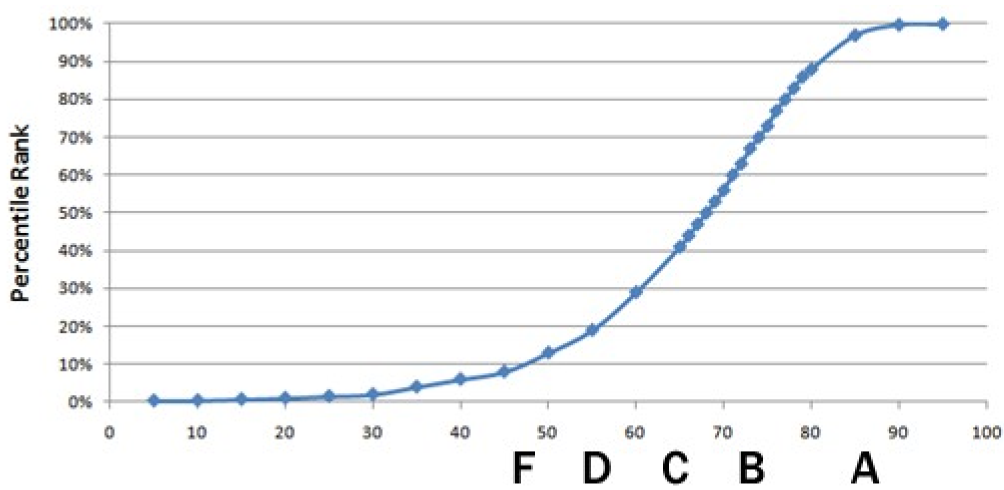

Figure 6. Percentile ranks associated with SUS score [19]

In order to get an A (the top 10% of scores) the SUS score needs to be above 80.3. Anything below a 51 score is an F (bottom 15%)[19].

Bangor et. al. have done an empirical evaluation of SUS. Their results from the analysis of a large number of SUS scores show that the SUS is a highly robust and versatile tool for usability professionals. [20]

3.4.2 Fun Toolkit

Fun Toolkit was originally designed as a usability survey for children. The Smileyometer presented in Figure 7 is basically visualised form of a Likert scale. [21] It is used as to describe how someone feels after using a system and its main features are that it is easy and quick to complete, requires limited reading ability, and requires no writing. [21] Even though the toolkit itself is designed to measure the fun in children-computer interaction the Smileyometer or shorter variation of it can be used as a short satisfaction questionnaire.

Figure 7. Fun Toolkit Smileyometer (from the left awful, not very good, good, really good, brilliant)[21]

4 Usability in healthcare information systems

This section goes through what healthcare information systems (HCIS) are and then continues to explain what specialities prenatal HCIS requires in regards of measuring usability. In the end of the section is a summarised view of different metrics.

Healthcare information systems are information systems designed to help medical personnel in their work by either giving knowledge base or to keep track of patient's medical record. In this work the latter is considered.

In Finland medical record was long recorded by hand on paper in a narrative form. By the end of 2007 all medical records were to be written in electronic format dictated by the Finnish Government Resolution[22].

Lehtokari [23] has done a research on the subject on how information is recorded both in Finnish healthcare and abroad. It gives indications that the transition to electronic format was done 1 to 1 with the old paper forms which is not necessary the best way of designing an electronic form. Lehtokari also found out that even though electronic forms reduce the time used in recording data and increase its accurateness compared to manually recording on paper they decrease the user's satisfaction. The users feel that they could potentially achieve better results with electronic recording than they do with the system they are currently using. [23]

4.1 Prenatal HCIS

Prenatal healthcare in Finland consists of regular public health nurse visits and two doctor appointments during the pregnancy, giving birth, puerperium and post inspection done by a doctor and public health nurse together. Table 3 presents the various environments and users the prenatal HCIS needs to cover and the timeframe in which they appear.

Table 3. Structure of prenatal healthcare appointments in one Finnish healthcare district [24]

8.-10. pregnancy week	Prenatal clinic	Public health nurse
11.-12. pregnancy week	Hospital	(Ultrasound) nurse
13.-18. pregnancy week	Prenatal clinic	Public health nurse
13.-18. pregnancy week	Prenatal clinic	Doctor
21.-22. pregnancy week	Hospital	(Ultrasound) nurse
22.-24. pregnancy week	Prenatal clinic	Public health nurse
27.-28. pregnancy week	Prenatal clinic	Public health nurse
30.-32. pregnancy week	Prenatal clinic	Public health nurse
35.-36. pregnancy week	Prenatal clinic	Doctor & Public health nurse
37.-41. pregnancy week	Prenatal clinic	Public health nurse
Giving birth	Hospital	Midwife & Doctor
Puerperium	Hospital	Midwife, Doctor & Nurse
Post check up	Prenatal clinic	Doctor & Public health nurse

Table 3 shows that there are 2 different environments where the prenatal HCIS is used: actual prenatal clinic and hospital. Also 4-5 user groups can be found: public health nurse, doctor, midwife, nurse/ultrasound nurse. Also worth noticing is the frequency in which the system is used as it requires the system to be memorable The whole system has been developed with this process structure in mind. Table 4 describes what the user groups need from the system.

Table 4. Description of user groups

User group	**Description**	**Special needs required from the system**
Public health nurse	Works at a prenatal clinic. Fills in the pregnancy form.	Needs the system to support workflow as in be fillable in the order they gather the pregnancy data(weight of the mother etc.).
Doctor	Does medical check ups on the expectant mother. Mainly works elsewhere.	System needs to be easy to use and easy to remember.
Midwife	Works at the hospital. Assists with labours.	Needs the system to be flexible and fillable in the “quiet periods” of labour.
Nurse	Works at the hospital. Responsible of ultrasounds and puerperium. Fills pregnancy and child forms.	Needs the system to be quick to use as they mainly work simultaneously with multiple patients/customers.

There are currently multiple information systems at use in prenatal healthcare. Mama, Haikara and I-Pana to name a few. They are used alongside the standard pregnancy form. [23] For example, I-Pana is a form of enterprise resource planning system designed to the needs of prenatal healthcare from prenatal clinic pregnancy control calls to hospitals puerperium. It is meant to be integratable with the other hospital information systems in order to keep concurrent information to a minimum. I-Pana also aims to unify and automate routine documentation. Rajala [25] followed its deployment and got some user comments that suggested that it is slow in some parts and complex and illogical in others. The comments also suggested that the system is not as integrated as it could be. [25]

4.2 Usability aspect in HCIS

Perhaps the main aspect HCIS adds on usability is that they might have a high impact on the patient's well being. If the users makes an error due to usability issues there might even be lives at risk. In the case of prenatal HCIS there is not only the well being of the mother but also the well being of the fetus (or fetuses) to be considered.

This does not bring anything new per se to the definition of usability. It gives a slight emphasis to the Errors variable of usability defined in section 2. So one might want to consider and classify errors more closely in the usability testing and minimize them in the developmental phase of the IS.

This speciality that healthcare adds on usability has been considered before. Kushniruk et. al. [26] present that there is three dimensions to HCIS usability[26]:

1. compatibility between clinical systems and physicians' tasks
2. support for information exchange, communication and collaboration in clinical work
3. interoperability and reliability

Item 1 is self-explanatory; the system needs to support what the user wants to do. As same data record is accessed by multiple users even from different user groups item 2 is important. Item 3 stems from the fact that there is often multiple different systems that the user needs to control simultaneously.

Healthcare Information and Management Systems Society (HIMSS) have made a usability maturity model proposition. Table 5 lists their view on how usability affects health organisations. [27] The value of usability to health organisations goes along the same lines as the value of usability in general listed in section 2.

Table 5. Elements of the value of usability to health organisations [27]

Increase Organisational Efficiencies	Increased Individual Effectiveness, Efficiencies
• Decreased Maintenance costs • Decreased Customer and Individual Training and Support Costs • Decreased Development Time/Costs	• Increased User Productivity/ Efficiency • Decreased User Errors/Increased Safety • Improved Cognitive Support
Improved patient, provider, organisational outcomes	

4.2.1 General measures in healthcare context

Kopanitsa et.al.[11] have made a literature review on what measures have been used to evaluate the usability of HCIS. Table 6 presents their findings in a summarised way giving examples of performance metrics and their possible applications. [11]

Table 6. Usability evaluation metrics for health records systems [11]

Type	Metric Class	Performance Metrics	Healthcare application
Efficiency	*Essential Efficiency* Estimates how closely a given user interface design approximates the ideal expressed in the use case model	Time to complete tasks: − % of tasks totally completed, − % of tasks half completed; − Comparison of task completion quality with/without software.	− Creating a Chart Note − Scheduling a Patient Visit − Prescribing a drug − Finding a Patient in the Data Base − Sending a Secure Message to a Patient
Effectiveness	*Layout Appropriateness* Favors arrangements where visual components that are most frequently used in succession are closer together, reducing the expected time of completing a mix of tasks	Surveys: − % of participants who respond they can always, most of the time, rarely, or never perform representative tasks.	− Prescribing a drug − Ordering a Lab Test − Handling a Drug-Drug Interaction Alert − Screening/Prevention − Finding a Patient in the Data Base
	Task Concordance Measures how well the expected frequencies of tasks match their difficulty, favors a design where more frequent tasks easier are made easier (e.g., fewer steps)	− % of participants able to complete tasks; − Comparison of task completion ability with software to task completion ability without software; − % of errors.	− Creating a Chart Note − Prescribing a drug − Screening/Prevention − Finding a Patient in the Data Base
Satisfaction	*Task Visibility* The proportion of interface objects or elements necessary to complete a task that are visible to the user	− number of positive/negative comments; − % of participants who made positive/negative comments;	−Scheduling a Patient Visit − Ordering a Lab Test − Handling a Drug-Drug Interaction Alert − Screening/Prevention − Sending a Secure Message to a Patient

4.2.2 General frameworks in healthcare context

Agency for Healthcare Research and Quality (AHRQ) have considered in their EHR Usability Toolkit what existing usability evaluation methods can be used in HCIS context. They point methods like heuristic evaluation, cognitive walkthrough and usability questionnaires as well as laboratory testing. [15]

They suggest that usability questionnaires would provide the basis of a toolkit to give insight into usability issues with systems currently in use. They continued to inspect different questionnaires. Table 7 shows a short review on their findings on the non proprietary ones. SUS is the best match for the case at hand as it measures the overall usability, is short, easily administered and has good reliability and validity. CSUQ(Computer System Usability Questionnaire) has higher reliability rate but as it is longer it would take more time to fill. ASQ would be even shorter than SUS and EUCS (End-User Computing Satisfaction Questionnaire) is almost as quick but as they do not measure the overall usability of the system they were discarded. USE(Usefulness, Satisfaction, and Ease of Use Questionnaire) and PUTQ were not considered due to their reliability rate and length. SUS is described as[15]:

- *SUS is a short non proprietary questionnaire, could be easily self-administered, and has good reliability and validity. However, it only measures overall usability and is not comprehensive enough for determining usability issues with an EHR.*

Which suggest that used alone it is not comprehensive enough but as a part of metrics it gives insight to the overall usability and other measures can be used to decipher the meaning of the SUS value.

Table 7. Review from Usability Toolkit[15]

Name	Length (number of items)	Reliability	Overall usability	Learnability	Efficiency	Effectiveness	Satisfaction
SUS	10	0.85-0.91	X	X	-	-	-
ASQ	3	0.93	-	-	X	X	X
USE	30	Not available	-	-	-	-	X
PUTQ	100	0.59-0.81	-	-	-	-	-
EUCS	12	0.92	-	-	X	X	-
CSUQ	19	0.95	X	-	-	-	-

4.2.3 Healthcare specific usability frameworks

There are frameworks designed healthcare in mind. In this work a framework called TURF is described more closely although there are others as well. For example, US National Institute of Standards and Technology (NIST) have a framework or protocol, as they name it, for measuring usability of electronic health record (EHR). However that framework does not give relative or comparable score of usability rather it gives a guideline on how to approach measuring usability of EHR and dictates that each case involves the development of unique metrics. [28]

TURF is a unified framework for electronic health record usability. The name comes from the four key components of usability it uses: Task, User, Representation and Function. TURF defines usability as "*how useful, usable and satisfying a system is for the intended users to accomplish goals in the work domain by performing certain sequences of tasks*". This differs slightly from the definition of usability from section 2 so TURF cannot be directly used in this context. [29] Table 8 shows the dimensions and corresponding measures TURF uses. Figure 8 shows how the framework is constructed and which components affect which part of usability.

Table 8. Dimensions and measures of usability under TURF [29]

Dimensions	Descriptions	Representative measures
Useful	A system is useful if it supports the work domain where the users accomplish the goals for their work, independent of how the system is implemented	• Across-model Domain Function Saturation: Percentage of domain functions in the EHR vs. all domain functions in the work domain • Within-model Domain Function Saturation: Percentage of domain functions over all functions (domain and non-domain) in the EHR
Usable	A system is usable if it is easy to learn, easy to use, and error-tolerant.	• Learnability • Number of trials to reach a certain performance level • Number of items that need to be memorized • Number of sequences of steps that need to be memorized • Efficiency • Time on task • Task steps • Task Success • Mental effort • Error Prevention and Recovery • Error occurrence rate • Error recovery rate
Satisfying	A system is satisfying to use if the users have good subjective impression of how useful, usable, and likable the system is	• Various ratings through survey, interview, and other instruments

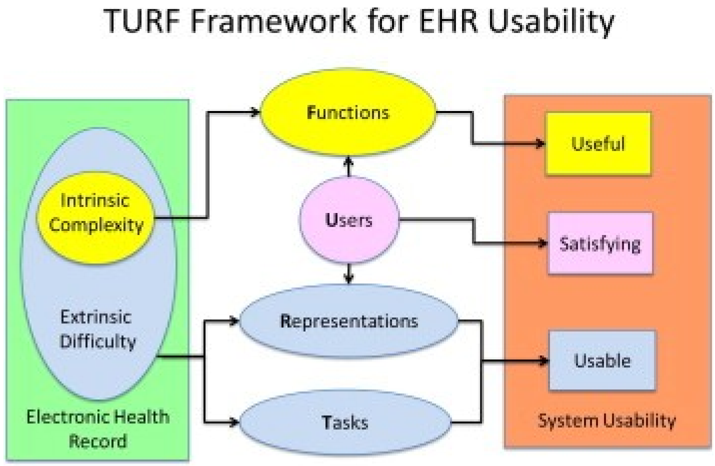

Figure 8. TURF framework [29]

4.3 Summary of healthcare usability metrics

Table 9 sums up the frameworks and general measures considered in this section. It shows how each described framework or measure corresponds to the types/parts of usability defined in section 2. As can be seen, none of them fit for all of the types. The general measures of table 6 and representative measures of TURF in table 7 have similarities. TURFs definition of usability and the types of it is different from the one used in this work so even though it seems to cover almost all parts it cannot be used as such.

Table 9. Summary of presented measures

	TURF	SUS	General measures
Efficiency	X	-	X
Effectiveness	-	-	X
Satisfaction	X	-	X
Errors	X	-	-
Memorability	X	-	-
Learnability	X	X	-
Overall	-	X	-

In this work it is in order to take a leaf out of NISTs book and combine a unique set of metrics based on the case needs. The general measures can be supplemented with ones from TURF and then combined with SUS to get an overall view of the usability of the system.

5 Usability metrics for prenatal HCIS of one Finnish healthcare district

Section 4 went through measuring usability in a healthcare information system context and gave a few examples of how it has been done. This section is about how that knowledge can be used in the context of prenatal HCIS of one Finnish healthcare district. This section proposes a set of metrics that can be used. Also a way of collecting them is laid out.

5.1 Case analysis

The setting of the case is prenatal healthcare. The section 4.1 presented that there are 4-5 possible user groups using the same system in two distinct environments. Of these user groups three were selected:

1. the public health nurses
2. doctors and
3. midwives.

Public health nurses were selected on the basis that they are the user group that most commonly uses the system. Nurses were discarded as they use the system rarely and resembling the way public health nurses use it. Doctors and midwives were selected because they use the system very differently from the public health nurses and they most often have overlapping interfaces/systems to use.

Task analysis and user observation resulted that the basic way to use the system is an appointment with the patient, in this named as “customer call” as there are no patients per se in the given context as the expecting mothers are not sick.

Figure 9 presents the flow of the customer call. First the user needs to check their own user info and if the workstation they use is correctly put for logging purposes. This is because the users do not have any specific workstation given to their use only and they might use any available computer. Then they need to access the patient record and check with the patient if the data is correct one.

The system prenatal healthcare uses is form based and used in other healthcare contexts as well. So the user needs to navigate through a file tree of different forms in order to access the pregnancy form they use. Then the customer call concentrates on getting the data that is recorded on the form and the form is filled while doing do. For example the weight of the mother and the heartbeat of the fetus is measured and then recorded by hand to the system in appropriate place on the form. So this interrupts the workflow and time is spent with the customer rather than using the system.

If applicable the user might need to access another system to check laboratory results. Finally the next customer call time is reserved (not present in the figure 9) and the patient record is saved and closed.

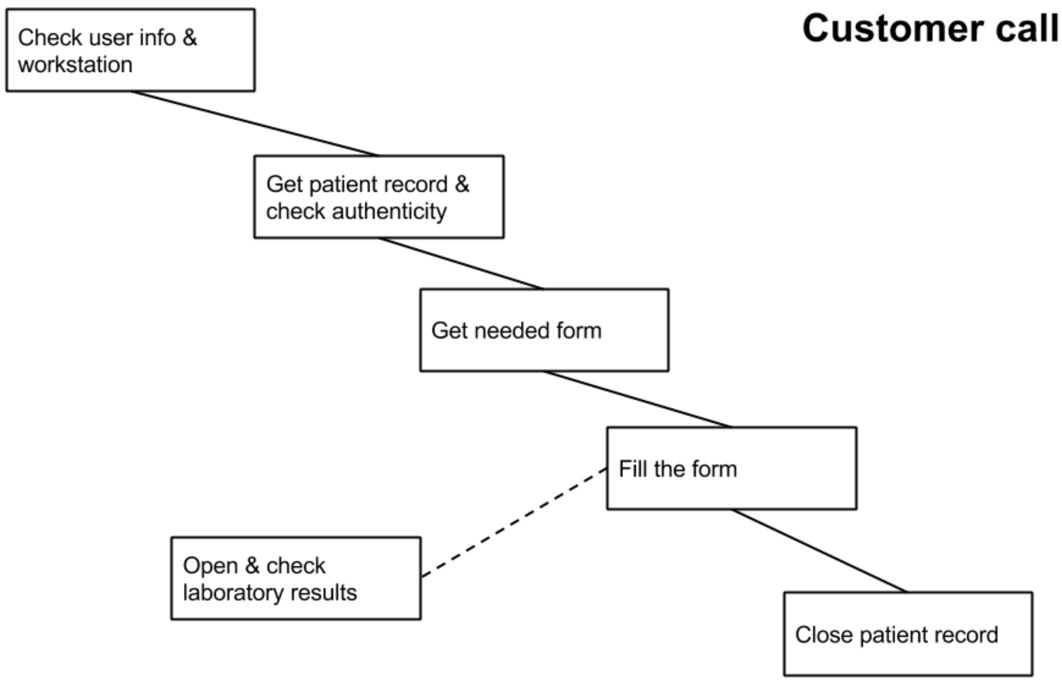

Figure 9. Task analysis of customer call

5.2 Proposed composition of metrics

Various measures are needed in this work as all parts of usability that were defined in section 2 (*effectiveness, efficiency, learnability, errors, memorability, satisfaction*) are wanted to be gathered. First a set of general metrics and the way they are to be measured is defined. Then some user questionnaires are needed to get the user perspective.

Any specific healthcare usability framework was not used. TURF defined in section 4.2.3 was considered but as its definition of usability differed from the one defined in section 2 it was not used as such. It is however used as a guideline for the general metrics picked along with the general metric listing of section 4.2.1.

5.2.1 General metrics

Table 10 presents proposed metrics based on section 4 and the case analysis of section 5.1. It shows the measure, method to measure it, the user group it focuses on and the type of usability it reflects.

So the focus groups, as in user groups the measure focus on, are:

- midwives(MW)
- doctors(D)
- public health nurse(PHN)

A task is a session with focus group person (MW, D or PHN) and consists of smaller tasks like inputting data to a form. Observing means that an observer is present at a customer call and observes the work of the focus group person present. Observer takes notes on paper with a pen (p&p) and if needed clocks time taken with a stopwatch while the user narrates their actions with think aloud protocol. Total cumulated time is estimated from other measures. Persistent error is an error that is present throughout a task or appears within a certain task every time the task is performed.

Basically all measures are aimed at each focus group. Satisfaction is also measured with questionnaire defined in section 5.2.2.

Table 10. Proposed metrics with measurement method, focus group and measure type

Measure	Measurement method	Focus group	Type
1. Time to perform a task	Observing, clock, p&p	MW/D/PHN	Efficiency
2. Number of key presses to perform a task*	Observing, p&p	MW/D/PHN	Efficiency
3. Time taken to correct errors while performing a task	Observing, clock, p&p	MW/D/PHN	Errors
4. Total cumulated time (/per use case/per day)	Estimate	MW/D/PHN	Efficiency
5. The number of different systems the user needs to log in to to perform a task	Observing, p&p	MW/D/PHN	Efficiency
6. The number of times the user needs to input concurrent data to perform a task**	Observing, p&p	MW/D/PHN	Efficiency, Effectiveness
7. Percentage of tasks completed successfully on first attempt	Calculated, observed, p&p	MW/D/PHN	Learnability, Effectiveness
8. Per cent of users who can carry out key tasks without guidance	Questionnaire	MW/D/PHN	Learnability, Memorability
9. Number of persistent errors	Observing, p&p	MW/D/PHN	Errors
10. Number of occurrence of persistent errors	Observing, p&p. estimate	MW/D/PHN	Errors
11. Per cent of users who feel "in control" of the product	Questionnaire	MW/D/PHN	Satisfaction
12. User rating on a 4-point scale anchored with "makes me more/less productive"	Questionnaire	MW/D/PHN	Efficiency

* Key presses that are used to navigate the system and forms

* *Excluding login-information

5.2.2 Usability questionnaire

Personnel at the healthcare district have had a “feeling” that the system they use is not satisfactory nor simple to use. They want data to back up this feeling but as they do not have time for long questionnaires so a quick and easy to answer questionnaire was needed to capture the data.

SUS presented in section 3.4.1 is a questionnaire fit for this task. It has been widely used in different context and has been proven to work in healthcare context as section 4.2.2 pointed out. As some visual measurement was also wanted Fun Toolkit with smileys is used as a simple but effective visual measure. It was added to the end of the standard SUS questionnaire.

Measures 8, 11 and 12 of general metrics from 5.2.1 are gathered as a simple, separate questionnaire after each customer call observation.

5.3 Testing set up

Testing is performed in real life customer call situation. As to not disrupt the work of medical personnel too much and manage the workload of the observer it was decided to divide the observed metrics into two different groups shown in table 11 by the similarity of the measures. Also this grouping lets the observer more time to take notes of the way the metrics work. The observer is present at the customer call and records all the measurement data by hand while the user narrates their usual tasks with think aloud method.

Table 11. Script of measuring appointments

Time	**Target**	**Focus group**
Appointment 1	1, 3, 9, 10	MW/D/PHN
Appointment 2	2, 5, 6, 7	MW/D/PHN
Appointment 3	1, 3, 9, 10	MW/D/PHN
Appointment 4	2, 5, 6, 7	MW/D/PHN
Appointment 5	1, 3, 9, 10	MW/D/PHN
Appointment 6	2, 5, 6, 7	MW/D/PHN

6 Results

This section presents the results of testing the metrics suggested in section 5. Based on the results of that testing a revised set of metrics for use of the healthcare district is suggested.

6.1 Metric testing

The designed metrics were tested in real life customer call situations. Testing aimed to capture 6-10 appointments from each focus group but due to the unpredictability of the testing environment this amount was not achieved.

The testing caught:

- 6 public health nurse appointments and
- 2 labours (midwife customer calls).

SUS questionnaire was distributed through personnel mailing lists but it didn’t receive any answers.

6.2 Revised metrics

Table 12 presents the revised metrics proposed for the use of prenatal healthcare unit of the Finnish healthcare district. Focus groups for these metrics are midwives and public health nurses. Doctors were ruled out as they were not available for the metric testing. SUS questionnaire was discarded from the metrics due to not providing any additional information.

Table 12. Metrics with measurement method, intended focus group and measure type

Measure	Measurement method	Focus group	Type
1. Time to perform a task	Observing, clock, p&p	MW/PHN	Efficiency
2. Number of key presses to perform a task*	Observing, p&p	MW/PHN	Efficiency
3. Time taken to correct errors while performing a task	Observing, clock, p&p	MW/PHN	Errors
4. Total cumulated time (/per use case/per day)	Estimate	MW/PHN	Efficiency
5. The number of different systems the user needs to log in to to perform a task	Observing, p&p	MW/PHN	Efficiency
6. The number of times the user needs to input concurrent data to perform a task**	Observing, p&p	MW/PHN	Efficiency, Effectiveness
7. Percentage of tasks completed successfully on first attempt	Calculated, observed, p&p	MW/PHN	Learnability, Effectiveness
8. Per cent of users who can carry out key tasks without guidance	Questionnaire	MW/PHN	Learnability, Memorability
9. Number of persistent errors	Observing, p&p	MW/PHN	Errors
10. Number of occurrence of persistent errors	Observing, p&p. estimate	MW/PHN	Errors
11. Per cent of users who feel "in control" of the product	Questionnaire	MW/PHN	Satisfaction
12. User rating on a 4-point scale anchored with "makes me more/less productive"	Questionnaire	MW/PHN	Efficiency

* Key presses that are used to navigate the system and forms

* *Excluding login-information

7 Discussion

This section goes through notes the observer made during customer calls and how the metrics suite the case. Also some future actions are proposed.

In observation it turned out that the defined customer call only a little resembles the actual way the users work. They do not for example check the user information nor do they check with the patient if the data is correct. Also the order in which they fill the data to the pregnancy form is completely wanton and depends on the person using the system.

In light of this the healthcare district would do well with standardising the whole practice first and then thinking how it is turned into a usable information system. Now every public health nurse and midwife fill in the system at different times of patient appointment and do every measurement (weighing etc.) at different times which do not necessarily abide by the order set by the forms resulting the possibility of time wasted on multiple clickings. Therefore in order to design proper metrics to measure the usability aspects of a system it is needed to make at least a rough sketch how a patient appointment is going to be performed.

The answer rate of SUS questionnaire was a disappointment. It is unclear whether the questionnaire reached the mailing lists it was aimed to or was it just that it got buried in the amount of other e-mail traffic. Might be that it reached the focus groups but it was discarded because not enough information was presented for the personnel about the questionnaire from the healthcare districts part. So nothing can be said of the suitability of the SUS itself as no data was managed to be gathered. But it is clear that the distribution method of the questionnaire needs to be changed. Perhaps it needs some preset testing event from the healthcare districts part. The participation cannot be left to the free will of the complete focus group as they perhaps do not have the opportunity to answers it among their daily work. As it is, it was decided to be discarded from the metrics.

If the testing is to be performed in real life situations there needs to be more time to reserved to gather the needed amount of appointments. Perhaps it could be done with getting a few of interested pregnant mothers and follow their appointments throughout their pregnancy. Also it needs to be better organised from the healthcare districts part as this time around most of the focus group personnel affected by this research were not aware of it. Every appointment the observer got to was a gamble as it was not certain if the customer was content with the observer being around.

The gathering method of observer noting everything down with pen&paper is not foolproof as there were distractions present at the appointments, for example customers older children wanted to play with the observer. This could be counterbalanced with increasing the control of the meetings/observations, for example by adding another observer or increasing the amount of appointments observed.

It was noted that the metrics might not gather all the errors present at the system. For example in the public health nurse appointments the system automatically had a tap on that every appointment was a first one even though they present the minority of appointment types. This is more like a design fault than an error so it slips past current metrics.

The most time used with the system was free form writing. It was noted that most of the observed personnel were using what might be called “two finger system” while typing with keyboard. Maybe the personnel training (or even the actual medical schooling) should include some sort of general computer skills training like a typing course.

Linked to the user's general computer skills was the use of keyboard and mouse in the navigation of the system. Some users used tabulator to navigate the fields of the form when others used mouse. It gathered to the number of clicks/keystrokes the user had while performing a task. It may affect the time the user uses but it needs more research.

It is also open to discussion whether the testing in real life environment reveal anything new or "better" information on usability faults compared to a testing done in a usability lab where a usability testing system such as Morae could be used to log the keystrokes etc.

To summarise research questions were as follows:

1. What kind of guidelines and frameworks are used to measure usability?
2. What usability metrics have been used in healthcare domain?
3. What usability metrics can be used in this particular healthcare instance?
   a. What different subparts does the suggested usability metrics consist of?
   b. In what scope can the measurements be made?
4. What metrics were successful in this study?

RQ1 was answered in section 3, RQ2 in section 4. Answer to RQ3 was presented in section 5. To answer RQ 4 the metrics collected by post observation questionnaire with public health nurses got the most definite answers. So the measures 9, 11 and 12 were successful in this study. But as they are only a small part it is still proposed to use the whole set of 12 metrics.

Mostly this research suffered from so called Chinese whispers. The information about the research did not reach the actual participating users, nor did the information about possible appointment times or changes to them reached the observer. Also the timetable was too rushed. So given more time and better informed participants these metrics might possibly work, at least with the public health nurse appointments.

As no doctors was caught on this testing, it is not clear if these metrics work for them. Midwives use the system when they can between the stages of labour, which can be hours in regards of time, the collecting method of these metrics is difficult for the observer.

8 Conclusion

This Thesis aimed to create a set of usability metrics that could be used in the prenatal healthcare unit of one Finnish healthcare district. First the concept of usability was defined, then the ways in which it can be measured. Then the healthcare information systems were defined and the work continued on to show ways in which the usability of these can be measured.

A metric set of 12 measures was created and supplemented with a System Usability Scale questionnaire and a visual Smileyometer from Fun Toolkit. The metrics were tested out in real life patient situations: 6 prenatal healthcare nurse customer calls and 2 midwife customer calls were observed. Based on the observations a revised set of metrics consisting purely of the initial 12 measures was proposed.

Nothing certain can be said about the suitability of these metrics. By increasing the personnel involvement, maintaining information flow and increasing the time available for testing more solid data could be gathered. In future it also needs to be defined if the real life situations bring out more usability faults compared to a laboratory testing.

REFERENCES

[1] International Organization for Standardization, “International Standard: ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 11: Guidance on Usability.” 1996.

[2] I. Sinkkonen, H. Kuoppala, J. Parkkinen, and R. Vastamäki, *Käytettävyyden psykologia*. .

[3] “Measuring the User Experience | 978-0-12-415781-1 | Elsevier.” [Online]. Available: https://www.elsevier.com/books/measuring-the-user-experience/albert/978-0-12-415781-1. [Accessed: 11-May-2015].

[4] J. Nielsen, *Usability Engineering*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[5] B. Middleton, M. Bloomrosen, M. A. Dente, B. Hashmat, R. Koppel, J. M. Overhage, T. H. Payne, S. T. Rosenbloom, C. Weaver, J. Zhang, and American Medical Informatics Association, “Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA,” *J. Am. Med. Inform. Assoc. JAMIA*, vol. 20, no. e1, pp. e2–8, Jun. 2013.

[6] R. Pereira, J. Duarte, M. Salazar, M. Santos, A. Abelha, and J. Machado, “Usability of an electronic health record,” in *Industrial Engineering and Engineering Management (IEEM), 2012 IEEE International Conference on*, 2012, pp. 1568–1572.

[7] P. J. Edwards, K. P. Moloney, J. A. Jacko, and F. Sainfort, “Evaluating usability of a commercial electronic health record: A case study,” *Int. J. Hum.-Comput. Stud.*, vol. 66, no. 10, pp. 718–728, Oct. 2008.

[8] J. Viitanen, H. Hyppönen, T. Lääveri, J. Vänskä, J. Reponen, and I. Winblad, “National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability,” *Int. J. Med. Inf.*, vol. 80, no. 10, pp. 708–725, Oct. 2011.

[9] J. Viitanen, A. Kuusisto, and P. Nykänen, “Usability of electronic nursing record systems: definition and results from an evaluation study in Finland,” *Stud. Health Technol. Inform.*, vol. 164, pp. 333–338, 2011.

[10] International Organization for Standardization, “International Standard: Ergonomics of Human-System Interaction, Part 210: Human-Centred Design for Interactive Systems.” 2010.

[11] G. Kopanitsa, Z. Tsvetkova, and H. Veseli, “Analysis of metrics for the usability evaluation of electronic health record systems,” *Stud. Health Technol. Inform.*, vol. 174, pp. 129–133, 2012.

[12] “Usability Testing Software | Morae from TechSmith.” [Online]. Available: https://www.techsmith.com/morae.html. [Accessed: 12-May-2015].

[13] M. W. M. Jaspers, “A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence,” *Int. J. Med. Inf.*, vol. 78, no. 5, pp. 340–353, May 2009.

[14] L. van Velsen, T. van der Geest, and R. Klaassen, “Testing the usability of a personalized system: comparing the use of interviews, questionnaires and thinking-aloud,” 2007, pp. 1–8.

[15] Johnson, C.M, Johnston, D, Crowley, P.K, Culbertson, H, Rippen, H.E, Damico, D.J, Plaisant, and C, "EHR Usability Toolkit: A background report on usability and electronic health records." [Online]. Available: http://www.rti.org/publications/abstract.cfm?pubid=18504. [Accessed: 11-May-2015].
[16] J. R. Lewis, "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use," *Int J Hum-Comput Interact*, vol. 7, no. 1, pp. 57–78, tammikuu 1995.
[17] H. X. Lin, Y.-Y. Choong, and G. Salvendy, "A proposed index of usability: A method for comparing the relative usability of different software systems," *Behav. Inf. Technol.*, vol. 16, no. 4–5, pp. 267–277, tammikuu 1997.
[18] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
[19] "Measuring Usability with the System Usability Scale (SUS): MeasuringU." [Online]. Available: http://www.measuringu.com/sus.php. [Accessed: 11-May-2015].
[20] A. Bangor, P. T. Kortum, and J. T. Miller, "An Empirical Evaluation of the System Usability Scale," *Int. J. Hum.-Comput. Interact.*, vol. 24, no. 6, pp. 574–594, Jul. 2008.
[21] J. C. Read, "Validating the Fun Toolkit: an instrument for measuring children's opinions of technology," *Cogn. Technol. Work*, vol. 10, no. 2, pp. 119–128, Apr. 2008.
[22] *Valtioneuvoston periaatepäätös terveydenhuollon tulevaisuuden turvaamiseksi*. Sosiaali- ja terveysministeriö, 2002.
[23] O. Lehtokari, "Hoitotyön tiedonhallinta äitiyshuollossa."
[24] "Eksote - Äitiysneuvolat." [Online]. Available: http://www.eksote.fi/fi/perhejasosiaalipalvelut/neuvolat/aitiysneuvolat/sivut/default.aspx. [Accessed: 11-May-2015].
[25] N. Rajala, "iPana-tietojärjestelmän ja RAFAELA-hoitoisuusluokituksen käyttöönotto ja toimivuus hoitohenkilökunnan kokemana," 2013.
[26] A. Kushniruk, J. Kaipio, M. Nieminen, C. Nøhr, and E. Borycki, "Comparing approaches to measuring the adoption and usability of electronic health records: lessons learned from Canada, Denmark and Finland," *Stud. Health Technol. Inform.*, vol. 192, pp. 367–371, 2013.
[27] N. Staggers and M. Rodney, "Promoting Usability in Organizations with a New Health Usability Model: Implications for Nursing Informatics," *NI 2012 Proc. 11th Int. Congr. Nurs. Inform.*, vol. 2012, Jun. 2012.
[28] S. Z. Lowry, M. T. Quinn, M. Ramaiah, R. M. Schumacher, E. S. Patterson, R. North, J. Zhang, M. C. Gibbons, and P. Abbott, "Technical evaluation, testing, and validation of the usability of electronic health records," *Natl. Inst. Stand. Technol.*, 2012.
[29] J. Zhang and M. F. Walji, "TURF: Toward a unified framework of EHR usability," *J. Biomed. Inform.*, vol. 44, no. 6, pp. 1056–1067, Dec. 2011.