UNIVERSITY OF VAASA

SCHOOL OF TECHNOLOGY AND INNOVATIONS
INFORMATION SYSTEMS

Oskar Kenttälä

EMOTIONAL INTERACTION DESIGN
Effects of Negative Affective Interactions on User Experience Design

Master's Thesis in
Information Systems

Master's Programme in Information Systems

VAASA 2019

PICTURES

TABLES

ABREVIATIONS

Human Computer Interaction	HCI
User Interface Design	UID
User Experience Design	UXD
Interaction Design	IxD
Affective Design	AD
School Information System	SIS
Management Information System	MIS
Graphical User Interface	GUI
Software as a Service	SaaS

UNIVERSITY OF VAASA
Faculty of Technology
Author: Oskar Kenttälä
Topic of the Master's Thesis: Emotional Interaction Design: Effects of Negative Affective Interactions on User Experience Design
Instructor: Tero Vartiainen
Degree: Master of Science in Economics and Business Administration
Major: Information Systems
Year of Entering the University: 2016
Year of Completing the Master's Thesis: 2019 **Pages:** 102

ABSTRACT

User interfaces and user experience is a modern phenomenon that has gained visibility with the rise of the globally connected internet era, especially with the many failings connected to poor usability. It is not uncommon to encounter applications, systems and products that are designed from a point of view which does not relate well to the user's requirements, causing negative emotions including frustration and aggravation. In this thesis the connection between insufficient user experience design and negative end user emotions is studied, along with how this eventually reflects upon a system in general. The hypothesis is that negative aspects cause negative emotions, but the eventual effect on the system is subjective.

The theoretical framework is based upon previous theories and research done in User Experience Design, with a focus on Interaction Design related subjects such as Usability and Affective Design. The research was conducted via a two-part interview, with the first part being Participatory Heuristic Evaluation based tasks, based on Usability research, and the second part being an interview, created with an Affective Design based focus in the questions. The goal of this research was to first discover if the test material had aspects that could cause negative emotions in the user, while the interview was designed to discover if the evaluators were affected negatively by the system and how this reflected upon the system. The test was conducted by five evaluators, non-professional in usability, considered expert users for the system. The test material was Wilma, a School Information System developed by Visma.

The research indicates that the test material had indeed several different aspects that were known, based upon contemporary research, to cause negative emotions in users and all of the evaluators were similarly considered to be negatively affected by the lack of proper User Experience Design. The effects on the system were as hypothesized, subjective, with aggravation and heaviness being the eventual outcomes. The research suggests that the user's emotions can be affected by UX design and the affective design approach works well as an evaluation method to discover general aspects of a system, albeit the subjective nature of emotions can cause inaccuracy.

KEYWORDS: User Experience, Affective Design, Interactive Design, Usability, School Information System

UNIVERSITY OF VAASA

Faculty of Technology

Author: Oskar Kenttälä

Topic of the Master's Thesis: Emotional Interaction Design: Effects of Negative Affective Interactions on User Experience Design

Instructor: Tero Vartiainen

Degree: Kauppatieteiden maisteri

Major: Tietojärjestelmätiede

Year of Entering the University: 2016

Year of Completing the Master's Thesis: 2019 **Pages:** 102

TIIVISTELMÄ

Käyttöliittymät ja käyttökokemus on moderni ilmiö, joka on saanut huomattavaa näkyvyyttä globaalin internetin aikakaudella, etenkin käyttökokemukseen liittyvien ongelmien ja epäonnistumisten muodossa. Ei ole ollenkaan epätavallista, että vastaan tulee laite, ohjelma tai järjestelmä, joka on suunniteltu sellaisesta näkökulmasta mikä ei millään tavalla vastaa käyttäjän tarpeita tai odotuksia, aiheuttaen negatiivisia tunteita kuten ärsytystä ja turhautumista. Tässä tutkimuksessa selvitetään puutteellisen käyttökokemussuunnittelun ja negatiivisten tunteiden yhteyttä käyttäjään, sekä sitä miten nämä lopulta heijastuvat järjestelmään. Hypoteesina on, että negatiiviset ominaisuudet aiheuttavat negatiivisia tunteita, mutta lopullinen vaikutus on hyvin subjektiivinen.

Teoreettinen viitekehys perustuu Käytettävyyssuunnittelun aiempaan tutkimukseen ja teorioihin, etenkin Interaktiiviseen suunnitteluun ja siihen liittyviin aiheisiin, kuten Käytettävyys ja Tunnesuunnittelu. Tutkimus toteutettiin kaksiosaisena haastatteluna, ensimmäisen osan ollessa Osallistuva heuristinen arvio, joka pohjautuu Käytettävyystutkimukseen, ja toisen osan ollessa Tunnesuunnitteluun perustuva haastattelu. Tutkimuksen tavoitteena on ensin arvioida, onko tutkimusmateriaalissa ominaisuuksia, jotka aiheuttavat negatiivisia tunteita, sitten tutkia, olivatko arvioijat negatiivisesti vaikuttuneita materiaalin ominaisuuksista sekä lopuksi tutkia, miten tämä vaikutti kohdejärjestelmään. Testauksen toteutti viisi arvioijaa, jotka olivat ei-ammattilaisia käytettävyydessä, mutta olivat kohdejärjestelmän kokeneita käyttäjiä. Koejärjestelmä oli Wilma, Visman kehittämä koulunhallintajärjestelmä.

Tutkimus osoitti, että koemateriaalissa oli useita erilaisia ominaisuuksia, jotka aiheuttivat negatiivisia tunteita, sillä kaikki arvioijat olivat saaneet negatiivisia tunteita järjestelmän puutteellisen käyttökokemussuunnittelun takia. Lopulliset vaikutukset järjestelmään olivat, hypoteesin arvion mukaisesti, subjektiivisia, ärsyyntymisen ja raskauden tunteita. Tutkimuksen perusteella voidaan päätellä, että kokemussuunnittelulla pystytään vaikuttamaan tunteisiin, mutta myös sen, että tällä tunnelähtöisellä lähestymistavalla voidaan arvioida järjestelmien yleisiä ominaisuuksia, joskin tunteiden henkilökohtainen luonne saattaa aiheuttaa epätarkkuutta tuloksissa.

AVAINSANAT: Käyttökokemus, Interaktiivinen suunnittelu, Tunnesuunnittelu, Käytettävyys

1. INTRODUCTION

User interfaces are the cornerstone of our modern, technology centered, societies: they are everywhere. You can expect to find user interface from anything that has an electronic component, for example, a phone, television or simply a LED flashlight, which in several cases has more than one lighting mode. However, user interfaces are, with a less academic but accurately describing word, “tricky” as they effectively combine technology, art and psychology in a package that is required to be simple, effective and pleasing (Preece, Rogers & Sharp 2002: 2). Creating this type of a combination of software and hardware is very challenging and the challenge increases exponentially with the complexity of the object or system the interface controls. From a technical point of view, the largest challenges are to combine software and hardware functionalities in a user interface that is error free, or the very least error tolerant, maintaining full functionality and keep everything as simple, concise and consistent as possible.

After the user interface works from a technical perspective, the user’s perspective and the user experience need to be similarly considered. The design, both graphical and physical, needs to be clear, informative and pleasing to the eye, but in addition a multitude of additional requirements, including legislature, safety and even intended location of the system, need to be taken into account (Cooper, Reimann & Cronin 2007: 24-41). To maximize the effects of a system, henceforth used to describe any object, application or otherwise with a clearly definable user interface, and to be able to provide the best possible results, the user’s properties, the persona, need to be considered and implemented in the interface. When considering the different user personas, you create not only an acceptable control interface, but are also able to predict and assess possible shortcomings, use-cases and error situations in the system (Cooper et al 2007: 75-108). However, more often than not, the system’s controls are far from optimal and even small details might make a system unusable or even dangerous. Similarly, growing complexity combined with insufficient user interface design often creates a convoluted,

distracting and infuriating result which directs energy from operating the system toward attempting to control the user interface itself.

User based emotional responses towards any type of system is inherently difficult to effectively measure due to their personal subjectivity. Variation is high from person to person as the reasons behind the feelings are diverse and personal background affects the results as do many other details (Lockner & Bonnardell 2014). Some elements are of course easy to discover as being annoying, but a general feeling of irritation might not be attributed to an exact object or aspect. This affects especially the case of user interfaces due to their nature: the positive aspects of a user interface are present only when they are not seen, but the negative aspects however are seen very well. Modern user interface development has been directed toward creating a more user centered, interactive design with the feeling at ease, comfort and user experience as the desired outcome (Preece et al 2002: 141). However, this does not change the reality that user interface design is an unthankful task as it is best when it is not noticed. When a user interface does not hinder the use of the system and creates an enjoyable experience the interface is considered acceptable and pleasant, but the system as a whole is credited, not the interface. However, when the user interface is not working, the shortcomings and problems are clearly visible.

Studies show that a poorly designed user interface causes frustration, mistakes and wastes time, energy and often money on trivial details that involve using the user interface, not the system (Shneiderman & Plaisant 2005: 454-455). While this obviously affects the entire user experience of the system negatively, as stated, how does it change the users', or in many cases, the customer's view towards the system and continued use of the system as a whole? In this thesis we are studying the usability and emotions towards a system and, if present, to discover the effects of these negative aspects and emotions have towards the system as a whole. Hypothesis is that the system in question for this thesis has some causes for frustration and aggravation, and these negative aspects will have a negative effect on the user experience of the system, however, the eventual overall effect on the system is subjectively varied, instead of being directly negative.

1.1 Objectives and Limitations

User interfaces are a central part of user experience of a system. User interfaces are the link between the system and the user and thus it is designed to control the functionalities of a system in a form that is as simple, effective and as pleasing as possible while creating a pleasant user experience from using the system. While all user interfaces are created to contain these positive aspects and aimed towards generating positive emotions altogether, their presence eventually yields limited information of the user interface and user experience as they suffer from a phenomenon that could be called the invisibility of positive features: a user interface is not the center of attention in the system, only the tool to operate the system, and thus it is seen only when it is not working properly. Of course, you can discover positive aspects of a user interface, but then you are specifically looking for these aspects and the discovery is in a sense artificial. Negative aspects that affect the user experience in user interfaces, on the other hand, are exceptionally easy to discover, along with the negative emotions they create. Thus this negative feedback is also the only natural feedback a user interface designer can expect to receive and, unfortunately, this feedback is often emotionally colored and locating the real issue from spiteful language is difficult. The objective of this thesis is to study the relation between aspects of negative user experience, the emotions it creates, and what their effect is to the system as a whole. In short, the objective of this thesis is to discover usability errors in a user interface through negative feedback. To help achieve these results three research questions are established:

1. What kind of aspects generate negative emotions in a user interfaces?
2. How do these negative aspects affect the user experience of the system?
3. What are the effects of these negative emotions towards the use of the system as a whole?

While the first and the second question can and will be studied as a part of the literature review, the main focus for these questions will also be a part of the empirical research explained more in depth in the following sub-chapter. The third question is a new approach and will be a significant theme in the results of the empirical research. In other

words, first a base is established to detect and understand what negative aspects are in systems, secondly reasons why it is negative will be considered. The last question will consider how the user perceives the overall status of the system and personal thoughts of the system will be considered. This research does heavily imply that something negative will be found from the material and due to the nature of the material in question, and the history of the instance that created the material, negative aspects will most likely be found. This will be discussed more in-depth in chapter 2. The results of this research are in a large part designed to detect and discuss the shortcomings of the target material's user interface and then used to improve the design. Finally, it should be mentioned, that while emotions are a notable part of this study and their research basis is established, they are not discussed from any kind of psychological point of view, but more of a humanistic point of view and as a mediator to understand why certain aspects are troublesome and what could be done to repair the situation.

1.2 Method

The research method of this thesis is a qualitative approach based case study conducted via interview, containing a questionnaire and heuristic evaluation, at the core of the method. The interview is designed to discover emotional aspects and usability issues in a system and is constructed using a heuristic evaluation and a questionnaire focused on making evaluations on the user experience in general. The target of this research is to discover negative emotions, such as frustration or aggravation, which affect the user experience of the system. The idea of this test is to find details or aspects in the system that cause difficulty for the user or are otherwise problematic while also giving personal, emotion based thoughts about the system. It should be noted, that due to the fickle and difficult to analyze nature of emotions, the results may vary and might differ greatly due to background, personality or personal preferences.

The theoretical framework for this thesis will be based upon previous research conducted on Human Computer Interaction with a focus on User Experience in user interfaces and their design with Interaction Design. The research will specifically be

based upon a combination of Affective Design, which introduces emotion as a point of view when designing user interfaces and user experience in systems, and Usability studies which will be used as a practical basis for the actual empirical research itself. Usability is used as a basis to discover different aspects on the user interface material with its powerful, well established tools while Affective Design is used as a standing point to evaluate these discoveries.

The select tool for this thesis is a questionnaire and Heuristic Evaluation as it is a powerful, malleable and simple tool to assess and evaluate the usability and functionalities of a system, especially user interface functionalities, from an intuitive, user based point of view. The goal of the heuristic evaluation method is to assess the system towards a list of acknowledged principles of usability, in other words the heuristics, (Nielsen 1993) and then make evaluations concerning the system, i.e. is the user interface good or not. In this thesis, we are using a modified list of heuristics, as normally the heuristic evaluation process is done by professionals, which is streamlined for non-expert users. The modification is based upon Participatory Heuristics (Muller, Matheson, Page & Gallup 1998) and then combined with Affective Design principles (Cooper et al 2007, Preece et al. 2002) for best results. The modified list will include themes of Clarity, User control, Consistency, Errors and Support (Muller et al 1998), with a special notion of irritability added to the list. The exact list can be found in chapter 5. Any found issues are categorized as per Nielsen's (1994) categorization method, although this is similarly modified to fit better into the Participatory heuristic framework and simplified for a base user.

Affective Interaction Design, or simply Affective Design, is a newer, modern point of view inside the field of Human Computer Interaction toward system design which combines psychology and emotion based design, i.e. affective design, with User Experience Design and user interfaces (Preece et al. 2002: 141). Affective design is usually described to be a part of the visual sphere of systems design, with a focus on graphical design, but it also contains the so called "Emotional Design" aspect of systems design. This emotional design part includes user responses, motivation, behavior and other user centered aspects that often need considering (Cooper et al.

2007: 89-92). Affective Design specifically focuses on generating emotional responses with a system and in the case of this thesis the focus is on studying undesired emotional responses and what their effect is on the system as a whole. Finding and evaluating specific, detailed emotional responses is easily added into the heuristics evaluation process, however, it should be noted that the results can be subjective in nature.

1.3 Structure of the Thesis

The first chapter functions as an introduction into the primary subjects, methods and goals of this thesis. The second chapter focuses on the theory concerning User Experience Design and User Interfaces, along with examples and design aspects, including Usability and Affective Design. The third chapter discusses Usability testing in general and the testing method used in this thesis, namely the Participatory Heuristics method. The fourth chapter introduces the subject material and explains the necessary background information of the material, Wilma, including general information on School information Systems. The fifth chapter presents the research method used in this thesis, a qualitative approach based case study with an interview. The exact testing procedure used in this thesis, along with background to the testing and test cases, is introduced in chapter six. The seventh chapter presents the results of the heuristic evaluation part of the testing itself, including different issues that were discovered. The results of the questionnaire and the results of the affective, emotional aspects are presented in chapter eight. The last, ninth chapter, contains the discussion, implications and conclusions of the research, along with further study possibilities and evaluation.

1.4 Wilma

The material for this thesis is a School Information System called Wilma, created by a Finnish IT company Starsoft and currently developed by Visma (Haapsaari 2018). More specifically, the research focus of this thesis is on the communications module of Wilma. Wilma is a modern School Information System software package that is

designed to provide electronic administrative capabilities for a school and includes significant databases for student information, communication capabilities, administrative tools, report and analysis tools, logistics tools and much more than is to be expected to be included in a School Information System.

Wilma was chosen for this thesis due to the software's history, as it has been developed two companies and the original design team had a significant deficiency: it lacked all coordinated and planned user interface and user experience design. This means that Wilma has some unconventional user interface solutions but also that internal cohesion of the user interface, not to mention the overall user experience, has been sometimes questionable (Haapsaari 2018). The current developer, Visma, has done significant improvements on both user interface and user experience design, with training and planned user experience design. However, some aspects of the original Wilma are still present, especially on the school administrative side, and feedback has sometimes been emotionally colored, although on a system and user base of this magnitude this is to be expected.

2. USER INTERFACE DESIGN AND USER EXPERIENCE DESIGN

User Interface Design (UID henceforth) and User Experience Design (UXD henceforth) are two base subjects that cover the interface layer between a human and a computer, also Human Computer Interaction, although their approach and execution of the subject differs greatly. The simplest way to describe the difference between UID and UXD is that UID concentrates more on the technical, executive side of a user interface with layout design, graphical design and functionality design being key concerns, while UXD focuses more on the user's point of view, considering needs of different users, also known as personas, how easy the interface is to use, usability, and how the functionality conveys to users (MacKenzie 2014). In a sense, they are two different sides of the same coin and thus they overlap and enforce each other in their approaches. This does however make the subject slightly convoluting and often UID and UXD are considered to be the same subject: user interface design.

UI and UX have both been studied extensively and their studies have been closely tied with the development of computer systems. Of these two, the user interface is older as an established subject and the development of Graphical User Interfaces, GUI's, in particular is considered to be the source of both UI and UX design (Akass 2001). User interface design and studies have historically concentrated mainly on the functionality, effectiveness and usability (ISO 9241: 2016) or, in other words, the functional and technical aspects. User experience studies, on the other hand, are much more recent and have become a subject for study first in the 1990's. The term was originally coined by Don Norman, after the advent of internet, when GUI's were becoming more commonplace and more user centric approach was needed to develop interaction between humans and computers (Merholz 2007). User experience design concentrates much more on the user of a system with interaction, visual design and usability in the center (Psomas 2007). Usability is a notable aspect in both UI and UX spheres of interest and they overlap significantly, but their point of view is different: UI usability is more technical, with a central focus on "does it work well" while UX usability is more user centric, with "is it pleasing to use" in the center.

2.1 User Experience and Product Design

User experience design in itself is not limited to simply digital systems, but can be found in all shapes and sizes: if a product creates a meaningful experience for the user, it is experience design (Interaction Design Foundation 2018). Consider a coffee brewer; a basic coffee brewer found in most homes is a simple appliance with limited, exact functionality and purpose. It has a pot for the coffee, filter holder for a filter and a water tank combined with a water heater. In its most basic form a coffee brewer is controlled by a single button: on or off. From a UI perspective this coffee brewer could be of different color, different shape and the single button could be of different forms. From a UX point of view the taste of the coffee could be made more varied or the physical design itself could be changed to suit different needs, much depending on what type of persona lies behind the target audience. The entire coffee experience could be changed altogether and it qualifies as UX design while UI design would focus more on how the interaction works between the user and machine.

For a more intricate example, let us consider a coffee brewer with added functionalities, such as a coffee bean grinder and a timer to start the brewer at a desired time. First of all, the added features require a much more elaborate user interface to accommodate the new features: a bean grinder is present so it needs to be added to the design, while the user interface has to have a clear option to start the grinder as well as clear settings for the grinder. The timer also needs an interface, again with controls, clear time display and a method to set the alarm. Thus with only two, relatively small, added features, the original on/off user interface has evolved into: on/off; on/off with beans; and on/off with a timer, with both of the new features requiring some form of setup and setting controls..

In Picture 3 (below) we have an example of an interface found on Electrolux (2018) EKAM200 which is a coffee maker with the option of grinding coffee and makeing

timed brews. With the addition of two new features, an interface of one on/off –button has turned into an interface that has nine buttons and a digital screen. The complexity of the user interface has increased, with a very rough estimate, by 1000%, while the features increased only with 200%. However, considering the different added features on the interface, the estimate is very conservative. In addition to this, what the picture does not show is that the coffee grinder is activated with an extra latch above the filter holder and is controlled by opening the hatch.

Picture 1. User interface of a coffee brewer with a coffee grinder and a timer (Electrolux EKAM200) (Picture by Oskar Kenttälä 2019)

The point of this example is not only to express the correlation between the complexity of a system in relation to the complexity of the user interface, but also to point out the importance of a user interface: it is the control center for the entire system and often the first single feature the user sees and familiarizes when starting to use the product. It is

also the single most important feature on any single product, being the most used feature on the product as it controls all the other features. Ben Shneiderman (2005: 4-11) describes user interfaces as being one of the most far reaching modern phenomenon in society with the ability to save lives or take lives. This is historically also true, as for instance, in the infamous Therac-25 accident (Computer 1993) two people died and four more were permanently injured due to an issue that eventually can be traced to poor user interface design, lax programming practices and the complete failure to understand the user's actions or requirements with the system.

While most systems are not quite as demanding and potentially lethal as an X-ray machine, a poorly designed and implemented user interface can frustrate a user and be severely detrimental to the user experience. This frustration does, however, eventually reflect poorly on the system itself and the user often is inclined to stop using the system or simply replace it with an easier to use alternative (Preece et al. 2002: 148). Common negative system features include confusing use of graphics or text; lack of feedback of your own actions; insufficient design, including poor layout, color options and uncommon key wording; and generally infuriating aspects that do not provide anything for the system (Preece & al 2002: 3). Poorly designed and implemented user interfaces make even simple systems difficult to use, but above all else, it frustrates the user and diverts attention from using the system to working the user interface. It should be mentioned, that some systems deliberately sacrifice user experience values for a more efficient outcome. Especially different Information Management Systems are notoriously frustrating to use, but their benefits in managing a company are so significant that it is much preferable to modify, that is teach, the user to operate in the system than to modify the system to accommodate the user (Topi, Lucas & Babaian 2005). This detail thus presents the main contradiction between UI and UX design: the necessity of the systems functionality versus the user's preferences and requirements.

Every system, be that a product, software, machine or otherwise has certain features and functions. The central idea of user interfaces, from both UI and UX perspective, is to have these features available in a simple, concise and easy to use form. This was established in the first paragraph of this thesis and is generally established as a central

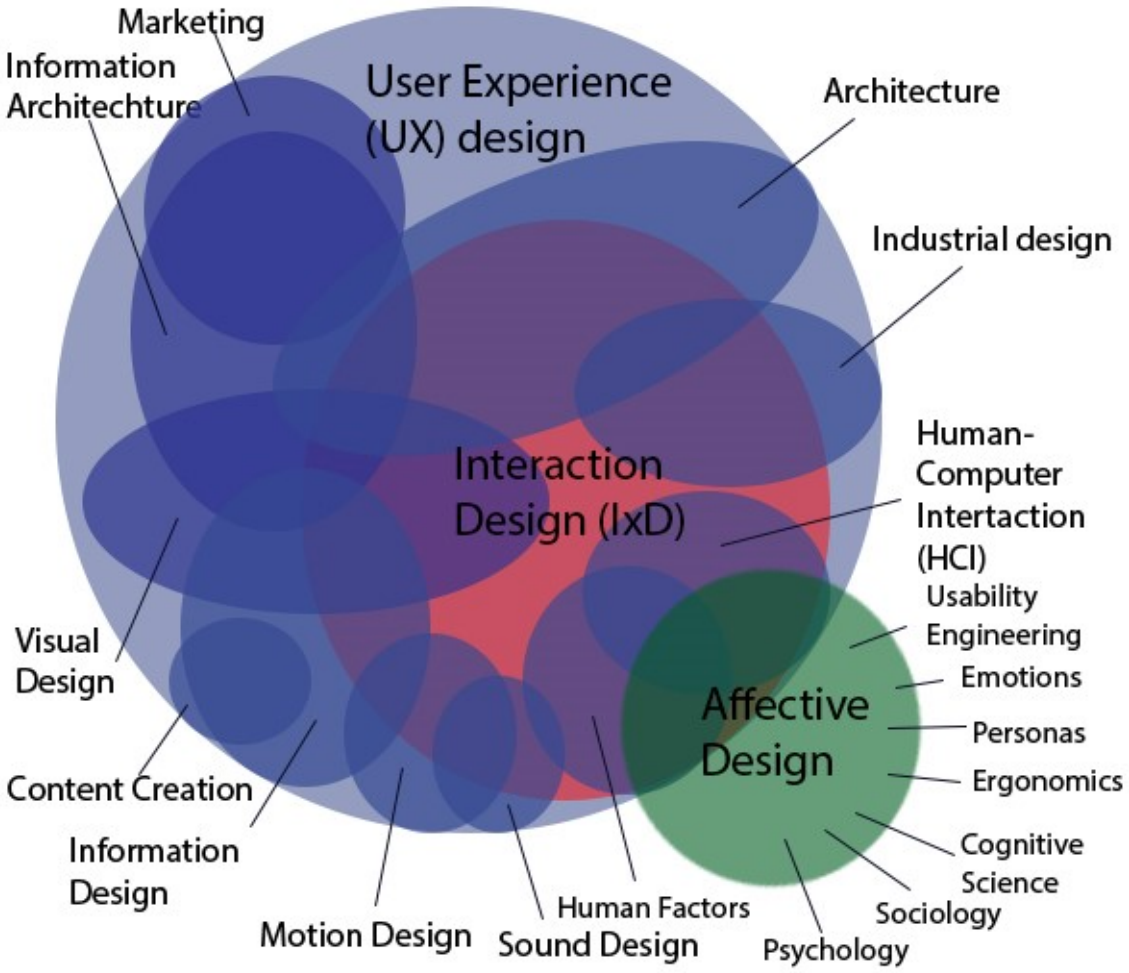
starting point for UX and UI design (Shneiderman 2005, Preece et al. 2002, Nielsen 1993). However, in several cases systems have dozens of features, requirements and interdependencies that require the user to have certain knowledge of the functionality and use of the system for safe operation. This brings about a situation where the minimum requirements for a system from a user interface design point of view, creates such a complicated set of controls or graphical user interface that the user experience suffers simply from the amount of features and controls.

Windows operating systems (Microsoft 1985), for instance, have a multitude of different features that require certain knowledge of the system to operate properly, but the requirements of the technology simply require certain, possibly inherently negative, features to be present as well. One such feature is error messages of all kinds, which incidentally, have been present in all version of Windows operating system. While critical information boxes, or pop-ups, are important for the correct working of complicated systems, they stop all activity, require attention and action from the user to be removed and are often meaningless for the user (Preece et al. 2002: 148-150). To prevent damage to the system itself, these error messages simply need to exist and this is an example area where UID and UXD severely clash with one another, while on the other hand the system needs to convey critical information to the user, but at the same time this critical information severely interferes with the user's experience of the system. However, User Experience Design has specifically been designed to balance out these requirements and Windows today is a much more pleasant experience altogether as much of the design focus has been diverted from simple functionality to accommodate the user better into the system. In a sense, the functionalities that are created by User Interface Design need to be further designed through the scope of User Experience Design for the best possible results.

2.2 Interaction Design

Interaction design, as most of computer science related subjects usually are, is a relatively new subject field that was originally coined in the 1980's. As a practical subject, interaction design is much older. In the center of the field is behavioral study of product use and this has been as long as there has been any type of product development (Friesen 2013). However, as a distinct subject field within information sciences, interaction design was recognized first in the 1990's and it has gained momentum first in the last decade as digital media products have become more common, along with the need to make proper design for these products. With roots in industrial design, interaction design shares many aspects with its parent with form and practicality being important aspects. Interaction design focuses instead on user behavior and through this creates interactive products, systems and services (Cooper et al. 2007). Interaction design is not limited to information systems or products but works equally well on physical products by studying how a user uses a product and then improving the design to fit better for the user and the user's requirements.

While the three main subjects of user experience design are form, content and behavior. Interaction design similarly has these three aspects, but has its main focus on behavior and especially on how form and content affect behavior, but similarly, how behavior can be used to affect form and content (Cooper et al. 2007: xxxii). In IxD, these three main aspects have further been divided into five specific subjects, or as called by Moggridge (2006) Five dimensions. These Five dimensions are: Words, Visual, Physical Space, Time and Behavior (Moggridge 2006, Silver 2007). The first three dimensions, Words, Visual and Physical space are considered to form the basis of any interaction: Words are what is said or done, Visual represents any and all visual cues and Physical Space represents either the object that is used or the space where something is interacted. The fourth and fifth dimension affect how the interaction is perceived: Time is a factor in any interaction as it defines the length or even when the interaction is available, while Behavior is represented by both how the interaction affects behavior but also what is the user's behavior before the interaction.



Picture 2. Interaction Design and Affective Design relations inside User Experience Design. (Saffer 2008) (Picture by Oskar Kenttälä 2019, adapted from Saffer 2008)

The greater User Experience Design field is constructed from several varying subjects and disciplines. Picture 2 above visualizes the entire UX field, including Information Architecture, Visual Design, Information Design, Motion Design and Interaction Design (henceforth IxD) being the most significant disciplines inside the field. Affective Design (henceforth AD) is an addition to the original picture describing the Disciplines of User Experience, by Dan Saffer (2008), as Affective design has traditionally been seen as being a set of several smaller disciplines, including Cognitive Science, Psychology and Sociology. However, they are in principle, aspects of Affective Design and this adapted picture links Affective Design, and the most important areas inside its sphere with the overall UXD field (Lockner & Bonardel 2014, Norman 2004, Saffer 2008).

Outside studying and evaluating interaction and experience, interaction design has one idea that needs to be specifically mentioned: the idea of the user, the Persona (Cooper et al. 2007:75-109, Preece et al. 2002: 73-105, Goodwin 2009: 113-299). While this at a glance sounds simple and self-explanatory, the idea of the Persona, the potential customer, user of the system, is a difficult and complex subject that easily is built and estimated erroneously. The idea of a Persona has basically two different points of view, both of which are equally important. First, the Persona represents a data gathering aspect, where all of the different features the potential customer has, that is goals, desires, expertise and, most importantly, behavior patterns, are used to make estimations of the system requirements and features (Cooper et al. 2007:76-77). The second part of Personas is the model users, specific customers that have more specified attributes which can be used to evaluate customer segment effects. For instance, one Persona could be a 7-year-old child and another the 35-year-old parent for the child. These two Personas are significantly diverging in their desires, expertise and behavioral patterns. When constructed properly, personas give relatively solid estimations of what to expect from their respective customer segments, although it should be mentioned, that the Persona, while a useful tool for making estimations, is rarely a perfect amalgamation of a real world user (Salminen, Kwak, An, Jung & Jansen 2018). As a contradiction against the usefulness of Personas, real users often use and need the system in unexpected ways. Thus the Personas, as well as the system, should be updated to accommodate this newfound information when conducting interaction design for maximum effect.

As IxD is considered to be a form of goal oriented design, with satisfying the user's needs being the goal, quality and understanding quality from the behavioral point of view is integral to understand and measure any type of results. Quality, however, in a field of study that is based mostly on personal experiences is difficult and contains a multitude of different subjects. Silver (2007) mentions that flow, responsiveness, context and appropriateness are the major qualities in IxD, while Löwgren & Stolterman (2004) recognizes several more in their research, including pliability, seductivity, playability, transparency, social action space, personal connectedness and parafunctionality to mention a few. This indicates that quality is not simple in

interaction design and they are strongly tied with usability studies and affective design, which both are a significant sphere inside Interaction design.

2.3 Usability

Usability is one of the more prominent modern fields of study inside the UX and IxD fields. As most of UI, UX and IxD, Usability is heavily linked to software development both historically and practically, but as a general subject it is far older. For example, a hammer with a leather grip could be considered rudimentary usability engineering as it improves and modifies a simple tool to better function for a human. Usability as a field of study was originally established in the 1980's with first usability tests being used to test computer usage, but it would be first in 1993 that it was clearly defined as a field of study when Jakob Nielsen published the book *Usability Engineering* (Sauro 2013). *Usability Engineering* is even today considered a base for established usability studies and it does introduce many of the key principles of usability that include the five defining aspects of Usability: Learnability, Efficiency of Use, Memorability, Few and Noncatastrophic Errors and Subjective Satisfaction.

Originally, Nielsen devised Usability to act as way to study and improve user interfaces between computers and human users, but it has been used to improve many types of human interaction including industrial design (March 1994), communication (Gaffney 2004) and even translations (Suojanen, Koskinen, Tuominen 2012). The reason for this is two-fold: first, Usability provides a specific user centric point of view which argues that products should be designed to be used as easily and efficiently as possible for the best results (Nielsen 2012) and second, Usability provides a large selection of different tools and methods to evaluate functionality (Nielsen 1993, Sauro 2013). While these tools have been designed specifically for software and user interfaces, they are often simple, malleable and easy to use themselves which in turn means that they are easy to modify for more specific uses and the user-centric perspective carries over to the new field.

One of the first tools to be specifically developed to study and evaluate software usability from the user's point of view is the Thinking aloud method developed in the

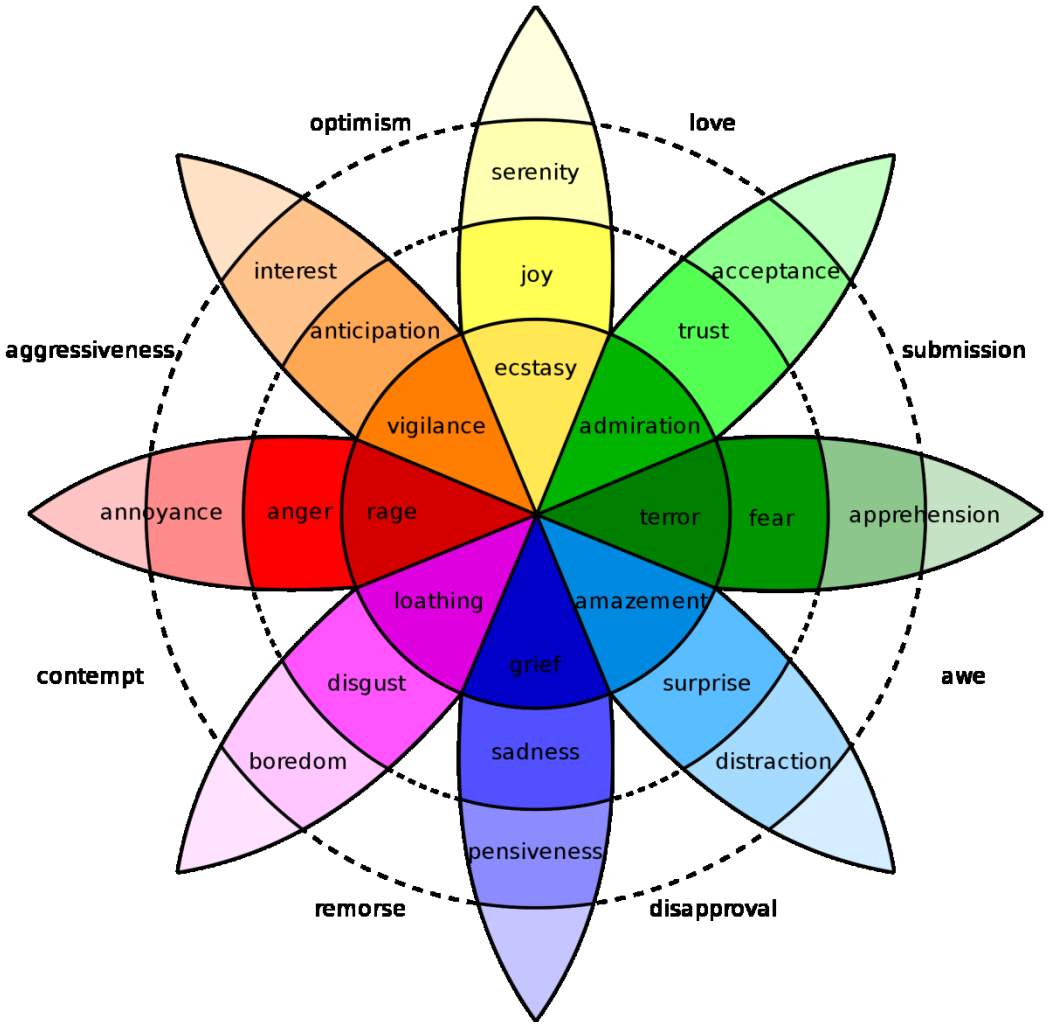
1980's (Eriksson & Simon1980, Lewis 1982). The basic idea of Thinking aloud is simply to ask the tester to speak aloud what he is doing and thinking while doing a test scenario and through this to make evaluations of the software. While not overall the first, considering that this type of testing was done in different occasions already at the beginning of the century, it was the first evaluation method, that had credible scientific leverage and it is commonly used even today, often in combination with other more complex tools (Sauro 2013). Other prominent testing methods include Nielsen's Heuristic Evaluation (1993), which is the select method in this thesis as well, more in Chapter 3, Cognitive Walkthrough (Sears 1998) and RITE (Medlock, Wixon, McGee & Welsh 2005) to name a few. These tools differ mainly in their central point of view, complexity and above all else, cost, both in money and time. As an example, Thinking aloud method is considered to be a low-cost alternative, only requiring a computer and a test person and the results are fairly good amalgams of the users experience. The problem with the Thinking aloud is that the test case dictates much of the results and thus their value is limited to the test case alone. Heuristic evaluation on the other hand is notably more complex and designed for professionals, but it is otherwise low-cost and it efficiently discovers the worst usability issues (Nielsen 2012).

While Usability in itself is very process derived and straight forward, it ties in with Interaction design as a form of quality, more exactly, functional quality. With a heavy focus on functionality and the ease-of-use of a system, Usability provides answers to basic problems and improves overall quality, interestingly both from a UI design point of view, but also from a UX point of view. The multitude of different tools enables designers to make evaluations of the system from varying stand points, simply by making slightly differing test cases. Usability often provide clear results in specific areas and questions, but offers less exact answers to general aspects and rarely gives any answers to questions like "How can it be repaired?" Usability only point out problems, not answers. Usability, and the tools it provides, is still a powerful method to improve any system and should not be ignored as such; it should be used as an additional tool to develop systems. From a User Experience point of view, Usability does not, however, provide the entire answer to quality as feelings and emotions are difficult to evaluate and assess. The emotional side of quality is better addressed by Affective Design.

2.4 Affective design

Affective design, or emotion affecting design, is one of the newer branches of research to be added to the IxD and UX spectrum of research. Originally a subject concerning mainly physical design and art, affective design has been included as an aspect of research in Interaction design first in the last few decades (Cooper et al. 2007: 76-108). However, affective design has been a major part in User Experience design since it was established as a field of research in the form of psychology, cognitive science and ergonomics, all of which can be seen in Picture 2 on page 18. In practicality, Affective design fills the other part of quality in the UX field, while Usability fills the other. Usability, in general, is concerned with functional quality with a focus on aspects such as concision, learnability, feedback and availability. Affective design focuses more on the aesthetic side of quality with user responses, motivation and behavioral patterns as the focus (Preece et al. 2002: 141-153).

Affective design focuses mainly on two distinct subjects: the visual sphere and the emotional sphere, although they are not in any way separate from one another but instead they overlap heavily. For instance, simple visual cues have long been known to cause emotional responses: a smiling icon creates a response of friendliness and thus a feeling of success (Preece et al 2002: 143). Thus visual aspects create emotional responses and, similarly, functional aspects create emotional responses, even without a visual component. The relationship between Affective design and Usability in creating quality is notably intertwined as research dictates that successful visual design reflects positively on Usability while Usability affects the emotional responses of a system (Tractinsky1997). While the main idea of Affective design is to create an effect on emotions and behavior, emotions, however, are very complicated and subjective. Due to this, it is never certain what emotions specific design might create in some user. For example, poor layout design such as unnecessary complexity, software errors and poor feedback most often creates negative emotions, mainly frustration and anger, but it could similarly create emotions of amusement (Preece et al. 2002: 147). The reason for this could be related partly to subjective experiences, situation or simply the user’s state of mind, but also to the vast variety of emotions humans have.



Picture 3. Plutchik's (2002) Wheel of Emotions. (Public Domain - Image from Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Plutchik-wheel.svg –)

Plutchik's (2002) Wheel of Emotions, in Picture 3 above, depicts the spectrum of human emotions and it recognizes eight main emotions and several secondary and tertiary emotions. The wheel is a three dimensional cone structure, with the primary emotions more separate and clear at the top, secondary emotions in the middle and tertiary emotions at the bottom of the cone and the further down the cone we go, the difference between the emotions is obscured. For instance, Rage and Vigilance are clearly different emotions, although they share certain similarities, but Love, Optimism and Submission begin to be notably less clear to discern. As a result, emotions are not only difficult to categorize and study, they are equally difficult to try to affect, although certain patterns produce somewhat expectable results (Norman 2004: 29-30). For instance, smiling faces create positive feelings while darkness is often connected to negative emotions, although these are subjective.

Don Norman (2004) is one of the premiere researchers in Affective design and he introduces the idea that products could affect the user's cognition and emotion through design and form. While he does not specifically discuss software or information systems with Affective design, he makes the connection that users respond emotionally towards products and brands, which are present in one form or another in most modern information systems (Cooper et al. 2007: 89). Norman (2004) introduces the idea that product design, system or physical, should address three levels of emotional and cognitive processing: Visceral, Behavioral and Reflective. He describes these levels as Visceral being a reaction toward something; shock, awe, infatuation and so forth, while Behavioral works on intermediate level of thought and emotion with minor adjustments to everyday behavior and finally, Reflective which is the slowest type of emotion and thought processing where you consciously consider something (Norman 2004, Cooper et al 2007). Out of these processes, the majority is Behavioral and design targeted towards these emotions and thoughts have the best chances of affecting the user, although Visceral and Reflective patterns can be designed as well, they are simply more difficult to create and the results are more unpredictable.

The marketing industry has, in fact been, using this emotional and behavioral affecting design for decades and several different examples can be found. For instance, United Colors of Benetton is known for creating shock advertising campaigns, an example can be found in Appendix 1, which is solely based upon a Visceral, shock effect from a controversial picture. The effect of this design is based upon creating discussion and potential feedback from a very negative experience to the viewer and in the marketing industry all publicity is considered "positive" publicity. An example from a Reflective design can be found from the Finnish YLE (Mickwitz 1986) and their infomercial warning about weak ice. This short commercial was originally shown to children as a warning to stay off weak ice and it combines dark music and cartoony characters in a very serious setting. The commercial very powerfully creates Reflective thoughts and emotions about ice by combining music, distressful emotions and likable characters. So effectively in fact, that some children have been reported to be traumatized by this

commercial (Kuosmanen 2018) which would also turn it into a form of visceral design, albeit perhaps unwittingly.

Behavioral level of cognitive processing is the middle level of human cognitive and emotional processing and also the most common form. Behavioral level is best described as being the unconscious, learned level of human action and thought, such as riding a bike, that does not necessary require active thinking (Norman 2004: 21-24). It is as such not quite at the level of a reaction to something but not either actual thinking and thus it contains a large portion of human action. In design, this is the most familiar area of the human cognition as it directly affects things such as what the user perceives as pleasure. Marketing has also in this area an endless amount of examples as the main idea of marketing is to connect positive thoughts to brands and this is often used with smiles, ideal settings and other positive connotations. A beautiful woman is considered the preferred subject in most instances as beauty nearly always brings good emotions. Interestingly, Behavioral patterns also enhance and inhibit lower Visceral and higher Reflective emotions, while at the same time they can affect Behavioral patterns in return (Cooper et al 2007). For instance, very fond memories of some product can leave a permanent positive feeling towards it while for instance getting frightened in relation to a product can leave a negative feeling towards it.

Emotions are eventually a very challenging subject in User Experience and Interaction Design. While some certain connections and patterns can be made, the inherent problem with emotions and cognition is their subjective nature. Certain aspects simply affect people differently and gaining a perfect result is always impossible. Certain aspects in UI/UX design, however, have been identified to consistently cause negative emotions, especially frustration, in users including crashes; insufficient process information and feedback; and inconsistent or unhelpful visual design (Preece, Rogers & Sharp 2002: 147). Also, the use of pop-up windows, especially in error situations, should always be carefully considered in all situations as they not only break the flow of the work, they easily aggravate the user. Similarly, certain positive aspects such as clear design, pleasant pictures and proper feedback bring positive thoughts in experience, but they are often invisible to the user if they are done correctly. More elaborate affective aspects

such as style and specific design is very difficult to convey correctly with desired outcomes, certain patterns can still be detected and established.

Overall, affective design is a powerful tool to consider when creating any type of design and this is no different in Interaction design. The challenge, however, comes in the form execution: how to create this emotion affecting design and what is expected to follow? This dilemma is further enforced in user interfaces which are not exactly the target of the user's attention, simply a means to utilize the software itself, and thus they suffer from a phenomenon that could be called the Invisibility of Positive Features: a user interface is usually completely invisible to the user until the moment it does not function correctly or the flow of operating the system is somehow interrupted. Therefore, most often when doing UI and UX design, the resulted emotions are often negative and developing the design is mostly removing negative features. This is not inherently counterproductive as these negative emotions very effectively pinpoint flaws in the system that can then be corrected, but unfortunately the task is otherwise for the designer thankless and often emotionally taxing. Especially as aggravated users are prone to sending their negative emotions to the designer in very colorful feedback.

3. HEURISTIC EVALUATION AND AFFECTIVE DESIGN

In this chapter we will discuss the Heuristic Evaluation including Nielsen's original Heuristic list, Participatory heuristic list and the modified list used in the thesis.

3.1 Heuristic Evaluation

Jakob Nielsen and Rolf Molich (1990) originally created the heuristic evaluation method as a simple, easy to approach tool to evaluate the quality and usability of computer software. Nielsen (1993, 1994, 2012) has since refined and improved the method on several occasions and currently it is considered to be one of the basic methods of evaluating software and user interfaces. Nielsen originally designed heuristic evaluation as an expert evaluation method to be used as an iterative system development tool where the evaluation would be done several times during the development cycle, especially after more significant changes. Heuristic evaluation is a fast, simple and cost effective method which usually finds the worst usability problems effortlessly (Korvenranta 2005: 115). Usually the evaluation is done by a group of three to five usability professionals as studies have shown that a single evaluator usually finds 35% of usability problems in a single round of evaluations and often the most notable issues are a part of this group (Kuutti 2005:47).

While heuristic evaluation does have notable benefits, it is not perfect and the most notable shortcoming is the complete lack of end-user feedback. Also, while the most notable issues are found easily, more complex issues or the outlying issue might stay hidden. For instance, a dark background picture on a website which hides black text is quickly flagged as a serious usability issue due to the difficulty of reading the text, but it does not necessarily point out that the website has a layout or theme problem. Finally, like most other usability testing tools, the method offers absolutely no solutions on how to correct the issue, simply, that it exists and thus, like in the example above, the real reason behind the problem might stay hidden.

Heuristic evaluation is done based on a set of rules or principles, often called heuristics, which are used to evaluate the target material. The list of heuristics is used much like a search algorithm, where the evaluator attempts to locate aspects in the material that do not comply with the rules (Nielsen & Molich 1990). Nielsen's list (1995) of heuristics is often used as a base for this list of rules, as it effectively covers most basic usability aspects present in a system. However, these rules can also be created specifically for a task, although the list should be kept to approximately ten rules and the list should be designed with care, as a poorly thought out list confuses the evaluator and yields bad results. Nielsen's (1994) list is derived from the principles of Usability: Learnability; Efficiency of Use; Memorability; Few and Noncatastrophic Errors; and Subjective Satisfaction and can be seen in Table 1 below (Nielsen 1995):

Table 1. Nielsen's Ten Heuristics (Nielsen 1995)

1. Visibility of system status
The system should give feedback to the user and functions should be transparent.
2. Match between system and the real world
The system should be understandable by the user with correct language and familiar phrases: technical jargon and unfamiliar concepts should be avoided. Also, the form should be logical and have a natural form.
3. User control and freedom
Redo and undo. The system should be designed for easy navigation and anticipate user errors in different stages, including going backwards or to exit completely.
4. Consistency and standards
The system should not substitute reality with its own: follow established rules of usage, actions, icons and words. If the system does introduce unconventional features or details (terminology, abbreviations etc.) use them consistently.
5. Error prevention
Instead of error-messages, design error prevention: automatic returns, advancement blocks, error removal and confirmation options all decrease the use of error messages. Less is more.
6. Recognition rather than recall
Actions and objects in the system should be recognizable when seen, not remembered. Instructions should be easy to find and access when so desired.
7. Flexibility and efficiency of use
Expert user tools like key combinations such as ctrl-c and ctrl-v should be supported. Design should in general be as simple as possible and have options for different user profiles.
8. Aesthetic and minimalist design

Following somewhat the earlier point, useless information should be avoided and the function of the system should be the first priority. All extra information reduces the value of the most important information.
9. Helping users recognize, diagnose, and recover from errors
Error messages should be as explanatory, simple as possible and preferably suggest a solution. They should not contain unintelligible codes that might tell much to a programmer but nothing to the user.
10. Help and documentation
While the optimal situation would be a system without the need for documentation, if it is needed it must be easily found and accessed, with a helpful form and search options. The help documentation is a system of its own and should follow all of the principles above.

After the evaluator has discovered an issue, the severity of the problem is classified in a 0 – 4 scale (Nielsen 1995). Three different factors contribute to the categorization: frequency, impact and persistency. Of these frequency and impact are somewhat self-explanatory as a single occurrence is not a significant issue, but ten times is and if a problem occurs only in certain specific conditions it is negligible while an erroneously triggering, consistent pop-up is a significant problem. Persistency is a combination of these: if the problem is easily ignored, even if it triggers often, it is of low persistency while a problem that triggers rarely, perhaps in certain conditions, but the user needs to for instance restart the system, the persistency is notable. Below the five different levels of severity (Nielsen 1995):

0 = I don't agree that this is a usability problem at all

1 = Cosmetic problem only: need not be fixed unless extra time is available on project

2 = Minor usability problem: fixing this should be given low priority

3 = Major usability problem: important to fix, so should be given high priority

4 = Usability catastrophe: imperative to fix this before product can be released

These five levels are somewhat explanatory, except the zero level and the fourth level. The zero level of the evaluation list points to a situation where a problem concerning the system has been discovered, but it is either so small that it simply does not warrant more

than a mention or is simply not a usability issue at all, simply a difference in opinion on how something should work. The last level of the list on the other hand refers to exactly the opposite with a problem that is so severe that it must be corrected for the system to be released. Such examples are usually rare and extreme, but they are present. Also, the definition of what is so severe that it must be repaired is situation, developer and evaluator dependent. Evaluator dependency is also one of the problems with the categorization as Nielsen (1995) states that due to the difficulty of assessing the severity of a problem, single evaluator's results should not be unanimously trusted but should be compared to other similar discoveries.

While this is relatively simple in itself, the evaluators own perception and "feeling" of the problems severity has notable impact. Novice users can consider a problem of high value while an expert user might consider it insignificant. Thus the persona of the evaluator is significant when doing an evaluation and due to this Nielsen (1994, 1995) considered expert users best to perform the evaluation for consistent results. However, confining the evaluators to only experts also cuts the end-user from the development, and the experts often have a wildly differing idea of the expectations and requirements the end-users have from the product.

3.2 Participatory heuristics

Participatory heuristics is at the same time a modification and a step away from Nielsen's Heuristic evaluation. While the method is in function and execution the same as in Heuristic evaluation, with an evaluator looking for details based on a list of heuristics and the rating the severity, the difference can be found in that the evaluation is not done by experts but by end-users of the system (Muller, Matheson, Page & Gallup 1998). This rather significant abbreviation from the original also requires that the list of heuristics to be modified for the purpose, with simpler terminology and easier-to-approach concepts. Otherwise, the Participatory list of heuristics is heavily based upon Nielsen's list. Table 2 below goes in detail on the list (Muller & al. 1998):

Table 2. Participatory heuristics (Muller, Matheson, Page & Gallup 1998).

Section 1: System status
1. System status
The system should give feedback to the user and functions should be transparent.
Section 2: User Control and Freedom
2. Task sequence
Users should be able to sequence their tasks in the order they want instead of the system dictating it. Have wizards as an option, not a requirement.
3. Emergency exits
Navigation backward and forward, exit from the system entirely, ability to repair mistakes. Avoid unnecessary dialogues, keep as simple and quick as possible.
4. Flexibility and efficiency of Use
Expert user tools like key combinations, should be supported. Options for different user profiles. Alternative access methods.
Section 3: Consistency and Relevancy
5. Match between system and real world
The system speaks in the users language, has logical order in messages and matches with the system in the real world.
6. Consistency and Standards
The system should follow established rules of usage, actions, icons and words. Logical order in messages. If the system does introduce unconventional features or details (terminology, abbreviations etc.) use them consistently.
7. Recognition rather than recall
Actions and objects in the system should be recognizable when seen, not remembered. Instructions should be easy to find and access when so desired. All objects should be clear and visible.
8. Aesthetic and minimalist design
Useless information should be avoided as all extra information reduces the value of the most important information.
9. Help and documentation
While the optimal situation would be a system without the need for documentation, if it is needed it must be easily found and accessed, with a helpful form and search options. The help documentation is a system of its own and should follow all of the principles above.
Section 4: Error Recognition and Recovery
10. Help users recognize, diagnose and recover from errors
Error messages should be as explanatory, simple as possible and preferably suggest a solution. They should not contain unintelligible codes that might tell much to a programmer but nothing to the user.
11. Error Prevention
Instead of error-messages, design error prevention: automatic returns, advancement blocks, error removal and confirmation options all decrease the use of error messages. Less is more.

Section 5: Task and Work Support
12. Skills
The system supports, adds or enhances the user's skills and knowledge. The system is there to help the user, not to replace.
13. Pleasurable and respectful interacting with the user
The system should reflect the user's professional role, identity and intention. The system should be functional but also aesthetically pleasing.
14. Quality work
The system should support the users work quality with timeliness, accuracy and completeness.
15. Privacy
The system should guard the user's, and clients, information.

As stated earlier, the list is notably similar to Nielsen's list, with only a few modifications on categorizing the subjects more efficiently, first into five sections and then moving some subjects, such as Error prevention and, finally, dividing some subjects. Only the last section, Task and Work Support, deviates significantly from Nielsen's list and is in fact a completely new, user centric addition to the heuristics. The list is somewhat long, which perhaps makes it slightly difficult for an untrained user to use, but otherwise it clarifies and simplifies many of the subjects in Nielsen's list.

Participatory heuristics as a method of studying system usability through users, instead of experts, has proven to be successful method (Schaarupa, Pape-Haugaarda, Hangaarda, Mihovskab, Hartvigsenc & Hejlesena 2015), however, it also adds more variables in the form of the persona of the evaluator. When expert users make the evaluation, they are usually of two very distinct personas, namely software engineers and usability experts, and they have similar technical expertise, understanding of usability and have the same view of the desired function of the system (Muller & al. 1998). When shifting the evaluation away from experts on end-users, all of this has much more variance and thus can have an effect to the results. Granted, with a larger set of backgrounds more diverse opinions and problems can be expected to be found, but their importance should be more carefully considered. Similarly, the persona of the test subjects should also be chosen carefully, so some expected variance could be eliminated.

3.3 Heuristic list for Affective design

The heuristic evaluation list for this thesis needed to be specifically created for the task as the target of the research is to consider usability factors through an emotional point of view. For this end, non-professional users were tasked with finding details, aspects and features of the target system that they either considered annoying or otherwise thought did not work. The list is not completely about emotions and the basic aspects of Usability as presented by Nielsen (1994) are present as well: Learnability; Efficiency of Use; Memorability; Few and Noncatastrophic Errors; and Subjective Satisfaction. However, Nielsen's list of heuristics does not exactly function with the task, as it is designed for professionals, and the Participatory heuristics based list (Muller et al. 1998), while more developed for an end-user, is still somewhat complicated and has aspects that do not add anything to this study. Thus, a customized list needed to be made with simplicity of use and affective design related aspects in mind. The finished list can be found in Table 3 below.

Table 3. Modified list of heuristics for this thesis.

1. Aesthetic and minimalist design
Presented information is relevant, there is sufficient amounts of it and it furthers the systems functionality.
Objects of the system (pictures, colors, names etc.) are of appropriate size, style and utility is consistent and clear.
Information is displayed in a pleasant way that furthers the functionality of the system.
2. Efficiency of use and standards
The concepts of the system are familiar and correspond to established standards.
The system should give feedback to the user and functions should be transparent.
The system has no distracting, excessive or unnecessary objects.
3. User control and freedom
Using the system is simple, clear and motivating. Navigation is clear and logical.
General computer tools, such as key combinations, are supported. Backwards movement and exits are available.
Using the system is pleasant, interesting and furthers the users goals.
4. System functionality
Functions progress logically, clearly and are not unnecessarily complicated.

Actions and objects in the system should be recognizable when seen, not remembered.
Systems functionalities do not frustrate or confuse.
5. Error prevention and Support
Faulty input can be repaired or modified.
Objects, functionalities and controls are error free and clear.
When required, help or support can be found easily, quickly and is useful.

The basis for this list is from both Nielsen's list (Nielsen 1995) and Participatory heuristics (Muller & al. 1998). First of all it is divided into five different subchapters much like in Participatory heuristics. This reduces the cognitive load for the evaluator and instead of trying to remember ten or more specific points the evaluator simply needs to remember the subchapter and what it roughly contained (Piernik 2017). This is important as the list is designed for non-professionals who are most likely not familiar with usability heuristics. The subchapters themselves are designed to represent a significant theme in the system: the first, Aesthetic and minimalist design, which is directly from both of Nielsen's list and Participatory heuristics, represents the general layout and aesthetic design of the system. Efficiency of use and standards is a combination of two of heuristics from previous lists, namely Flexibility and efficiency of Use and Consistency and standards, and focuses on general usability and standards. User control and freedom is directly from established heuristics and focuses on control and navigation. System functionality is similarly created from previous heuristic lists and concentrates on functionalities. Lastly, Error prevention and support combines the different error related subjects.

In total the heuristic evaluation list contains a total of 16 heuristics, which are more or less based on established heuristics found in either Nielsen's list or Participatory heuristics list or in both. Some parts of the list are slightly modified or thematically close to these lists, but are directed more towards simplicity of use, so a non-expert can understand the subject more efficiently.

Affective design aspects in the list are represented in three specific extra additions which deviate from the established lists:

- Information is displayed in a pleasant way that furthers the functionality of the system
- Using the system is pleasant, interesting and furthers the users goals
- Systems functionalities do not frustrate or confuse**.**

These three heuristics are located in the end of the 1st, 3rd and 4th subchapter and are designed to guide the evaluator to consider both emotionally positive aspects and negative aspects. These three heuristics are somewhat similar with each other, albeit they are pointing at different aspects of the system, but this was specifically designed so their mildly repetitive nature remains with the evaluator better. Overall the affective aspects are few when compared with the usability aspects in the heuristic list, however adding more would be overly repetitive and would interfere with the evaluation in general. Thus the heuristic evaluation is supplemented with a questionnaire that will focus more on emotional thoughts concerning the system and together with the evaluation a solid data set is created where conclusions regarding affective design in the material can be done.

As a final note, it should be mentioned, that due to the non-expert nature of the evaluation, the severity rating list from 0 – 5 has been modified for this evaluation process by removing the 0 level, only retaining the levels 1 through 4, meaning that the final severity list is:

1 = Cosmetic problem only: need not be fixed unless extra time is available on project
2 = Minor usability problem: fixing this should be given low priority
3 = Major usability problem: important to fix, so should be given high priority
4 = Usability catastrophe: imperative to fix this before product can be released

The reason for this is, that the 0 level gives unnecessary and distracting tasks to the non-expert evaluator as the zero level is basically designed for more in-depth expert evaluation, where the zero level represents an idea where an evaluator thinks something

is either wrong with a certain detail, but he perhaps cannot exactly point out what it is, or simply to point out that the expert evaluator thinks this should be done differently. This abbreviation is similarly encouraged to be done in Participatory heuristics which is similarly designed for non-experts (Schaarupa, Pape-Haugaarda, Hangaarda, Mihovskab, Hartvigsenc & Hejlesena 2015).

4. SCHOOL MANAGEMENT SYSTEMS

In this chapter we will shortly discuss the background of the material, namely Management Information Systems and their specialized form designed for schools, School Information Systems. While this is perhaps somewhat secondary information concerning this thesis as a whole, the background of the material is good to be familiar with as Wilma, the primary material, is a School Information System.

4.1 Management Information Systems

Student Information Systems (henceforth SIS), or School Management Systems are a specialized type of Management Information Systems that is specifically designed for the requirements of a school. Management Information Systems (MIS henceforth) are systems designed to assist in decision making of companies by controlling information, products, people and resources in the best possible way (Laudon & Laudon 2014: 32-61). The importance of MIS systems has been increasingly important in the modern world as the amount of available information is increasing, competition is harder and the need to reduce cost is constant. MIS systems provide companies the necessary tools to increase performance, reduce costs and gain more accurate information of the company, thus providing the possibilities to make faster and more accurate decisions. However, most MIS systems are complicated, difficult to implement into companies and without careful and proper use, create more problems than benefits, including performance issues and erroneous information (Singh & Twalo 2014). MIS systems are thus categorized into several subtypes to better differentiate them for their intended use in business. These categories include Decision Support Systems, Human Resource Management Systems and Enterprise Resource Management. Enterprise Resource Management systems in particular are considered to be paramount for modern companies that specialize in any kind of large scale production, as it enables information gathering and control to all levels of the corporation. School Information

Systems are similarly one of these categories and much like the rest of them, have distinct requirements and features that differentiate it from other MIS systems.

Historically MIS systems are a relatively new branch of science as they are directly tied to the development of computers (Laudon & Laudon 2014: 198). The first era MIS systems were originally from the 1960's, when mainframes controlled most data and slaved minicomputers worked as operators, most MIS systems were rudimentary and resource intensive and only the largest companies owned them. The second and third era belong to the advent of personal computers. Originally individual stations enabled powerful data management at a fraction of the cost of the old mainframes, but the need share information eventually developed into small client/server networks that enable true MIS software to take root in enterprises (Boykin 2017). The current eras, fourth and fifth, contain the modern information system architectures. The advent of internet enabled the previous, basic client/server networks to merge on an enterprise level, into larger structures where enterprise servers and databases contain most information and applications while client computers use this information anywhere in the world. Now, with high-speed data transfer being the norm, the newer cloud computing solutions have taken root and now the enterprise does not necessarily need to own or control any of the server side capabilities: they simply buy access to the software from a service provider that takes care of everything else except data management itself (Laudon & Laudon 2014: 200).

SIS systems are often fourth or fifth era cloud computing or enterprise client/server solutions or hybrids of the two that combines these features in one form or another. Cloud computing solutions in their simplest form only require the users, both students and teachers, to download an end-user application or uses an internet browser based interface to access the software. Enterprise level Client/server solutions are slightly less common, although they are prevalent in public education, where the cost of maintaining expensive servers are outweighed by access and network security when dealing with a very large amount of students (Kumar 2000). Both of these software types have their advantages and disadvantages, with client/server solutions having the benefit of greater overall administrative control over the software and data speed due to owned dedicated

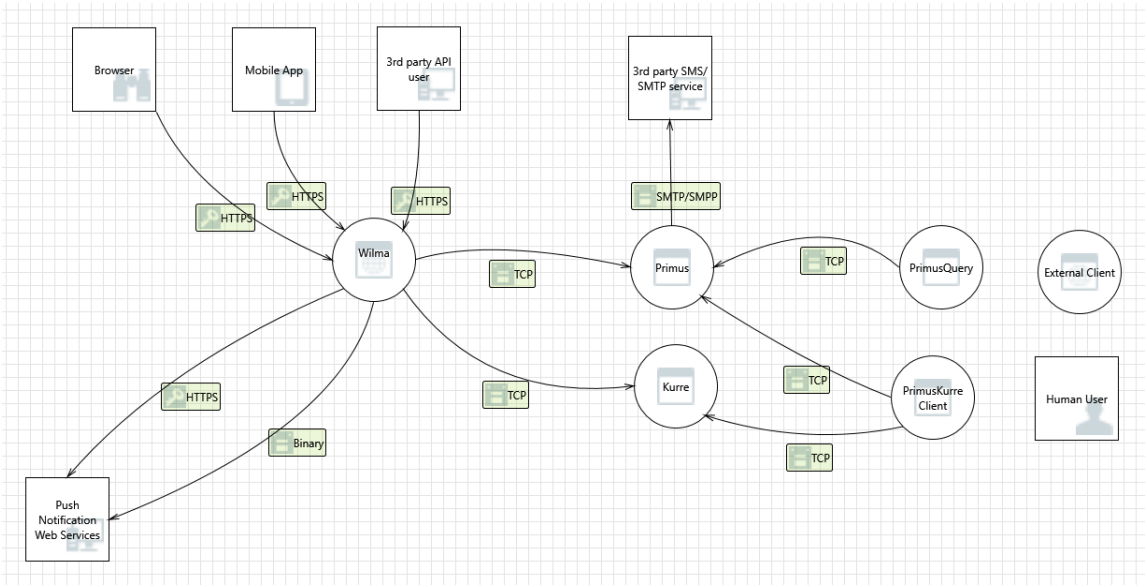
servers but requiring much more resources in the form of IT personnel, servers and ultimately, money. Cloud computing solutions on the other hand are often much cheaper in every discernable context, but their customization options are often limited, expensive and complicated, as they affect not only the entire software at once, but also every other user (Larkin 2018).

4.2 School Management System Features

School Information Systems are, as earlier stated, a specialized form of a Management Information System. The general features and requirements of a School Management Systems include different forms of course management, including enrolling, grading, feedback and results; student management, including profiles with attendance, grades, contacts etc.; and general school management, including communication, administration and so forth. All SIS systems do not provide all of these features as some simply provide a virtual learning environment, which are called Learning Management Systems and offer less school specific tools, but often overlap in features with SIS systems, while some systems simply provide an electronic database for the school's students. This creates a system that must be able to manage massive amounts of data and this data must be stored efficiently and be easily accessed, while data security is kept at a maximum (Kumar 200).

While data movement is often a defining feature for all MIS systems, in the school world this data is geared towards user data management and analysis which creates unique requirements where the database capacity must be unusually large. The data transfer speed and availability similarly need to be significant enough to be able to accommodate an ever increasing amount of users and finally, this data needs to be available for a long time after the user's active role in the system ends. This creates challenges in software architecture, integration and maintenance, especially when this data needs to be available for the users all the time. SIS systems also have two very different user types: school staff and students. Both of these groups have differing requirements and expertise which need to be filled. The school staff needs

administrative rights of differing levels, analysis tools, communications and different logistical tools. The students need to be able to access this information easily, especially communication with the administrative side of the system.



Picture 4. Wilma SIS architecture (Visma 2018) Permission to use the picture given (Kenttälä, 2018, personal communication via email, <19.11.>)

Wilma's (Visma 2018) architecture example above in Picture 4 visualizes a basic structure of an SIS system. In this case the system is handled so that most of the user interface layers are handled in Wilma, with communication, course feedback and user data modification as primary tools, while Primus serves as the primary teacher tool with most teacher related administrative tools and reports, and Kurre acts as the primary administrative tool with data management, analysis and reporting tools. These three software bundles then work in unison to serve the different users of a school. While this picture is specific to Wilma, it provides a solid example of the different program relations inside an SIS system, as well as, the user interface layers a SIS system can have.

Like most modern fifth era SIS systems, a notable feature of Wilma is that the system can be either self-hosted as an on-site server/client structure or as a cloud computing solution provided. The main questions when deciding whether to maintain a server/client structure or to take a fully automated SaaS solution is dependent on

available infrastructure, school size, IT personnel and local internet connectivity, although eventually the cost of maintaining the system is the deciding factor. For instance a very small school of 50 people only needs one server while a still relatively small school of 500 people requires two. Once the school size grows to 1000 to 5000 students the server structure requires 4 computers with one computer dedicated to data control alone (UNESCO 2012). This quickly increases cost as you need the necessary infrastructure, resources and technical capabilities to maintain the servers, but you also need to have available a stable, high speed internet connection as a 5000 student school would require a constant connection of 10 MB/s (Haapsaari 2018). Also, data security is an issue that has to be solved and maintained. When taking a SaaS solution, all of these details are handled by service provider and the school only needs client side computers. Finally, implementation of the SIS to the school is also a factor, as both students and teachers need to be educated to the use of the system.

SIS systems are now days becoming increasingly common as the requirements of the networked, modern world simply demands electronic solutions to school management, student data management and communications channels with the school. At the same time Learning Management Systems are similarly evolving and becoming more common, as they enable certain studies to be done in a virtual environment, reducing the resources that need to be placed on the studies. These systems together highlight a phenomenon where studies and schooling is more moving into the cloud computing era as it provides cost reductions and improves study possibilities overall. Before the cloud computing era, most of learning and school management solutions were expensive, difficult to maintain and required special IT knowledge. However, in smaller schools, these server/client solutions are still viable as in a small scale they are easier to maintain and implement compared to cloud computing alternatives.

4.3 Wilma

The material for this thesis is a School Information System called Wilma, created by a Finnish IT company Starsoft and currently developed by Visma (Haapsaari 2018).

More specifically, the research focus of this thesis is on the communications module of Wilma. Wilma was originally published in 2001 and was developed under Starsoft until 2016 when the entire company was bought by Visma and then integrated into Visma and its different subsidiaries. Visma is a software company that specializes in corporate software and software services in the Nordic countries, Romania, Netherlands and Latvia (Visma 2018). While a part of Starsoft, 50 people were actively developing Wilma and currently, under Visma, the amount of developers has increased to 60. Wilma currently has a significant market share in the Finnish education system with 95% of all the different grade schools (In the Finnish education system: Ala-aste, Ylä-aste, Lukio) and 75% of vocational schools (Finland: ammattikoulu) totaling in a 90% market share in the public sector and an extra 10% in the private sector, along with a few schools outside Finland (Haapsaari 2018).

Wilma itself is a modern School Information System software package that is designed to provide electronic administrative capabilities for a school and includes significant databases for student information, communication capabilities, administrative tools, report and analysis tools, logistics tools and much more than is to be expected to be included in a School Information System. Wilma, as a fifth era Management Information System is designed to work with most modern portable devices including smartphones and computer desktops, but also provides options for managing vital user information, both students and teachers, a key factor in Student Information Systems. A client school has the option to either maintain their own database server, that is then linked to Wilma via internet, or then it can buy the necessary database infrastructure directly from Visma as a Software as a Service solution. Both of these options have their advantages and disadvantages which are closely related to the cost and size of a client school. See Chapter 3 for more background information on SIS systems and the cost/technology discussion.

The reason Wilma works well for this thesis is due to the software's history and market share. When Wilma was originally developed by Starsoft, the design team had a significant deficiency: it lacked all coordinated and planned user interface and user experience design and this translates so that Wilma has some unconventional user

interface solutions but also that internal cohesion of the user interface has been sometimes in question (Haapsaari 2018). Currently, under Visma, there have been significant improvements on both user interface and user experience design, with training and planned user experience design. However, some aspects of the original Wilma are still present, especially on the school administrative side, and feedback has sometimes been “colorful”, although on a system and user base of this magnitude this is to be expected. Due to the outright commonness of the software, feedback is also abundant and thus a multitude of different subjects and points of view can easily be found, both from student and teacher side of the user spectrum. This feedback could similarly be considered coming from expert users as Wilma is an integral part of modern Finnish education. Finally, Wilma’s communication module was specifically chosen to be studied in this thesis due to the fact that it is simply used the most and thus possible shortcomings and improvement suggestions are likeliest to be found as well.

5 RESEARCH DESIGN

The research method of this thesis is a qualitative approach based case study with an interview, containing a questionnaire and heuristic evaluation, at the core of the method. The aim is to study the software's communication module for usability and affective aspects. To help guide this study, three research questions were set:

1. What kind of aspects generate negative emotions in a user interface?
2. How do these negative aspects affect the use of the system?
3. What are the effects of these negative emotions towards the use of the system as a whole?

Questions one and two have been discussed in Chapter 2 of this thesis with some examples of possible negative emotions along with general effects of these emotions. Question three however, while discussed and considered at the end of chapter 2, is eventually the target question of this thesis and the basis of the case study. To reach this goal, a two-part questionnaire and a heuristic evaluation guided by a use case scenario is created for a set number of test subjects. The test subjects, henceforth called evaluator to better describe their function in the thesis, are chosen to represent a certain segment of the target software's user base. In this case, the users are parents of children in the Wilma system and teachers. These two segments are the largest and most significant for the system.

The questionnaire is done as a three-phase interview, where the evaluator, guided through the questions and the evaluation. The first phase of the interview is a preliminary information questionnaire, with basic user information to create a background, available as Appendix 2 in Finnish and Appendix 3 in English, followed by a brief introduction to the heuristic evaluation task along with an introduction to the list of heuristics created for this thesis. It should be noted, that the interview is conducted in Finnish, as all of the evaluators are Finnish speakers, and thus the material is given to them in Finnish as well. All material used for the interview is available in both

languages in the Appendix section. The questions are not given to the evaluator, with the exception of the list of heuristics to keep the task in mind, as the interviewer asks the questions.

After the preparatory tasks for the evaluation has been done, the evaluator is presented with four test cases, available as Appendix 4 in Finnish and Appendix 5 in English, where the evaluator is tasked with using different features of the communications module. These tasks are designed to go through the most important features of the system, including messaging, group messaging, inbox and search functionalities. The evaluation is done in the Speaking Aloud (Lewis 1982) method in which the evaluator is speaking actively about what he or she is thinking and doing while the interviewer makes notes about different details that come up during the evaluation. This helps understand the different thoughts and emotions the evaluator is going through but also keeps the situation alive and makes the evaluator more comfortable during the evaluation task. The main task for the evaluator at this phase is to conduct the heuristic evaluation by finding usability problems and then categorizing them.

The third and last phase of the evaluation is affective design questionnaire, which is designed to discover different affective aspects present in the software. The questionnaire, available as Appendix 5 in Finnish and Appendix 6 in English, contain questions designed to not only gain thoughts about how the evaluator feels about the software, but also the general thoughts about the evaluation process. The affective design portion of the research is reinforced with this second questionnaire, as finding emotional affecting aspects with a heuristic evaluation alone is difficult. With the help of a questionnaire, aimed at finding general, emotional thoughts, the affective aspects of the material can be assessed and considered more clearly.

The goal of this interview is to discover if there are any emotional aspects present in the target material, along with any usability problems, and discover how they affect the usability as well as system as a whole. The hypothesis of this thesis was that some causes for frustration and aggravation can be found in the system and they will have a negative effect on the overall user experience of the system, albeit exact results may

vary. The interview is designed so that the heuristic evaluation process helps discover any potential emotion affecting features while the questionnaire is designed to find insight into how these features affect the system as a whole. To get different types of backgrounds in the research, the evaluators are selected from both teachers and student parent users. In this way the thesis has some variation, but is also limited to adult users to limit variation. The questionnaire is eventually rather simple, but with a combined research where the heuristic evaluation is designed to find exact features and the questionnaire is designed to give less exact emotional insight to these discoveries, the results should be solid enough to create some conclusions, even if emotions are always difficult to assess.

6. DATA GATHERING

The main purpose of the research of this thesis is to evaluate the effects of negative emotions towards software. This, however, is not a straightforward task as while you could ask any user of the software how she or he thinks about it the answer would not be exactly scientifically relevant in any way. For the answer to be relevant, first we must establish the quality of the software itself: does the software have aspects that create negative emotions? To answer this, this thesis uses a UX point of view, which has all the necessary tools to discover possible flaws in the software. Once a credible level of quality is established, the possible emotions the user has must be discovered and to this, the Affective design inspired questionnaire is conducted. After that, you can ask the emotional response the software creates and know the answer has scientific value. Thus, the material of this thesis is based upon a heuristic evaluation, which is designed to test the software negative aspects, and a questionnaire, that is designed to evaluate the if these aspects create negative emotions and, finally, together these can be used to evaluate negative aspects affect on the system as a whole. The following chapter discusses in detail the data gathering process.

6.1 Test Cases and Questionnaires

The research method for this thesis is an interview that contains a questionnaire and a heuristic evaluation with four test cases. All of these can be found in the Appendix section, Appendices 2 to 6 presented in the order they are presented to the evaluator. The interview begins with a background information questionnaire, which is designed to find out what kind of history the evaluator has with the system: how often it is used, what type of user he or she is, how the evaluator would rank the system[1] and what are the general thoughts of the system. All of this is important information when considering the persona of the evaluator, albeit the background of the evaluator is allowed to be otherwise relatively free, with the exception being of general user types,

[1] Grade between 4 to 10, the Finnish lower grade school grading system

which are set to parent of a student in Wilma and teacher using Wilma. Also, of specific interest in the questionnaire is the question concerning previous thoughts of the system: personal ranking and thoughts of the system. These questions establish any possible bias towards the system which could emotionally affect the test results. While biases are allowed to exist, they should somehow be noted to establish possible discrepancies in the results. Finally, bias towards the system also already tells something of the success of the system: if there are biases, there is also possibly some kind of user experience affecting issue, not only a usability issue.

The second part of the interview is the heuristic evaluation, which requires first a briefing of the evaluation task. The evaluators are all non-professionals in usability and thus they require education in the subject. The heuristic evaluation itself is a simple and intuitive method, so the evaluators only need to know the basic functionality of the method along with a few examples to understand it. Also, the interviewer can help with functional matters, although care must be taken not to affect the evaluation otherwise. The four test cases otherwise revolve around the main features of the communications module of Wilma: login, sending messages, receiving messages and search. The test cases are thus aimed towards them.

The first test case is basically a warm up task where the user needs to login on to the system and find the communications module. In the second task the user needs to send messages to certain people and groups. It should be noted, that the possibility to send group messages is only available to teacher users, thus the feature is new to parent users. Thus on this mission there can be some discrepancies found between the evaluators as it is a new feature for half the group. This part of the test was still decided to be left even for non-teachers for the sake of keeping the tasks unanimous; it is enough to keep in mind that if there are notable differences in the results it is due to the unfamiliarity of this feature. The third tasks is getting messages and answering to them. The last task is the search functionality, where the evaluator needs to find a certain message with certain parameters given in the task. This task is the most complicated and thus the results might be similarly most varied. Overall the tasks are relatively open ended and short with the exception of the last search task. They should be sufficient in

discovering the most notable usability problems but also any emotional aspects that specifically affects the evaluator. Most notable discrepancies are most likely found in emotional aspects of the tasks which can be further analyzed by the second questionnaire.

The second questionnaire asks a set of open ended questions to establish general thoughts about the software. The questions themselves have been divided into two sections: Test related and system related questions with both sections having a few questions particularly focused on emotions. The questions focused on the test itself, begin with a general estimate of the usability of the system, general thoughts of the test and if the evaluator found something that specifically felt confusing. These questions are designed to be simple starting questions which are there to support the heuristic evaluation. Also, the last question, which asks for a specifically confusing detail, is also the first question concerning the emotional spectrum.

The second category asks simply what do you like about the system, what should be improved or what is missing, is there something in the system that frustrates you specifically and what is the first thought or feeling you get when using or talking about using the system. This last question is perhaps the most important question in the set, as it basically asks straight out the hypothesis question of this thesis: what is the emotion that best describes the entire user experience of the system. The answer for this is obviously very subjective, but in combination with all the previous questions, it allows a relatively good insight to the evaluators feeling of the entire system. While these answers in general will vary greatly, they will add extra information to the heuristic evaluation and together are designed to discover if there are any affective aspects in the system and how they affect the user experience.

The last two questions are a free word question, where the evaluator can sum up what he or she thinks of the system and test while the last question is possible extra questions that the interviewer might ask during the interview. Any details or questions the interviewer finds relevant are also placed here. When dealing with emotions and doing an interview, there might turn out to be important details or questions during the

interview that are good to record: emotions are not always simple questions and dynamically interviewing a person might yield unexpected results.

It should be noted that all of these questions are more or less questions concerning the idea “how do you feel” and there are not expected to be any exact answers. This questionnaire does, however, somewhat rely on the heuristic evaluation discovering some issues or problems in the system, but even if there are none, only positive answers, these also tell something about the user experience of the system. Positive aspects, especially on user interfaces, are difficult to see as it is only a means to an end, but positive user experience can be detected from a feeling of success or other positive emotions. These can be detected with the questionnaire and give results, even if the heuristic evaluation would not. Eventually these two test methods would also work solo, without the other, but they give much clearer results together: the key weakness of the heuristic evaluation is that it cannot see a forest from the trees: it concentrates specifically on detailed issues and the big picture is left hidden, while a questionnaire, no matter how detailed, often leaves so vague or general answers that they do not eventually tell anything specific of the subject, much more so when discussing emotions. Together they can, however, tell not only more about the state of the system, but also how it affects emotions or how emotions affect the system: details in the questionnaire can potentially be directly linked to a specific error in the system. Still, as a final note, both the heuristic evaluation, and the questionnaire supporting it, are only as good or effective as the questions guiding the action.

6.2 Interview Method

Before going into detail about the results itself, it is good to mention a few details of the interview itself. First of all, the process itself was the same in all interviews: the interview begins with a small background questionnaire to determine a rough Persona – type for each user, followed by the heuristic evaluation and then ending in the affective design questionnaire. All of the evaluators were given a short briefing how the entire questionnaire will be conducted and specifically about the functionality of the heuristic

evaluation although the five category heuristic evaluation with a total of fifteen sub categories was relatively difficult for a non-professional UX subject to handle so quickly. However, with the help of the interviewer, the heuristic evaluation was successful: The idea of the interview was that the evaluator used the Speaking-Out-Loud –method, where the evaluator constantly spoke what he thought and did during the evaluation, and whenever a problem was discovered, the interview paused and the evaluator and the interviewer, with sufficient UX understanding, stopped briefly to consider what kind of an issue it was and how severe. When the type and severity was agreed upon by both, the evaluation continued. This assisted evaluation process worked admirably, as while the evaluator might not have had a significant understanding of the type of the error that was discovered, he or she knew full well when some aspect of the software was not working correctly or to his or her liking. Also, a good estimate of the severity was similarly easy to discern. Thus, with a little assistance from a UX professional interviewer, a person with no knowledge with any UX related subject was able to conduct an efficient heuristic evaluation with clear results.

This approach might be slightly unorthodox, but in effect the task of the heuristic evaluation is to find problems in the target material. What the actual type of problem is, is only significant to the overall dataset when considered with the results of other heuristic evaluations. Thus the idea of creating categorized results with the assistance of the data manager, or in this case, the interviewer, is acceptable for the end results. A non-professional UX evaluator does have the disadvantage of not exactly knowing what type of issues to look for, but this makes the results found by the evaluator much more spontaneous and in a way more realistic as the results have no UX filter: the evaluator finds problems that he or she honestly thinks work incorrectly.

6.3 Evaluator selection

The test subjects, or evaluators, of this thesis were five people, as suggested by Kuutti (2003), regularly using the material software Wilma. Users from all of Wilma's three different user groups were chosen: students, teachers and guardians of students. There

were only two additional rules in evaluator selection: the evaluator must be over 18 years of age, or in other words an adult, and must be familiar with the software. These rules were set mainly to clarify the test procedure and to streamline the results. The first rule of adults was set because under-aged users of Wilma have a tendency to concentrate on details that are not relevant for the software, but affect them personally. Such details are for instance personal notes about misbehavior, which are usually plentiful for younger children. Adults do not have a bias towards such notes, or they at the very least understand, that this is a result of their own actions, not so much actions of the software and are able to ignore such details with greater efficiency.

Secondly, the reason the selected evaluators were required to have previous knowledge of the software is due to the inherent complexity of an SIS. A user with no previous experience with such complex software is bound to have biased and wildly varying perceptions towards the system and thus they would differ greatly from those of even a beginner level user. Taking the perceptions of a complete stranger to a system is in itself an interesting subject, but clearly outside the scope of this research. There is bound to be some level of variation between the users, especially between the categories as teachers are bound to have greater knowledge of the system than guardians, but this in itself is an acceptable deviation as the target is to research emotions, not Wilma itself. Similarly, to have three different user types on a heuristic evaluation, with vastly different requirements and motivations for the software, would normally be unacceptable due to the significant variation in results. However, as the target is not to evaluate the material per say, but to discover the emotions concerning the software, having all three groups represented is more than acceptable.

6.4 Evaluator personas

The five different evaluators of this thesis were categorized into Personas to establish their age, user category, approximate abilities with the software and their initial thoughts about the software. This categorization is done to establish the evaluators

required attributes for the research and, more importantly, reveals if they are biased towards the software in some way. This is an important detail, as if an evaluator has a specific bias towards the system or a detail of the system, it affects the results of the research, giving unnecessary weight towards certain aspects in the system. The details for the evaluators can be found in Table 4 below.

Table 4. Evaluator personas division

Role / ID	Age	Use per week	Score	Use reason	Expertiese	Persona type
Guardian 1	53	Mixed, more in the past	6,5	Daughter was in school, no active use now	Advanced, several years of experience	Advanced guardian
Guardian 2	49	Mixed, more in the past	7	Daughter was in school, no active use now	Advanced, several years of experience	Advanced guardian
Student	30	3 times a week, ~30 minutes	7	In use at own school, vocational studies	Advanced, several years of experience	Advanced student
Teacher 1	33	Several hours per day	7,5	Teacher, extensive daily use	Expert, professional user, several years of experience	Professional teacher
Teacher 2	42	Several hours per day	8	Teacher, extensive daily use	Expert, professional user, several years of experience	Professional teacher

The persona allocation, Table 4 above, is divided into six different categories: User role and identification code, Age, Use per week, Preliminary score before the test, Reason for use, Approximated expertise and eventual Persona type. The user role is combined with identification to name the evaluators and along with stating their user type in the system. Age is technically irrelevant for the results of the research however it provides some background about the evaluators and user groups and to be allowed for testing you had to be a legal adult or 18 years of age. The reason for this is that under-aged evaluators have a high probability of not concentrating on relevant aspects in the

system. Preliminary score is to test if the users are biased in some way and to give some initial idea on how they perceive the system. Use reason details simply why they use Wilma and provides again some background about the evaluators. Expertise is basically deducted from Use per week and Use reason: for instance, Guardian 2, while having currently reduced, mixed use of the system due to her daughter being older and able to use the system solely, can still be considered an advanced user due to having been using Wilma for approximately 5 to 10 years or the time her daughter was in the lower grades and upper grades of the Finnish education system. All of the users were eventually at minimum advanced users and this was basically a prerequisite to be taken into the interview and heuristic evaluation: the results would be notably different and thus compromised if the testers were beginner or basic level users.

The last category was the eventual persona of the evaluator and it should be noted that the eventual persona is narrow in the sense that there was eventually only three attributes set to each persona: user type in system, expertise and age. This is however sufficient for the research as it focuses less on the groups themselves and more on very personal details. For instance, in the case of a normal heuristic evaluation, mixing two user groups with wildly different requirements and expectations from a system would be completely impossible due to the significantly different results: the groups would require tailor made tasks and this was noted especially in the answer between the two guardians and two teachers. However, as the research eventually focuses on the combination between the heuristic evaluation and the emotional aspects, mixing the groups is not only acceptable, but it in fact gives a broader spectrum of results. Eventually, all the evaluators were categorized as a minimum of being an advanced user, having several years of experience with the system and being above the age of 18.

6.5 Evaluator Preliminary Insights

The last detail that could have an effect towards the research results were the preliminary emotions and thoughts the evaluators might have towards the test system. This was established with two simple questions: what is your overall grade of the system and free word before the evaluation. While these questions are open ended and

simple, the evaluator is most likely going to talk about a certain detail that they remember the most or have thought about the most. The overall average grade of the system between the evaluators was approximately 7 although one of the evaluators changed the value during the first test. Also, more than one evaluator gave comments of the overall grade, including one saying that they did not think it was particularly effective.

The free word section on the other hand gave much more varied thoughts from the evaluators, as expected. Two of the evaluators stated outright, that the system was difficult to approach, albeit you learn to use it quickly. The teachers stated that they considered it a useful tool. The preliminary answers were relatively ambiguous, which would suggest that the evaluators had no special thoughts concerning the system. Overall, the evaluators were deemed sufficient and acceptable for the purposes of the research, even with one of the guardians having a minor bias towards the system.

7 RESULTS OF THE HEURISTIC EVALUATION

The heuristic evaluation was done in this thesis to gain evidence and insight to the quality of the user experience in the target material, more specifically to make assessments if the target material has aspects or details that cause negative emotions. The evaluation itself was done with a custom made list of heuristics, available on page 34, by five different non-professional evaluators with slightly varying base Personas. Once a problem was discovered it was rated in severity between one to four, with one representing minor cosmetic problems and four representing major usability problems. As the evaluators were non-professionals, they were assisted slightly in selecting the problem categories.

The results of the heuristic evaluation can be found on the next page as Table 4. All results are first named by their feature, the problem function is stated, how many times they were found by different evaluators, their assessed severity is stated and, finally, their problem represented category is mentioned. It should be noted, that the severity rating is colored for severity in the results for easier recognition: green are level one severity, yellow level two, orange level three and red represents level four. Categories that had mixed feelings in severity are also mentioned and are always colored towards the higher level. When categorizing problems, the evaluators were only tasked to state the primary error category of the list of heuristic. These five main points were Aesthetic and minimalist design, Efficiency of use and standards, User control and freedom, System functionality and Error prevention and Support. The letters A, B and C were added after the interview to represent the exact category. As a final note, the results of the heuristic evaluation are not the main focus of this thesis, simply the first part, and therefore only a general overview is done of the material.

Table 5. Overview of problems found in the heuristic evaluations

Feature	Problem type	Times found	Severity	Category
Recipient selection Feedback	Feedback	1	1	2b
Course message vague	Feedback	2	3	4b
Selected recipient vague	Feedback	1	2	4b
Search functionality guidance	Guidance	1	3	5c
Student personal info lacking	Information	3	3	1a
Recipient selection too narrow	Information	3	2 / 3	1a
Seach field auto-empty	Function / Information	1	2	2b
Memorizing information	Information	2	2	1a
Difficult to find information	Information	3	2 / 3	1c
Escape -functionality issue	Function / Nonfunctional	1	4	2a
Message selecetion issue	Function / Non-standard	3	1 / 2	3a
General unclear funtionalities	Function / Non-standard	1	3	4a
Lack of organization	Function / Control	2	2	1b
Hourly marking issue	Function / Control	1	3	4c
Notification settings issue	Function / Control	1	3	4c
Multiple recipient selection issue	Function / Control	2	2	2b
Message chains issue	Function / Navigation	2	2 / 3	2a
Navigation difficult	Function / Navigation	1	3	4b
Memorizing of functionalities	Function / Non-standard	3	3	4b
Unnecessary complexity	Function / Non-standard	3	2	4c
Lack of Draft messages	Function / Non-standard	1	3	2a
Page selection issue	Function / Non-standard	1	2	2b
No Attachments function	Function / Missing	1	2	1b
Teacher list illogical	Function / Information	1	3	1c
Search functionality unclear	Function / Information	3	2 / 3	2a
Teacher abreviation issue	Function / Information	1	2	2a
Non-standard message functionality	Function / Non-standard	1	3	2a
Message finding unclear	Function / Information	2	2	2b
Jump to newest function	Funtion / Unnecessary	1	2	2c
Layout	Visual design	1	1	1b
Site objects unfitting	Visual design	1	3	1b
Static message chains	Function / Control	1	2	4c
Slow connection issue	Error Free	1	2	5b

7.1 Overview of the Discoveries

The five evaluators discovered 53 problems in total, out of which 33 were unique problems. No single issue was discovered by all evaluators, albeit thematically they identified all the same issues present in Wilma. Out of these errors, three significant problem themes could be detected: problems relating to information management, problems relating to the general communication framework and finally, to a slightly lesser degree, problems relating to feedback and visual details. Most of these issues were rated to be of level 2 or 3 in severity, with only three level 1 problems and one level 4 problem. The heuristic categories themselves were represented relatively unevenly: Aesthetic and minimalist design; Efficiency of use and standards; and System Functionality gained most of the problems, with 17, 16 and 13 problems respectively found in these categories. User Control and Freedom gained only four issues and Error prevention and Support gained two problems.

Overall the results found by the evaluators are solid and somewhat follow the theoretical framework: each evaluator found between 7 and 14 errors, which is roughly between 20 to 40 percent of the total unique errors found, which follows somewhat the amount Kuutti (2005:47) suggests, and the most severe usability problems were thematically discovered by all evaluators, very similar to what Korvenranta (2005: 115) stated. It should be noted, that although no single problem was flagged above all else, a problem with general search functionality in Wilma was found by all evaluators. This same problem can be found in several different search functionalities around the software and the evaluators found the same problem, simply from a different area of the program. The largest problem type was in different kind of functionalities with two thirds being categorized as different kind of issues with single functionalities or general functionalities. This is, however, to be expected as non-professional evaluators are more prone to looking at details instead of general themes, with a focus on personal preferences. This in itself is not a detrimental aspect in the evaluation as the focus of the entire evaluation is to assess and discover subjectively affecting emotional aspects in the material.

7.2 Categories and Severity

53 problems were discovered in total by the five evaluators and there were 35 unique problems and the vast majority of these errors could be found in three categories: Aesthetic and minimalist design (category 1 henceforth), Efficiency of use and standards (category 2 henceforth), and System Functionality (category 4 henceforth). The other two categories, User control and freedom category (category 3 henceforth) and Error prevention and Support category (category 5 henceforth) gained practically negligible amounts of errors, with five and two errors respectively. Table 5 found in the next page shows these problems divided in the list of heuristics used for the evaluation.

The reason for this particular division of the results into these categories can at least partly be attributed to the non-professional nature of the evaluators: categories 1, 2 and 4 are somewhat clearer and thus easier to assess problems as they are based on visuals, effectiveness of use and system functionality which are familiar concepts for non-professional usability evaluators. Freedom of use and software control is a somewhat unfamiliar concept for a non-professional. For instance, category 1 problems relating to information gathering were easy to discover as were issues with standardized functionalities from category 2, but only extreme cases of difficult navigation were detected by the evaluators, despite the fact that from the interviewer's expert point of view, the navigation was unusually heavy in several situations.

Interestingly, the absence of category 5 problems, Error prevention and Support, was notable as only two errors were discovered, both relatively minor issues as well. You would expect that assistance would be wished for more in conjunction with complicated software such as an SIS, but this is perhaps a phenomenon in itself in conjunction with this type of software in general or perhaps simply with Wilma, as the software is notably lacking in any significant help files. It should be mentioned that direct software errors were rare, with only one discovery during low latency conditions.

Table 6. Errors divided according to Heuristic categories

1. Aesthetic and minimalist design	**Total:**	**17**
a. Presented information is relevant, there is sufficiently of it and it furthers the systems functionality.		8
b. Objects of the system (pictures, colors, names etc.) are of appropriate size, style and utility is consistent and clear.		5
c. Information is displayed in a pleasant way that furthers the functionality of the system.		4
2. Efficiency of use and standards	**Total:**	**16**
a. The concepts of the system are familiar and correspond to established standards.		8
b. The system should give feedback to the user and functions should be transparent.		7
c. The system has no distracting, excessive or unnecessary objects.		1
3. User control and freedom	**Total:**	**5**
a. Using the system is simple, clear and motivating. Navigation is clear and logical.		4
b. General computer tools, such as key combinations, are supported. Backwards movement and exits are available.		0
c. Using the system is pleasant, interesting and furthers the users goals.		1
4. System functionality	**Total:**	**12**
a. Functions progress logically, clearly and are not unnecessarily complicated.		1
b. Actions and objects in the system should be recognizable when seen, not remembered.		6
c. Systems functionalities do not frustrate or confuse.		5
5. Error prevention and Support	**Total:**	**2**
a. Faulty input can be repaired or modified.		0
b. Objects, functionalities and controls are error free and clear.		1
c. When required, help or support can be found easily, quickly and is useful.		1

Severity ranking was in general between levels 2 and 3, with only three problems being identified as cosmetic level 1 issues and only one problem being ranked up to level 4. The rest were split evenly between levels 2 and 3, with 25 problems being ranked to level 2 and 24 as level 3. Overall, these results are also as expected: non-professionals

have difficulties assessing what a simple cosmetic error would exactly be and similarly assessing level 4 errors is felt as extreme. The single level 4 problem was, in fact, ranked as a 4 simply because the functionality did not work at all as intended, according to the evaluator. The level 2 and 3 problems were easier to assess as the evaluators unanimously searched for problems according what they thought was incorrect: usually less severe problems were ranked 2 and more severe issues were level 3.

Category 1

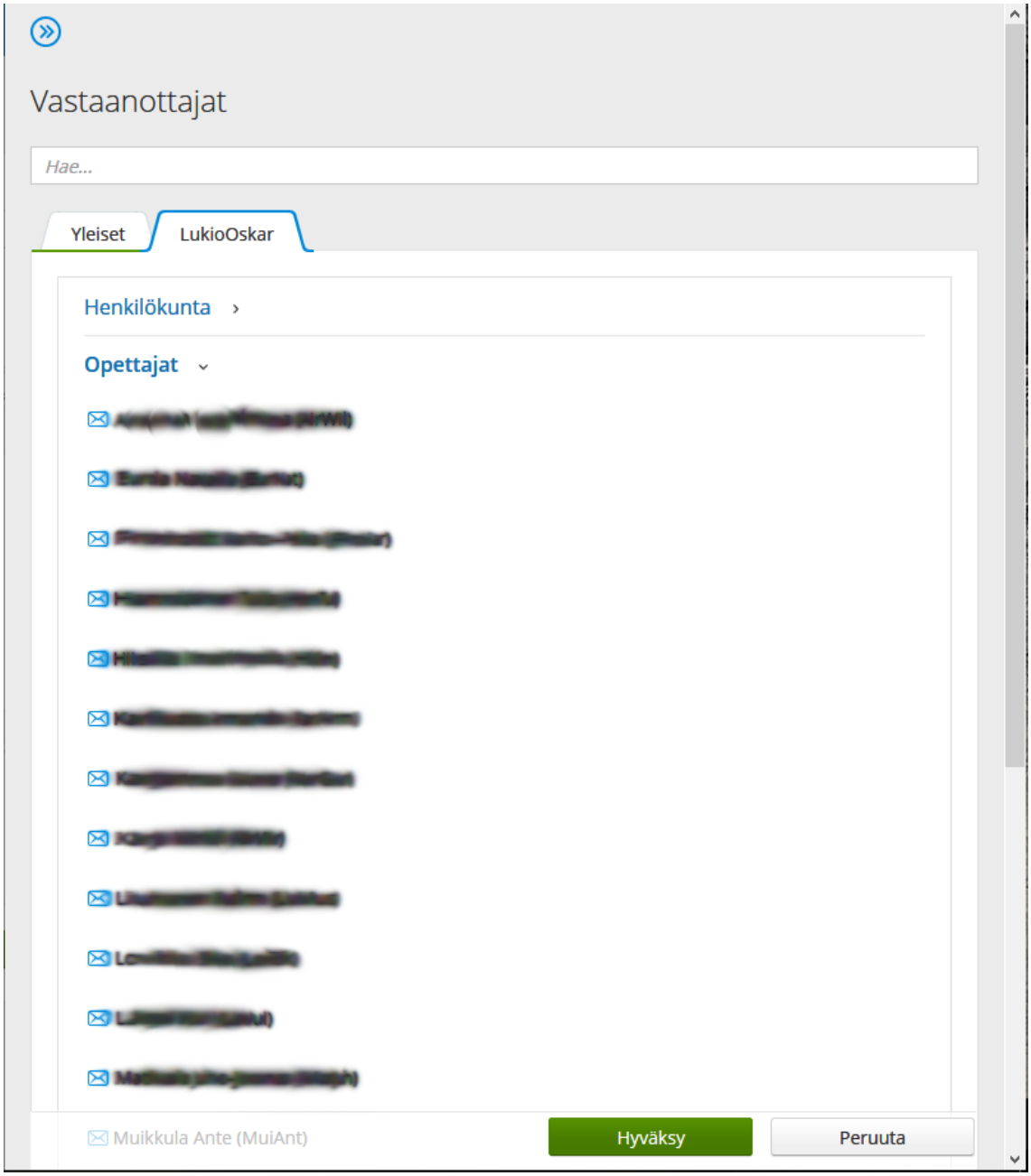
The three categories with the most problems, namely category 1 with 17 discoveries, category 2 with 16 issues and category 4 with 12 problems were in contents very varied, although thematically the discoveries were consistent. Category 1 with the most discovered problems, for instance, almost exclusively focuses on information design and management related issues, although the problems themselves vary in detail with problems relating to the lack of organization options to finding information. Missing features were also placed in category 1, such as the lack of file attachments in the messaging system, although they are perhaps more of a System functionality related issue. This in itself is perhaps a shortcoming in the heuristic list's design, but for the purposes of the evaluation, the result is clear none the less.

A detail concerning the emotional spectrum of the research can also be found in category 1c, or more accurately "Information is displayed in a pleasant way that furthers the functionality of the system", was specifically created and placed in the list of heuristics to evaluate emotions and nearly all evaluators found problems relating to this category which were considered to rank often in level 3 on the severity rating. All of the discoveries are also of information design and management related features, which would already suggest that the software does indeed create negative emotions and at least one of the causes is in insufficiently presented information. As a final note concerning category 1, an unusual detail can be seen in the results is a notable lack of issues relating to visual design, such as the use of objects or other visual ques. There were a few cases, but they were of low severity ranking and mentioned only a few times.

Category 2

Category 2, or Efficiency of use and standards, had the second most discoveries with 16 problems, focused significantly on functions that worked incorrectly, often in a non-standard way. Also, the issues with information design were discovered in this category, but in a different form. For instance, different non-standard search functionality problems were found in several cases, a few navigation issues relating to information gathering and feedback issues relating to information were discovered. In fact, over 75% of the discoveries related in one way or another to information design or displaying information and the rest of the cases were mainly focused on non-standard functionalities. Severity ratings were nearly all level 2 or 3, with a single level 1 problem and the only level 4 issue was considered to be a non-functional feature.

Overall, category 2 presented similar finding to category 1: the main issues with the software focuses on information design and management related issues with a notable presence of non-standard functionalities contributing to this issue. The missing functionalities, the shortcoming mentioned earlier in the heuristic list, were placed in this category. For instance, the unfinished messages, or so called “Draft” messages, were thought to be missing and the issue was placed in category 2 while it is more relating perhaps to category 4. Similarly, the only severity 4 issue discovered in the entire system was placed in category 2, seen as a non-functional feature, although it could be considered a category 4 issue. Both of these functions do, however, relate to non-standard functionalities as well, thus the outcome is somewhat ambiguous. Interestingly, the Draft –messages functionality is present in the software, but the user simply did not find the functionality. This suggests that this is not perhaps a missing functionality issue, but an issue relating again to insufficient, non-standard information design feature, as normally draft messages are used via a pop-up prompt asking if the user wishes to save the unfinished message.



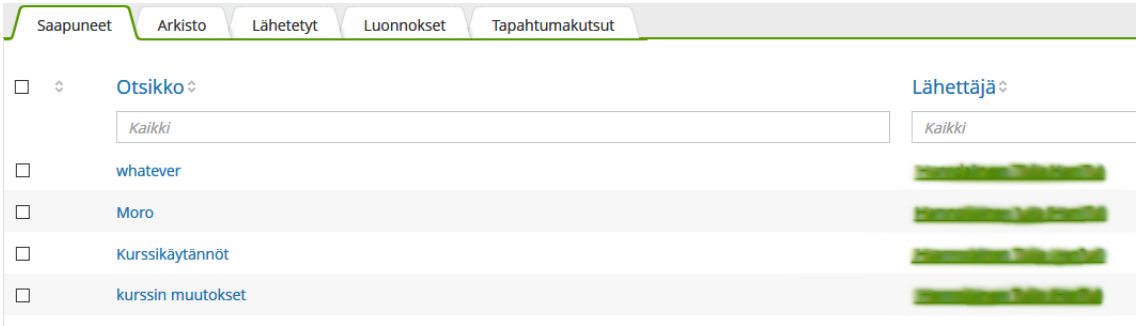
Picture 5. Message recipient selection screen in *Wilma*. (Visma 2019)
Permission to use the picture given (Kenttälä, 2019, personal communication via email, <1.7.>)

Picture 5 above displays the message recipient selection screen found in Wilma. This particular feature was pointed out by all evaluators to have several different issues present in both category 1 and 2. Category 1 related issues included insufficient recipient selection, as the feature was found to be lacking in information design. The list

only displays the name and system tag for the recipient, but displays no other information. The system does have a feature which displays all current teachers, but even then you have to remember who the teacher is and there is no additional information displayed. For the evaluators to find information you had to go to other areas in the system and this was considered troublesome and frustrating by the evaluators. Picture 5 also displays the search feature that was found by all evaluators. This issue was placed in category 2 as a non-standard or unclear feature. The issue with search functionality points to the system not having an activation button to start a search, instead to start the search you need to press the "Enter" –key from the keyboard. This confused all of the evaluators, although it should be mentioned that two of the evaluators found the same problem from another area in the system.

Category 3

Category 3, User control and freedom, had only five discoveries from the evaluators and three of these discoveries were referring to a single unique problem. This specific problem is a functionality where selecting a message in the inbox of the messaging system might take the user to the profile of the message's sender instead of the message and thus this is confusing to the user. This selection screen can be seen in Picture 6 below.



Picture 6. Picture of *Wilma's* inbox. (Visma 2019)
Permission to use the picture given (Kenttälä, 2019, personal communication via email, <1.7.>)

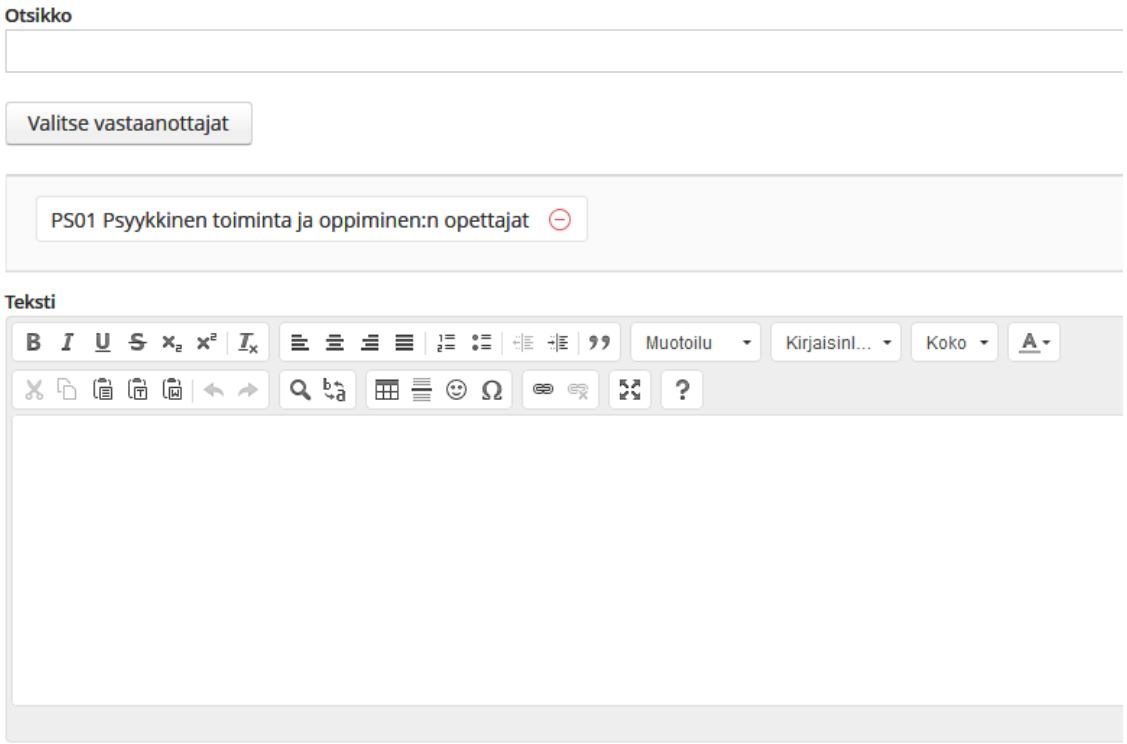
This specific problem was ranked as level 1 or a 2, depending on the evaluator, and is thus a relatively minor problem, although it does give clues of a more significant problem in the whole software, especially when considering the pervious findings and their themes. The two other discoveries were an insight into how the navigation is difficult and a similar note on the general complexity of the program. In a way, all of these issues are referring to already established themes as there are many non-standard functionalities and design choices in functions that are otherwise generally very standardized. It is unusual to find several problems in simple search functionalities or situation where three different evaluators with expert or above user status get confused and make navigation errors when selecting a message.

Interestingly, the general insights found in this category, namely the navigation is difficulty and general complexity, are also two of the major causes of negative emotions and refer to the same basic problem: the user has to use the system itself instead of use the system to get a task done (Preece, Rogers & Sharp 2002: 147) and in fact, category 3c heuristic, or otherwise "Using the system is pleasant, interesting and furthers the users goals", has been partially designed from this affective detail in software. Otherwise this category had very little issues and most likely this refers to how the non-expert evaluators were unfamiliar with what user-control and freedom in software and UX exactly contains. The problems that were discovered were however significant from the general perspective as they gave strong clues to the nature of the overall problem.

Category 4

Category 4, System functionality, had the third most issues in the overall evaluation and, as expected, they mostly focused on details concerning functionalities. Most notable in this category are issues concerning basic usability related subjects including an insight, which was found by no less than three of the evaluators, where many functionalities in the program need to be remembered, even after you have done the task several times. This alone is against Nielsen's (1995) 6th basic heuristic, Recognition rather than recall, where he states specifically that actions and objects need to be recognized, not remembered. There are other similar insights as well including a point

about the unnecessary complexity of the software and vague functionalities. An example of these vague functionalities, could be found for instance in a situation, where the evaluator tried to send a message to a teacher via a course screen, but the eventual recipient was shown in the message to be “Teachers of course PS01” (in the picture in Finnish: “Opettajat”), seen in Picture 7 below, and the evaluator became confused due to wanting only to send a message to a single teacher, not several teachers. Incidentally, both of the Guardian evaluators made this mistake.



Picture 7. Picture of the message screen, with an ambiguous recipient (Visma 2019) Permission to use the picture given (Kenttälä, 2019, personal communication via email, <1.7.>)

There were also several detailed problems found including a few feedback problems and two peculiar issues mentioned by two different evaluators that pointed towards functionalities that they could not find or thought did not exist, even if they in fact were available in the software. These would suggest that there is an underlying problem with overall freedom and control in the system, as other similar “missing” features were complained about in other categories. Overall, in this category, the lack of consistent

and planned user experience design begins to show: there are several different basic usability related discoveries along with detailed discoveries in functionalities that could have been corrected with end-user feedback. Finally, all severity ratings were of level 2 or 3 and thus they can be considered significant.

7.3 Conclusions and Themes

The purpose of the heuristic evaluation was to make assessments if there are any aspects that could cause negative emotions in a user. The results suggest that there are several such aspects and some of these affective factors are even clearly pointed out by contemporary researchers to likely cause frustration and aggravation in a user. These include the general complexity of the system, lack of fulfillment of user expectation, lack of freedom and control and requirement to recall instead than recognize processes (Nielsen 1995, Preece, Rogers & Sharp 2002: 147). Thematically the heuristic evaluation show primarily significant information design related problems, especially in how the information is presented to the user, but there are also different function related issues as well, both in single functions and in general design. These problems include non-standard design in features that are relatively standardized in modern systems, such as search functionalities and e-mail, feedback problems in certain functions, again in search functionalities, and certain expected features. There also seems to be a general issue with the lack of control and freedom, as there are several settings and features in Wilma which do exist, but are simply left invisible to the user, causing significant frustration to the user.

The heuristic evaluation was, eventually, not perfect and during the interviews there were a few detail subjects that were somewhat left ambiguous where they would go in the heuristic list. One such subject was completely missing features, as in several cases they were placed in categories that were less than optimal for their placement. Similarly, there was a discrepancy between the different user groups, as the teachers had significantly higher expertise with the system compared to the other groups. Due to this the teacher’s answers were at the same time more generally inclined and, at the same

time, focused on specific details they encounter in their work. Normally, this kind of combination of personas with wildly differing expertise and requirements would be a significant problem for the entire heuristic evaluation as the results between the different groups would need to be evaluated separately. In this case, as the idea of the heuristic evaluation is to assess the presence of aspects affecting emotions negatively, this does not present a problem as the teachers only express any shortcoming in Wilma from their own point of view.

Overall the heuristic evaluation, even with a few shortcomings in the list of heuristics, was sufficient for the task and clearly pointed out that the software does have aspects that affect the user negatively, namely frustrates and confuses. In fact, almost all of the evaluators became seemingly aggravated during the evaluation, especially when trying to find necessary information regarding the teachers in the system and on the last task, which was specifically designed to be vague and difficult, the frustration of the evaluators was clearly noted. In fact, in one case, the evaluator found the correct answer to the task by pure chance. Only one of the teachers remained calm during the tests, although even he got confused in a few instances, while all of the other four evaluators were clearly aggravated to some degree. From a completely usability point of view, the evaluation was also a success as it demonstrated that there are clear problems in information design and management, overall usability and in several functionalities. This was eventually an expected result, considering that Wilma has not had any kind of planned UX or UI design during the first ten years of development. In a massive and complex software such as Wilma, or any SIS system, this inevitably creates problems and in this case it is seen as hidden functionalities for the user, fragmented information placement, overly complicated features or, inversely, insufficient features and the need to learn the software over and over again. Unfortunately, this causes change resistance, a detail also mentioned by one of the teachers, towards an otherwise powerful and useful tool that is in a sense, necessary in a modern educational world.

8 AFFECTIVE ASPECTS INTERVIEW

The second part of the research was an interview which was designed to find affective design related details and insights of the target material, Wilma. In other words, when the heuristic evaluation was designed to discover if the material had details that have the potential to cause negative emotions, affect the user, the interview is designed to discover if the evaluators were negatively affected by the target material. This was achieved by asking questions concerning general thoughts, emotions concerning and specific details, positive and negative, about the software. The entire list of questions can be found in the Appendix section, both in English and Finnish.

8.1 Thoughts of the Test

The first set of questions in the questionnaire, were again very simple questions about the test itself. These questions were designed to be simple to approach, warm up questions that ease the interviewee in to the second part of the evaluation. The questions were to give the system a general grade of usability, along with possible comments, and to give insights into the evaluation method. These questions function as a way to assess if the heuristic evaluation was a working solution in their point of view to this type of research. As mentioned earlier, the evaluators were non-professionals and if in their liking the entire method is nonsensical, the validity of the entire method can come into question. For the purposes of this research, this would have no such implications, as it would be conducted despite dislike for the method, but it would work as a reference point for further research.

The first question was to grade the usability, using the Finnish lower grade school grading method between 4 being a fail and 10 being full score, and the results were surprisingly even, with everyone giving a grade of 7, except one evaluator giving it a grade of 6,5 totaling at a median of 6,9. Eventually, all of the evaluators considered the usability to be mediocre at best. This somewhat relates to the results of the heuristic

evaluation, although it could have been expected to be less. Most likely, the non-expert nature of the evaluators had an impact here, as they are perhaps not quite as familiar with what constitutes good or bad usability to that matter. None the less, the results of the overall grading match the overall findings. Also, as the first question in the questionnaire, it worked admirably as a warm-up question. It is important not to overly tax interview subjects or at the very least, incrementally increase the difficulty and complexity of the subjects.

The second question was to give thoughts concerning the heuristic evaluation itself. Here there was a notable difference in answers between the three different persona categories as the guardians and student thought that the heuristic evaluation was eventually a sensible way to investigate the software, although they expressed dislike to certain details, such as the aggravating nature of the last part of the test. The teachers on the other hand specifically expressed how they liked how the test showed new ways to operate the system and how the test was designed in general. Overall, all of the evaluators thought that the heuristic evaluation was meaningful way to test a software like this and considering the results of the evaluation, with the results following expected patterns and with the discoveries being accurate, it can be said, with certain certainty that this type of non-professional, assisted, heuristic evaluation works as a method to assess software from an end-user point of view.

8.2 Positive, Negative and Improvements

The first two questions in the questionnaire that introduces affective design in the research are two questions focusing on negative and positive aspects of the material. The first of these questions asks what the evaluator considered annoying, confusing or if there was something otherwise wrong in the system. The answers varied between the evaluators, although they all talked about specific details that affected their life the most when using the system. For instance, the general complexity of the system and the lack of organization options were mentioned specifically to be a problem for the guardians, the single student evaluator mentioned the illogical functions of the messaging system

and the teachers mentioned details about the information management system and the shortcomings it has. Overall, these details followed quite specifically the discoveries in the heuristic evaluation, although both of the teachers discussed specific details they encounter in their work. The second teacher evaluator, simply named Teacher 2, summed up this accurately[2]:

> The program has several frustrating in-functionalities and functionalities that require too long routes to enter. The chaffy nature frustrates. Heavy. The software feels like it was built simple and then expanded all too big. It is the user's responsibility to find all the parts.

Overall, the evaluators summed up the main thematic issue they thought was in their personal point of view the most negative and affected them the most. The precise results varied somewhat, such as aggravation and confusion, but one specific theme began to show: heaviness, which is also mentioned in the excerpt above. More than one evaluator pointed out here, how everything works in its own way and how you must specifically work the system to gain the desired outcome and this directly links to insufficient usability design as this is specifically mentioned by Nielsen (1995) to be a core requirement in all systems to be recognized, not memorized.

The second affective design related question was to describe what they like in the system. Much like the previous question, the evaluators describe details that affected them the most in a positive fashion using the system. However, in this question the difference between persona categories was more prominent, as the guardians and student describe very briefly how they appreciate the function of the system with the ease of communication, information on on-going events and general functions of the system as a tool. The teachers on the other hand go into much further detail how Wilma helps their daily work, for instance, with long term surveillance, communication with parents, as an aide in decision making and into some specific smaller details. The other teacher evaluator, Teacher 1, describes the situation as follows:

[2] All excerpts are available in their original Finnish form as Appendix 8

> Looks nice, works fast, no silly phrases. Communication is much easier than in other forms, especially compared to not having the system at all. Communication to students is very fast. In certain instances it is possible to have the insight from several teachers and making long term assessments and evaluations is easier. Especially when making yearly assessments.

This insight from Teacher 1 basically sums up nearly everything all of the evaluators mentioned as positive: easy to approach layout and visual representation; communication capabilities; and different functional possibilities. Especially the main functionalities and communication gained high praise from nearly all evaluators, notably for different reasons, as the guardians enjoyed how they can see more about their children's education, the student enjoyed the ease of communication and the teachers enjoyed basically all the different aspects the systems brings. A notable detail about the positive answers, and in part positive feelings, is the lack of it. This is seen in every other answer except in the answers from both of the Teachers, who work closely with the program and appreciate the different abilities it brings, but even their answers were at best only half as long compared to the negative aspects and the teachers mention no details, only general functions and possibilities the system enables. The other user categories answer only with a few lines and even then the answers are limited to describing general functions.

The first two affective design related questions in the questionnaire were designed to be simple, but they did give significant insight into the system otherwise. The negative aspects were very similar to the ones found in the heuristic evaluation and all of the evaluators expressed how these details aggravate them and how they need to use too much time to solve these issues. Already from the first question some slight conclusions can be made that the different negative aspects in the software do affect them and it can be tangibly seen in how much time they use on navigating the quirks of the system. Incidentally, the question about the positive aspect revealed how much the evaluators on the other hand like the function of the system as they do understand that the system offers several different tools to follow their children, or their own, studies and contact the school. Functions that otherwise require significant amounts of manual work. Similarly, the teachers liked the system for exactly the same reasons, albeit from their

very different point of view, along with the possibility to follow-up for students on a longer time. Yet again, the teachers similarly expressed as a notable negative aspect how much extra time they use during a day on the system writing notes, placing homework and other details. They do however point out that they would probably need to this anyway, although in a different way. Curiously, both of the teachers mentioned how Wilma has changed the educational field otherwise as well, stating that students might, for instance, point out that homework was not mentioned in Wilma and that is why they did not do it. In a sense, Wilma has become a sort of an authority that demands actions and time from both the students and teachers, and the teachers did not like this.

After the negative and positive details, a natural follow-up was to ask what should be improved upon or what was missing. Here the answers diverged notably between the evaluators as the guardians mentioned the "quirks" of the software and how you need to work around them, but also mention that much of the quality of the software is dependent on how the teachers use it. They also mention the lack of general feedback options to the school. The student was slightly on the same lines on this question and mentions the problems with information management. The teachers yet again go into more details in their answers, mentioning the lack of attachments on messages but also the lack of personal customizations on the software. Guardian 2 had a more personal, albeit specific detail about the system:

> The teacher's use of Wilma could sometimes be improved upon, especially with communications to the parents. Giving feedback from the school's use of Wilma could also be something to develop, since it seems that certain details are overlooked.

This is a very specific detail about the incorrect use of the system by the teachers and how there should be a system to provide feedback is in itself very relevant, as the system does not have such a possibility at the moment, at best a person can give feedback about the system. This however also shows how using the system is not always easy from the teachers point of view and reminds us that any system is only as effective as the people using it.

Overall these three questions follow the lines established by the heuristic evaluation, although by making generalizations of the software. The negative details mentioned are almost exactly mirrored in the heuristic evaluation, although often from a more general point of view. Similarly, the missing or improvable aspects follow these lines with a specific personal touch: the teachers wish more customizations or features that ease their work; the guardians want improved communication and generally easier functions to the software; and the student wished for improved information gathering and management options. In a sense, the aspects that have been confirmed to cause negative emotions and frustration in the system would be desired to be removed.

The positive aspect question on the other hand brought new information as all of the evaluators honestly liked the general function, the purpose, of the software. All of the user categories, guardians, students and teachers alike, appreciated and valued the ease of communication from the sense that you do not need to remember phone numbers or email addresses, even if using the communication system in general is considered tedious. Guardians and the student on the other hand enjoyed the ability to follow studies more efficiently, be that your own or your child's, with up-to-date homework and information about events. The guardians valued the ability to gain more information about their children's studies, especially when considering that without Wilma, information coming from the school can be limited. The teachers on the other hand liked the many decision assisting features and the ability to make long term follow-up on students which Wilma enables, although yet again they mention specifically the problems with usability as working with Wilma takes considerable amounts of time due to the different issues it has.

From a purely emotional point of view, these two questions show a general theme of negative emotions being somewhat clearly visible while positive emotions are only notable in their absence. Of course, this varies greatly between evaluators, however, the evaluators are much more confused, aggravated or feel daunted by the system, instead of, for instance, feeling a sense of ease from how the system makes their life easier, which it eventually is designed to do. In fact, to gain any kind of positive feedback, the

evaluators must be taken out of the procedural world of the system itself and taken to a level where they consider the overall benefits of the software. Naturally, due to the invisibility of positive aspects phenomenon, this is somewhat expected, as even if the system would work easily and flawlessly, the ease-of-use would not be noticed in itself, but in such a case the aspects that would gain negative feedback, would be significantly smaller issues compared to the general observation that navigation and sending a message is a daunting task.

8.3 Affective Aspects

The affective aspects were mainly concentrated on three questions in the questionnaire along with the Open word and interviewers notes giving extra details. The three questions were: was there something that frustrated you in the system; was the user experience pleasant; and what is the first thing that comes to mind, when thinking about the system? These three questions were specifically created to discover, what the evaluators felt about the system itself and in combination with the details from the other questions an estimate could be made if the user had negative emotions toward the system. The last question of these three, what comes to mind, is the most relevant of these questions in conjunction with the hypothesis. It is basically the end result of the user experience that the users have gained during a longer period.

The first affective question, was there something that frustrated you in the system, gave unusually varying answers from the different evaluators, as some were very detailed and had a multitude of different negative details and problems while some answers were plain and simple. Interestingly, one of the users, a guardian, even stated that she thought that there were no frustrating aspects in the system, even if this specific evaluator was very frustrated in several situations during the evaluation itself and made several specific remarks about the flaws the notification system has, even if it was outside the test. The other answers were more in line with their discoveries during the evaluations and seeming emotions during the evaluation and for instance, Guardian 1 plainly stated that “The communication system was irritating and difficult to handle”. The student

similarly made a remark that “The communication system is irritating and difficult to use as the teachers are hard to find in the system”. The student evaluator also mentioned that she had a problem with Wilma’s mobile app, although this was not a part of this research frame per say. The teachers had again much more to say in this question as well, giving general details that frustrate their daily life such as the general lack of usability and clarity, differences in organizational level of use and the amount of extra work Wilma causes them. Interestingly, both of the teachers mention how Wilma is changing the overall educational field mentioning that students even demand more information in Wilma. They both also end this question by mentioning specifically how making notes about students is difficult and time consuming, although they mention it in different context.

The second question, was the user experience enjoyable, gave diverse answers from the different users, so diverse, that a specific theme is difficult to conclude. The answers were however very short and did not go into much detail. The guardians stated that the user experience is mediocre, although they both add negative details in their answers, the first mentioning that the system is at times agonizing to use and the other mentioning how the users and the organizational practices cause problems. The student on the other hand mentioned specifically how sending a message to a teacher is easy and simplified by the system, but sending the actual message is complicated and difficult due to the complicated interface, stating basically the already mentioned problem where you have to use the system and not the function it provides. Excerpt below:

> It is easy to send a message to a teacher with the system, for instance for not being able to attend, but certain details, such as the 12.00 time limit to inform about not attending, make matters very complicated, especially as an evening student. The system rarely functions as you would expect it to function, instead it has a completely own way of doing things, which is hard to remember.

Interestingly, both of the teachers, while normally very vocal and detailed in their thoughts of the system, were in this case almost tight-lipped. The first simply says that user experience is basic and the other directly states it is not good and it is heavy, adding that it is not designed with the user’s needs in mind.

These two questions, as said, were designed specifically to discover if the users were affected negatively by the software. While the answers were diverse and subjective, as expected, they do give somewhat clear results that the users were negatively affected by the software, although to different degrees. This combined with the notes from the interviewer, not written but seen during the interview, where more than half of the evaluators were notably aggravated during the evaluations, support this outcome. Similarly, the open word section at the end of the interview further supports this: all of the evaluators, in one way or another state that the software is difficult to use, but is necessary in the modern world. Obviously, to what degree they were negatively affected and in what particular way remain ambiguous due to the subjective nature of the user experience.

Most notable details supporting the fact that they were negatively affected, at least to some degree, eventually cannot be seen from anything the evaluators exactly pointed out but is more seen as a general theme in the entire material, which could already be seen from the basic questions of the system: the lack of positive feedback. The users do understand the function of Wilma, they appreciate the functionality the software provides and they wish to continue using it in the future due to this. However, when we go in any direction after this base functionality, we almost unanimously encounter nothing but negative feedback and negative emotions: aggravation, confusion and heaviness. This can be best seen in the affective design questions: the negative feedback is three to four times more expansive than the positive feedback and the positive feedback almost unanimously contains a negative detail. Of course, due to the invisibility of positive aspects, this is somewhat to be expected, but even after this, the evaluators seemingly struggled to find a positive detail to mention, other than the function. Finally, all of the evaluators mentioned specific personal negative details that aggravated them, including email notifications or issues in class notes, which were never mentioned by other evaluators, implicating that they were affected on a personal level by dysfunctional features that were carried over into the heuristic evaluation and interview. While there is no exact place where a specific answer could be pointed out, there is however sufficient overall evidence to support the result that the evaluators

were negatively affected by the software as negative emotions are clearly much more prevalent than positive feedback.

8.4 Effects of the Negative Aspects

The last question in the questionnaire was already mentioned in the previous sub-chapter: what is the first thing that comes to mind, when thinking about the system? This single question is the most important question in the questionnaire as it basically links the material to the hypothesis of this thesis and the research questions. The idea behind this question is to test what is the intuitive feeling considering Wilma and while it is by no means exact or accurate on an emotional level it indicates the base feeling towards the system: is it negative or positive? Does it have a specific connotation? Does it have a specific theme? Is it linked to something specific? The answer is always very subjective, but it is an honest feeling towards a subject and it can be evaluated and assessed. As a simple example: consider Google and the first thing that comes in mind. The answer is subjective, has varying thoughts and connotations, but it at least in the answerers mind is relatively clear and tells something about Google.

The answers themselves for this question were invariably negative, although they had variance in subject and severity. The first guardian stated that she felt simple dislike towards the system and she would not like to use it. The second guardian simply said she gets aggravated. The sole student similarly explained she has this emotion that she does not want to do it, it feels heavy. The teachers followed the same line: the first teachers simply said “Work” and the second said “Aggravation”. The second teacher did mention that he appreciates the usefulness, but the software simply repulses him.

The answers were, as requested in the question, short, but in conjunction with everything else done during the heuristic evaluation and affective interview, extremely accurate. They also support all of the findings in the previous steps of the research and even in a Persona level of the different evaluators condones the results. The answers did have some variance, however two answers were simple notions of negative emotions of

dislike, two answers were plainly aggravation and one answer stated work. In a sense the last answer of Wilma is also the clearest and most honest, although it perhaps scientifically is not accurate. However, as a general theme the answers do indicate the following things: the software feels heavy, it is complicated to use and if a user has to use it on his or her free time, they would rather not do it.

Finally, the most prevalent emotions encountered during the questionnaire, were more or less negative or neutral, while nearly all emotions encountered during the evaluation itself was invariably negative, if any were shown. For instance, Teacher 1, who was perhaps most familiar with using the system, stating that he had experience working with Wilma in a large school, small school and without Wilma altogether at a third school, became confused and aggravated in a few situations during the heuristic evaluation but afterwards, during the interview, he was mainly neutral and unaffected about everything. In fact, Teacher 1 was so accustomed to using Wilma, that he barely found any issues, outside the usual navigation and communication difficulties, and even he had eventually very little good to say, stating basically very neutral details and his eventual feeling was “Work” when thinking about the system. All of the other evaluators were much more vocal about problems during the heuristic evaluation and during the interview and were incidentally much more negatively affected as well. Eventually, the results speak for themselves as it was relatively clearly pointed out by the heuristic evaluation that there were aspects that can cause negative emotions in a user and comparing these results to the results of the interview, where there are negative emotions to be seen, most notably aggravation, confusion and heaviness, a conclusion can be made that the evaluators were negatively affected, to a subjective, varying degree.

These results combined with the previous findings, where the evaluators were at least to some degree negatively affected by the negative aspects of the target software there is an answer to the overall question of this thesis: how do negative aspects in user experience design affect the user? In short, it makes the user not want to use the software. In a way this was an expected result, however in the case of Wilma, where software’s functionality is appreciated and the users understand the usefulness and need

of the software and, thus they want to use it, the answer is more complicated. The users continue to use it, but it feels heavy and it feels like work. Work they do not wish to do in their free time. Eventually, the best overall answer in the case of Wilma was stated by one of the teacher evaluators: the complex and difficult to approach nature of Wilma creates unnecessary change resistance in the user base. Change resistance that could be avoided with sufficient user experience design.

9. DISCUSSION

The goal of this research was to study the relationship between aspects of negative user experience and the emotions they create, along with what their effect eventually is towards the system. To achieve the research goal, a set of research questions were created: what kind of aspects generate negative emotions in a user interfaces; how do these negative aspects affect the user experience of the system; and what are the effects of these negative emotions towards the use of the system as a whole. In the research a two part usability study was conducted: a participatory heuristic evaluation and an interview. For the use of the heuristic evaluation, a list of heuristics was created based upon contemporary usability research and the theoretical user experience framework. The heuristic list had five base heuristics and 15 more accurate heuristics to aid the non-professional evaluators in the task. The heuristic evaluation was done by five UX/UI non-professionals, who were considered to be advanced or better users of the target material, the communication module of Wilma, a school information system used widely in Finland, representing all three of the user categories of the software, namely teachers, students and guardians of students. Finally, a hypothesis was established, which stated that in the system in question for this thesis, some causes for frustration and aggravation can be found and they will have a negative effect on the overall user experience of the system, albeit exact results may vary.

9.1 Related research

The heuristic evaluation was designed to evaluate the material itself and to make conclusions if the material had negative aspects that have the potential to affect the user. The evaluators found 53 different issues, out of which 32 were unique. The results from the heuristic evaluation were eventually quite clear, as all evaluators found a number of different aspects that were in their liking negative on a personal level, but were also confirmed to be negative in comparison with existing literature. These aspects included, for instance, several usability aspects Nielsen (1995) states are requirements for good

user experience, but are lacking, such as Consistency and standard, Recognition rather than recall and user control and freedom. Similarly, aspects that are known to affect the user negatively were also found, such as the general issue in complexity, where the user has to use the system itself instead of using the system to get a task done (Preece, Rogers & Sharp 2002: 147-152). Eventually, while there were other discoveries as well, these were the most prominent issues affecting the system and the ones considered to most cause negative emotions.

The second part of the research was an interview designed to evaluate if the five evaluators were emotionally negatively affected. The interview consisted of nine questions designed to assess the thoughts and emotions of the evaluators towards the target system. The results of the emotional interview were somewhat less clear than the results of the heuristic evaluation, which was to be expected, as emotions are first of all, subjective, meaning that they can vary significantly from person to person and can even be situation dependent, but secondly, affecting emotions, due to the same reasons, is complicated and the results are uncertain, although certain patterns, actions and aspects are known to produce somewhat expectable results (Norman 2004: 29-30). However, most of the evaluators were negatively affected by the software, stating even during the heuristic evaluation that they felt confused, aggravated or frustrated due to the complicated nature of the target material. These are mentioned to be the most known emotional results of insufficient UX/UI design by contemporary literature (Preece, Rogers & Sharp 2002: 147, Cooper, Reimann & Cronin 2007: 152). The questions themselves yielded similar results, as many of the evaluators mentioned several of the same details, which are known to cause confusion, frustration and aggravation, along with specific personal details that caused negative emotions. It should be mentioned, that one of the evaluators, an expert level teacher, was nearly completely un-phased by the material and showed only once confusion during the evaluation and during the interview basically stated details that did not work in the system. This worked as an example on how not everyone reacts in the same way toward the same affective stimuli.

Eventually, the established hypothesis described the situation relatively well: the material had several negative user experience aspects that caused negative feelings,

albeit the subjective nature of emotions was notable in the results. Eventually, these negative emotions manifested in the users as a feeling of difficulty and heaviness, as a desire not to use the system. This is, in the case of the material, an unfortunate result, as the evaluated system is a School Information System designed to assist the communication and flow of information between the school and people connected to the school, creating a situation where the users of the system want to use the system, but due to the negative user experience, the system simply feels like work they have to do in their free time. This also eventually reflects negatively on Wilma, and indirectly to the creator Visma, as a brand, as users easily respond emotionally towards products and brands (Cooper et al. 2007: 89). The feeling of heaviness or dislike is not the emotion you wish your product or brand to generate and thus the results are less than optimal.

Considering the overall method and results of this research, where a basic usability evaluation method was combined with an emotional evaluation method, the results were solid and work as a reference point for further research. The research demonstrates that Affective design, which is a newer theoretical framework in User Experience Design, works efficiently in amending certain short-comings inherent in several Usability evaluation methods, for instance in the case of this thesis, Jakob Nielsen's (1995) Heuristic Evaluation was used, which is effective at finding single detailed results, but struggles in finding they key reasons behind the issues. The Affective Design based questionnaire used as a secondary method in this thesis efficiently combined the results from the Heuristic evaluation and provided excellent results and insights into the key challenges and issues in the target material.

In a sense, affective design worked as an opposite to Nielsen's Heuristic evaluation as emotion based approach to User Experience evaluation worked best at discovering general, macro level thoughts from users, while Heuristic evaluation has its strengths in discovering specific details. Together these could be used to assess clearer overall themes and aspects in the material, such as the detail discovered during the research, where the evaluators complained about different features being complicated and confusing, especially search functionalities, turned out to be a memorization issue after conducting the interview, as much of the information in the system is placed in a

fragmented way and to find certain information, you have to do significant navigation and specifically remember details yourself. This type of detail is difficult to discern from simply doing a heuristic evaluation.

9.2 Limitations

The emotions themselves are, however, a very difficult subject, due to their subjective nature and even daily events playing their part in the equation, and this is perhaps the area where more research could be done in relation to the discoveries of this thesis. Heuristic evaluation itself is an old and well known method, but the affective design based ideas are new in the UX sphere or research. Psychology itself is not new to UX but exact cause and effect related questions have not been studied much in conjunction with UI and UX and this thesis is somewhat lacking in this sense as well. No significant or deep psychological background was added to this thesis, mainly a method to recognize basic emotions was used, and the exact effects of the different emotions were similarly not studied. Such a research would be well beyond the scope of a thesis and the method used in this thesis, where results using a proven method are combined with a new method based on basic emotions to create base level results, is sufficient. For further research, this would clearly be the strongest suggestion with detailed research, for instance, into what kind of emotions certain features or functions create is a possibility; more detailed research into how to affect emotions through user experience; and how mechanical, physical design affect emotions or inversely, how affective design could change how physical objects could be designed.

The research was eventually not without its flaws: the list of heuristics used in the research, which was designed to combine some affective details with established usability heuristics, proved be less than perfect, and as a result, certain discoveries were difficult to place. Similarly, the questionnaire had some flaws as well, with overlaps and ambiguities, and thus certain answers were not always as useful as desired. If a similar research was conducted again, these would have to be addressed accordingly. If a certain bias should be mentioned concerning this thesis, it could perhaps be that some

questions are perhaps leading and the entire research eventually expects that the material has negative details. Wilma was specifically chosen as a material for this thesis due to the background, where no realistic UX design was done for 10 years of development and this combined with the massive size of the software creates a situation where user experience is bound to have issues, more than less. The question is, would this same research be viable, with a material that has better overall UX design? Most likely yes, the results would be more likely less precise as they were now, but there would most likely be some results, especially as emotional aspects are combined with the evaluation.

Finally, the validity, reliability and truthfulness of the results, research method and eventual conclusions should be evaluated, as while the answers from the interviewed test subjects are most likely truthful, their reliability and validity is in slight question as the subject is eventually emotions. Emotions can be very different already between two different days as such a small things as bad weather can have an effect. However, as the subject is a heuristic evaluation and an emotional interview, the eventual results are most likely not significantly affected. The eventual research and conclusions from the research should also be evaluated, as they were still done by only one researcher.

9.3 Conclusions

According to the research, negative user experience creates subjective, negative feelings, however, how this eventually reflects on the system, varies from person to person, although the result is towards negative. While as a concept this is simple and straightforward, the context and purpose of the system creates notable variance in exact results. In the case of this thesis, the evaluated system was a School Information System that is designed to assist the communication and flow of information between the school and people attending the school, creating a situation where the users of the system want to use the system, but due to the negative user experience, they feel the use of the system is hard and difficult; it feels like work they have to do on their free time.

Considering the overall method and results of this research, where a basic usability evaluation method was combined with an emotional evaluation method, the results were solid and work as a reference point for further research. The research demonstrates that Affective design, which is a newer theoretical framework in User Experience Design, works efficiently in amending certain short-comings inherent in several Usability evaluation methods, for instance in the case of this thesis Jakob Nielsen's (1995) Heuristic Evaluation was used, which is effective at finding single detailed results, but struggles in combining these results into a cohesive whole. The Affective Design based questionnaire worked and was efficient in discovering general, macro level themes and issues in the system and combined with the results of the Heuristic evaluation, provided excellent overall results and insights into the key challenges and issues in the target material, Wilma.

The research was however not without its flaws: the list of heuristics used in the research, which was designed to combine some affective details with established usability heuristics, proved be less than perfect, as certain results were difficult to place, for instance. Similarly, the questionnaire had some flaws as well, with overlaps and ambiguities, but in eventuality, the results were sufficient enough to provide acceptable answers: Affective Design, or in other words, psychological user details, can be effectively used to evaluate Usability and in fact it can be used to bridge technical results more efficiently with user based results. Based from this, the possibilities for further research are significant as psychological based usability research is yet a rather new field of study. For instance, detailed research into what kind of emotions certain features or functions create is a possibility; more detailed research into how to affect emotions through user experience; and how mechanical, physical design affect emotions or inversely, how affective design could change how physical objects could be designed.

SOURCES

Akass, Clive (2001). *The men who really invented the GUI.* Computer Active. [online]. [14.11.2018]. Available at: https://web.archive.org/web/20110816031619/http://www.computeractive.co.uk/pcw/pc-help/1925325/the-invented-gui

Boykin, George (2017). *The History of Management Information Systems* [online]. [3.11.2018]. Available at: https://bizfluent.com/about-5444925-history-management-information-systems.html

Cooper, Alan, Reimann, Robert & David Cronin (2007). *About Face 3: The Essentials of Interaction Design.* Wiley Publishing Inc. Indianapolis, U.S.A.

E A (2018). *error-operation-completed-succesfully*, <https://www.flickr.com/photos/thecomputerdude/5501251932>. Cited: 11.10.2018

Electrolux (2018). *Coffee maker EKAM200*, <https://www.electrolux.fi/support/user-manuals/?q=ekam200>. Cited: 4.10.2018

Ericsson, K. Anders & Herbert A. Simon (1980). Verbal Reports as Data. *Psychological Review.* 87:3. American Psychological Association. Washington D.C., U.S.A. [online]. [11.12.2018]. Available at: http://www.psy.cmu.edu/~siegler/Ericsson-Simon80.pdf

Friesen, Jeff (2013). *Design patterns, the big picture, Part 3: Beyond software design patterns.* JavaWorld. [online]. [11.11.2018]. Available at: https://www.javaworld.com/article/2071208/core-java/design-patterns--the-big-picture--part-3--beyond-software-design-patterns.html

Gaffney, Gerry (2004). *Usability for technical communicators*. Information & Design. Melbourne, Australia. [online]. [10.12.2018]. Available at: https://infodesign.com.au/usabilityresources/articles/technicalcommunicators/

Goodwin, Kim (2009). *Designing for the Digital Age: How to Create Human-Centered Products and Services.* Wiley Publishing, Inc. Indianapolis, U.S.A.

Haapsaari, Janne (2018). Kysymyksiä graduun. Sähköpostiviesti Janne Haapsaarelle 16.11.2018.

Interaction Design Foundation (2018). *User Experience (UX) Design* < https://www.interaction-design.org/literature/topics/ux-design>. Cited: 18.5.2019

ISO 9241 (2016). Ergonomics of human-system interaction, International Organization for Standardization, Switzerland.

Korvenranta, Heli (2005). Asiantuntija-arvioinnit. In: *Käytettävyystutkimuksen menetelmät.* Edited by Ovaska, Aula & Majaranta, 2005, 111–124. Tampere, Finland.

Kumar, Kamal S. (2000). *Higher Education: A Lesson in Mainframe Server Automation* [online]. [3.11.2018]. Available at: https://web.archive.org/web/20061111130412/http://www.esj.com/article.aspx?ID=380045556PM

Kuosmanen, Ilkka (2018). "Varokaa heikkoa jäätä, mmmm" - Pikku Kakkosen nallen lausahdus traumatisoi lapseni. *Etelä-Suomen Sanomat.* [online] [21.12.2018]. Available at: https://www.ess.fi/Mielipide/esalaiset/art2452380

Kuutti, Wille (2003). *Käytettävyys, suunnittelu ja arviointi.* Talentum. Helsinki, Finland.

Larkin, Andrew (2018). *Disadvantages of Cloud Computing* [online]. [6.11.2018]. Available at: https://cloudacademy.com/blog/disadvantages-of-cloud-computing/

Laudon, Kenneth C. & Jane P. Laudon (2014) *Management Information Systems - Managing the Digital Firm.* 13th edition. London. Pearson Education, Inc.

Leveson, Nancy & Clarck Turner (1993). An Investigation of the Therac-25 Accidents. *Computer.* [online]. [7.10.2018]. Available at: https://web.stanford.edu/class/cs240/old/sp2014/readings/therac-25.pdf

Lewis, Clayton (1982). *Using the "Thinking Aloud" Method In Cognitive Interface Design* IBM. Yorktown Heights, U.S.A.

Lockner, Damiens & Nathalie Bonnardel (2014). Emotion and Interface Design. How to measure interface design emotional effect?. *International Conference on Kansei Engineering and Emotion research*. Linköping. [online]. [20.10.2018]. Available at: https://www.researchgate.net/publication/278963569_Emotion_and_Interface_Design_How_to_measure_interface_design_emotional_effect

Löwgren, Jonas & Erik Stolterman (2004). *Thoughtful Interaction Design.* MIT Press. Massachusetts, U.S.A.

MacKenzie, Deb (2014). *What's the Difference Between UX and UI Design?*. Center for Digital Media. Vancouver, Canada. [online]. [9.11.2018]. Available at: https://thecdm.ca/news/whats-the-difference-between-ux-and-ui-design

March, Artemis (1993). Usability: The New Dimension of Product Design. *Harvard Business review.* September – October 1994. Harvard Business Publishing. Massachusetts, U.S.A. [online]. [10.12.2018]. Available at: https://hbr.org/1994/09/usability-the-new-dimension-of-product-design

Medlock, Michael C., Wixon, Dennis, McGee, Mick & Dan Welsh (2005). The Rapid Iterative Test and Evaluation Method: Better Products in Less Time. In: *Cost-Justifying Usability.* 489-497. Eds: Randolph G. Bias & Deborah J. Mayhew. Morgan Kauffmann Publishers. San Francisco, U.S.A.

Merholz, Peter (2007). *Peter in Conversation with Don Norman About UX & Innovation.* Adaptive Path. [online]. [14.11.2018]. Available at: http://www.adaptivepath.org/ideas/e000862/

Mickwitz, Camilla 1986. *Varokaa heikkoa jäätä,* YLE. [online] [21.12.2018]. Available at: https://yle.fi/aihe/artikkeli/2007/03/12/pikku-kakkosen-jaavaroitus

Moggridge, Bill (2006). *Designing Interactions.* MIT Press. Massachusetts, U.S.A.

Muller, Michael J., Matheson, Lisa, Page, Colleen & Robert Gallup (1998). Methods & tools: participatory heuristic evaluation. *interactions.* 5: 5, 13-18, ACM New York, New York, U.S.A.

Nielsen, Jakob & Rolf Molich (1990). Heuristic Evaluation of User Interfaces. *CHI 90 Proceedings.* ACM. U.S.A.

Nielsen, Jakob (1993). *Usability Engineering*. Academic Press, Boston, USA.

Nielsen, Jakob (1994). Enhancing the explanatory power of usability heuristics. *Proceedings of Human Factors in Computing Systems*, ACM Press, 152–158.

Nielsen, Jakob (1995). *10 Usability Heuristics for User Interface Design.* Nielsen Norman Group. [online]. [10.1.2019]. Available at: https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, Jakob (2012). *Usability 101: Introduction to Usability* [online]. [21.10.2018]. Available at: https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Norman, Donald A. (2004). *Emotional design: why we love (or hate) everyday things.* Basic Books. New York, U.S.A.

OpenEMIS (2018). *What is User Experience (UX) Design?*. Interaction Design Foundation (IDF). Denmark. [online]. [13.11.2018]. Available at: https://www.interaction-design.org/literature/topics/ux-design/

Piernik, Malgorzata (2017). *8 ways to reduce cognitive load: Part 2*. UX Planet. [online] [12.1.2019]. Available at: https://uxplanet.org/8-ways-to-reduce-cognitive-load-part-2-4b0f9d8ef5ad

Plutchik, Robert. (2002) *Emotions and Life: Perspectives from Psychology, Biology, and Evolution,* 1st edition. American Psychological Association. Washington, U.S.A.

Preece, Jennifer, Yvonne Rogers & Helen Sharp (2002). *Interaction design : beyond human- computer interaction.* John Wiley & Sons, Inc. U.S.A.

Psomas, Steve (2007). *The Five Competencies of User Experience Design* UX Matters. [online]. [15.11.2018]. Available at: https://www.uxmatters.com/mt/archives/2007/11/the-five-competencies-of-user-experience-design.php

Saffer, Dan (2018). *Disciplines of User Experience*. [online]. [13.11.2018]. Available at: http://www.kickerstudio.com/2008/12/the-disciplines-of-user-experience/

Salminen, Joni, Kwak, Haewoon, An, Jisun, Jung, Soon-gyo & Bernard Jansen (2018). Are personas done? Evaluating their usefulness in the age of digital analytics. *Persona Studies*. 4, 47–65. [online]. [30.11.2018]. Available at: http://www.kickerstudio.com/blog/2008/12/the-disciplines-of-user-experience/

Sauro, Jeff (2013). *A Brief History of Usability.* MeasuringU. Colorado, U.S.A. [online]. [10.12.2018]. Available at: https://measuringu.com/usability-history/

Schaarupa, Clara, Pape-Haugaarda, Louise, Hangaarda, Stine V., Mihovskab, Albena Hartvigsenc, Gunnar & Ole K. Hejlesena (2015). Using Participatory Heuristic Evaluation as a Collaborative Backbone in Large-Scale Projects – preliminary experience from the eWALL EU-Project. *Proceedings of the 13th Scandinavian Conference on Health Informatics*, 15-17, Tromsø, Norway [online] [12.1.2019]. Available at: http://www.ep.liu.se/ecp/115/004/ecp15115004.pdf

Shneiderman, Ben & Catherine Plaisant (2005). *Designing the User Interface*. 4th edition. London. Pearson Education, Inc.

Sears, Andrew (1998). The Effect of Task Description Detail on Evaluator Performance with Cognitive Walkthroughs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 259–260. Los Angeles, U.S.A.

Silver, Kevin (2007). *What Puts the Design in Interaction Design.* UX Matters. [online]. [29.11.2018]. Available at: https://www.uxmatters.com/mt/archives/2007/07/what-puts-the-design-in-interaction-design.php

Singh, Prakash & Thembinoksi G. Twalo (2014). Effects Of Poorly Implemented Performance Management Systems On The Job Behavior And Performance Of Employees. *International Business & Economics Research Journal.*14:1, 47-60, Clute Institute, Colorado, U.S.A.

Spring/Summer 1992, Picture: *Murder* Concept: Oliviero Toscani Photo: Franco Zecchini/Magnum, United Colors of Benetton

Suojanen, Tytti, Kwak, Koskinen, Kaisa & Tiina Tuominen (2012). *Käyttäjäkeskeinen kääntäminen*. University of Tampere. Tampere, Finland. [online]. [10.12.2018]. Available at: https://tampub.uta.fi/bitstream/handle/10024/66333/kayttajakeskeinen_kaantaminen_2012.pdf?sequence=1

Topi, Heikki, Lucas, Wendy & Tamara Babaian (2005). Idnetifying Usability Issues With an ERP Implementation. *In Proceedings of ICEIS'05* [online]. [29.11.2018]. Available at: http://cis.bentley.edu/tbabaian/papers/ICEIS-05.pdf

Tractinsky, N. (1997) Aesthetics and apparent us- ability: empirically assessing cultural and methodological issues. *In Proceedings of CHI'97*,115-122. ACM Press. Atlanta, U.S.A.

UNESCO (2012). *OpenEMIS.* UNESCO. Paris, France. [online]. [8.11.2018]. Available at: https://www.openemis.org/

UNESCO (2012). *OpenEMIS Infrastructure Diagram.* UNESCO. Paris, France. [online]. [8.11.2018]. Available at: https://www.openemis.org/wp-content/uploads/2018/04/OpenEMIS_Infrastructure_Diagram_en.pdf

Visma (2018). *Tietoa Vismasta.* Visma. Oslo, Norway. [online]. [19.11.2018]. Available at: https://www.visma.fi/tietoa-vismasta/

APPENDICES

APPENDIX 1. Example of a Visceral commercial design (Toscani & Zecchini 1992)

APPENDIX 2. User information sheet, Finnish version

Henkilötietolomake

Nimi:

Ikä:

Käyttäjätyyppi (Huoltaja/Opiskelija/Opettaja):

Käyttö per viikko (oma vapaa arvio):

Yleisarvosana järjestelmästä (4-10):

Yleinen vapaa sana Wilmasta (ennen testiä, muista, antamaasi tietoa käytetään järjestelmän parantamiseen, joten se merkitsee):

APPENDIX 3. User information sheet, English version

User information sheet

Name:

Age:

User type (Student/Guardian/Teacher):

Usage per week (rough estimate):

General rating of the system (4-10):

Free word (Before the test, remember, your information will be used to improve the system):

APPENDIX 4: Test Cases, Finnish version

Tehtävänanto:

Tehtävä 1: Kirjautuminen

- Kirjaudu sisään käyttäen tunnusta: mattiina.ankerman@demo.local
- Salasana: §123qwer
- Navigoi viesteihin.

Tehtävä 2: Viestien lähetys

- Lähetä lyhyt viesti koulun IT-vastaavalle Wilhelmina Airasmaalle. Käytä aiheena ”IT-ongelmat”.
- Poikasi on saanut lapun kotiin kurssin PS01 opettajalta, syy lapulle on mielestäsi aiheeton ja haluat valittaa aiheesta. Lähetä viesti PS01 kurssin opettajalle, sekä poikasi ryhmänohjaajalle.
- Vaihto opettajaksi (haastattelija tekee vaihdon puolestasi).
- Olet saanut viestin opiskelijasi vanhemmalta. Havaitset, että kurssin käytäntöjä pitää muuttaa viestin perusteella. Lähetä aiheesta viesti kurssiesi PS01 ja PS04 oppilaille sekä huoltajille.

Tehtävä 3: Viestiin vastaaminen

- Olet saanut viestin poikasi opettajalta, joka koskee aiemmin keskustelemastasi aiheesta. Vastaa opettajan viestiin ja kiitä häntä.

Tehtävä 4: Vanhojen viestin hakeminen

- Lähetit aiemmin veistin koulun IT-vastaavalle. Nyt muistat saaneesi linkkejä oppilaiden tietoturvasta. Etsi tämä viesti.

APPENDIX 5: Test Cases, English version

Task insctructions:

Task 1: Log in

- Log in using the username: mattiina.ankerman@demo.local
- Password: §123qwer
- Navigate to messages.

Task 2: Sending a message

- Send a short message to the school's IT-manager Wilhelmina Airasmaalle. Use the subject "IT-problems".
- Your son has received a note from the teacher of course PS01. The reason for this note is in your mind unjustified and you wish to complain. Send a message to the teacher of course PS01 and your son's group instructor.
- Change to Teacher user (Done by the interviewer).
- You have received a message from the guardian of one of your students. You notice you need to change the practices in one of your courses according to the message. Send a note about the changes to the guardians and teachers of your courses PS01 and PS04.

Task 3: Answering a message

- You have received a message from your son's teacher, concerning an earlier subject. Answer this message and thank him.

Task 4: Searching old messages

- You sent earlier a message to the schools IT-manager. You now remember recieving a message with links concerning student information security. Find this message.

APPENDIX 6: Affective questionnaire, Finnish version

Arviointilomake: Nimi:

Yleisarvosana käytettävyydestä (4-10):

Mitä pidit testistä (oliko testi järkevä)?

Löysitkö jotain joka ärsytti, oli hämmentävää tai ei sinusta tuntunut oikealta?

Mistä pidit järjestelmässä?

Mitä pitäisi parantaa? Puuttuuko jokin?

Onko järjestelmässä jotain joka ärsyttää erityisesti?

Oliko käyttökokemus miellyttävä?

Mikä mielikuva tai tunne tulee järjestelmästä tai kuulet puhuttavan siitä?

Vapaa sana: (Jälkeen testin, voit kommentoida myös testiä)

APPENDIX 7: Affective questionnaire, English version

Affective questionnaire: Name:

Overall grade of the usability (4-10):

What did you think about the test (did you find it sensible)?

Did you find something that was annoying, confusing or otherwise did not feel right?

What did you like in the system?

What should be improved? Was something missing?

Was there something in the system that particularly aggravated you?

Did you find the user experience enjoyable?

What comes in mind when you think about the system or hear somebody talking about it?

Free word: (After the test, you may also comment the test itself)

APPENDIX 8: Interview excerpts – Original in Finnish

Excerpt 1: Teacher 2

Ohjelmassa on aika paljon turhauttavia toimimattomuuksia, vaikean kautta ja toimintoihin mennään liian pitkiä reittejä. Silppumaisuus turhauttaa paljon. Raskas. Tuntuu kuin se olisi rakennettu yksinkertaiseksi ja sitten laajennettu liian suureksi. Käyttäjän vastuulla on löytää palikat.

Excerpt 2: Teacher 1

Kivan näköinen, toimii nopeasti, ei typeriä fraaseja. Viestiminen on huomattavan paljon helpompaa kuin muilla tavoilla. Etenkin vrt. jos olisi ilman. Oppilaille viestintä on erityisen nopeaa. Tietyissä asioissa, on mahdollista saada usean opettajan näkemys joihinkin asioihin, auttaa tukena päätöksiin. Oppilaiden merkinnät (aktiivisuus ja häiriöt) näkyvät pitkään, etenkin vuosiarvioinnissa auttaa.

Excerpt 3: Guardian 2

Opettajien ohjelman käytössä välillä parannettavaa, etenkin ulos -viestinnän suhteen. Palautteen anto myös koulun Wilman käytöstä olisi hyvä kehittää, sillä välillä tiettyjä asioita ei Wilman käytössä oteta huomioon.

Excerpt 4: Guardian 1 and Student

Guardian 1:

Viestit ovat vaikeita ja järjestelmä on ärsyttävä.

Student:

Viestin lähettely on ärsyttävää ja vaikeaa: opettajia vaikea löytää.

Excerpt 5: Student

Järjestelmässä on helppoa laittaa opettajalle viesti esim. poissaolosta, mutta tietyt yksityiskohdat, kuten kello 12.00 takaraja ilmoittaa poissaolosta, tekevät siitä välillä monimutkaista. Järjestelmä ei monestikaan toimi niin kuin sen odottaisi toimivan, vaan sillä on aivan oma tapa toimia, jota on vaikea muistaa.