TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Jari Varsaluoma

Approaches to Improve User Experience in Product Development: UX Goals, Long-Term Evaluations and Usage Data Logging

Julkaisu 1585 • Publication 1585

Tampere 2018

Tampereen teknillinen yliopisto. Julkaisu 1585
Tampere University of Technology. Publication 1585

Jari Varsaluoma

Approaches to Improve User Experience in Product Development: UX Goals, Long-Term Evaluations and Usage Data Logging

Thesis for the degree of Doctor of Philosophy to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB216, at Tampere University of Technology, on the 2nd of November 2018, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2018

Doctoral candidate: Jari Varsaluoma
Laboratory of Pervasive Computing – Human-Centered Technology
Faculty of Computing and Electrical Engineering
Tampere University of Technology
Finland

Supervisors: Professor Kaisa Väänänen
Laboratory of Pervasive Computing – Human-Centered Technology
Faculty of Computing and Electrical Engineering
Tampere University of Technology
Finland

Dr. Heli Väätäjä
Laboratory of Pervasive Computing – Human-Centered Technology
Faculty of Computing and Electrical Engineering
Tampere University of Technology
Finland

Pre-examiners: Professor Minna Isomursu
Department of Digital Design
IT University of Copenhagen
Denmark

Assistant Professor Evangelos Karapanos
Persuasive Technologies Lab
Department of Communication and Internet Studies
Cyprus University of Technology
Cyprus

Opponent: Professor Philippe Palanque
The Interactive Critical Systems (ICS) team
Informatics Research Institute of Toulouse (IRIT)
University Toulouse III – Paul Sabatier
France

ISBN 978-952-15-4226-8 (printed)
ISBN 978-952-15-4251-0 (PDF)
ISSN 1459-2045

Abstract

Companies that develop digital products are increasingly interested in designing products and services that provide positive experiences for users. User-centered approaches can support product development teams in creating products that are more likely to offer the intended user experience (UX) to the customers. Instructions and guidance based on empirical studies in real-life contexts can provide valuable information for product development teams, including UX designers and other UX experts working in the industry. User-centered design and evaluation methods are used widely. However, there is a lack of specific guidance on how to achieve and evaluate the desired UX during the product development process of digital products.

This thesis explores three different approaches for supporting product development activities towards the desired user experience: setting UX goals, long-term UX evaluations, and utilizing usage data logging. This research work was motivated by the lack of empirical studies regarding the utilization of UX goals and long-term UX evaluation methods in product development of digital products. In addition, little is known about how product development teams perceive the utilization of usage data logging, especially in manufacturing industry. This thesis is based on research work conducted in five case studies and published in six scientific publications. Four of the studies were conducted in collaboration with product development practitioners from companies and included real or potential end-users (e.g., customers or company employees). Three of the studies evaluated UX of real products on the market. In total, during all the case studies, data from 185 survey responses and 20 interviews were analyzed, while log data were collected and analyzed from 61 participants.

First, the elicitation process of UX goals is explored as a part of the experience design process, where the intended experiences are set as a starting point of the design. Findings from nine design cases are synthesized as characteristics of a good UX goal, instructions for defining UX goals and as a theoretical Experience Goal Elicitation Process. Second, the benefits and challenges of specific, long-term UX evaluation methods are summarized. The perceptions of product development practitioners reveal that long-term UX evaluations can 1) help in understanding change in UX over time, 2) confirm expectations based on other data sources, 3) support updating of current products, and 4) support the conceptualization and development of future products. Third, the expected benefits by practitioners for usage data logging include the following: 1) the data show what users really do; 2) data can be collected without disturbing users; 3) findings from the data can inform user interface design decisions and 4) justify more qualitative user studies. Furthermore, 5) logged usage data can provide new business opportunities for supplier companies in the manufacturing industry. The utilization of usage data logging can be supported through collaborative development of visual data analytics tools between researchers from academia and practitioners from industry. Guidelines to support such collaborative processes are provided.

This research work provides theoretical contributions in the form of the Experience Goal Elicitation Process. Methodological contributions include a new understanding of benefits and challenges of long-term UX evaluation methods. Furthermore, the case study descriptions and the

instructions for defining UX goals contribute to product development activities on a practical level. These results can support product development teams and especially UX designers investigating the feasibility of UX goals and long-term evaluation methods for their product design and evaluation activities. The expected benefits of long-term UX evaluations and usage data logging can inform UX designers and managers when considering how to utilize these approaches. The guidelines for developing visual data analytics tools can support analytics tool developers in academia and industry. Future studies should develop the studied approaches further by investigating their utilization in different consumer and business-to-business contexts.

Preface

During this journey towards a doctoral degree, I have had a privilege to work with many wonderful, inspiring and insightful people. First, my utmost thanks go to my two excellent supervisors, Prof. Kaisa Väänänen and Dr. Heli Väätäjä, whose encouragement and expertise made this work possible. Their support has been invaluable especially during the last year of the dissertation process.

I would like to express my gratitude to all those colleagues from IHTE who have been working in the same research projects over the years. Jarmo Palviainen, Jani Sundström and Dr. Kati Kuusinen from sharing the experiences during my first research project at TUT from 2008. Good memories were made and studied also during the SUXES and DELUX projects, with Prof. Sari Kujala, Prof. Timo Saari, Dr. Tanja Walsh, Piia Nurkka, Tiina Koponen, Liinu Helkiö, and Farrukh Sahar. Special thanks go to Ville Kentta and Tommi Kanervo, the versatile full stack developers of the DrawUX tool, and Armand Djappi Noubissie for the testing.

I wish to thank all my co-authors of the included publications for their work and valuable insights, especially Dr. Eija Kaasinen, Hannu Karvonen, Dr. Yichen Lu, Dr. Tomi Heimonen, Dr. Jaakko Hakulinen, Katariina Tiitinen, Prof. Markku Turunen, and Harri Nieminen. Furthermore, as the majority of this dissertation work was done in close collaboration with industry, I wish to thank all practitioners from companies who participated in the research projects and provided their insights to make these studies happen.

I would also like to express my sincere appreciation to others who have shared the days in (and outside of) the office. There are so many of you that please forgive me for not to mention you all. The long list of dear colleagues includes Adjunct Prof. Thomas Olsson, Dr. Jarno Ojala, Dr. Teija Vainio, Dr. Satu Jumisko-Pyykkö, Dr. Timo Partala, Dr. Petri Ihantola, Dr. Aino Ahtinen, Dr. Kirsikka Kaipainen, Dr. Laura Hokkanen, Dr. Pradthana Jarusriboonchai, Katja Laine, Jaakko Luoto, Parisa Pour Rezaei, Elina Hildén, Susanna Paasovaara, Ekaterina Olshannikova, Emilia Pesonen, Chelsea Kelling, Otto Kauhanen, Iikka Pietilä, Aparajita Chowdhury, Aleksi Hiltunen, Pouya Eghbali and many more. Special thank you for Hilkka Losoi, Tiina Sävilahti and Jari “Jaffa” Salo for all the support with practicalities over the years.

I wish to thank my father Kari and my brother Jukka, together with Tanja, for your continuous support and encouragement during these years. Mum, I am certain you would also be proud.

Finally, my deepest thanks go to my own family. Otso, my favorite fireman-lumberjack-astronaut, may your spaceships and chainsaws never run out of fuel! Matilda, my precious little sunshine, may you reach your goals in life as successfully as you already reach the top of the children’s slide. Piia, my much-beloved wife, your patience, strength and positivity during this journey mean more to me than I can express with words. Never stop being you, for that is how I love you the most. For you, I dedicate this dissertation, for my heart is already yours.

Tampere, November 2018
Jari Varsaluoma

Contents

List of Publications

The thesis consists of a summary and the following original publications, referred in the thesis as P (publication) and a number 1-6. The publications are reproduced with the permission of the publishers. The candidate's contributions to each publication are explicated after each publication's reference information.

P1. Varsaluoma, J., Väätäjä, H. Kaasinen, E., Karvonen, H. and Lu, Y. 2015. The Fuzzy Front End of Experience Design: Eliciting and Communicating Experience Goals. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction* (OzCHI '15). ACM New York. pp. 324-332.

Varsaluoma and Väätäjä led the planning of the study that was carried out parallel to an academic workshop planned by all the authors. Varsaluoma conducted the data collection, the majority of the analysis and was the principal author of the publication.

P2. Varsaluoma, J., Väätäjä, H. and Walsh, T. 2016. Exploring Motivational Aspects and User Experience of Mobile Mathematics Learning Service in South Africa. In *Proceedings of the 20th International Academic MindTrek Conference* (Academic MindTrek'16). ACM, New York. pp. 159-168.

Varsaluoma led the planning of the study, conducted the data collection and majority of the data analysis. He was the principal author of the publication, planning and writing the majority of the paper.

P3. Varsaluoma, J. and Sahar, F. 2014. Usefulness of Long-Term User Experience Evaluation to Product Development: Practitioners' Views from Three Case Studies. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (NordiCHI'14). ACM, New York. pp. 79-88.

Varsaluoma led the planning of the study and performed the majority of the data collection and analysis. He planned and wrote the majority of the paper.

P4. Varsaluoma, J. and Sahar, F. 2014. Measuring Retrospective User Experience of Non-Powered Hand Tools: An Exploratory Remote Study with UX Curve. In *Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content & Services* (Academic MindTrek'14). ACM, New York. pp. 40-47.

Varsaluoma led the planning of the study, data collection, and analysis. He was the principal author, planning and writing the majority of the paper.

P5. Varsaluoma, J., Väätäjä, H., Heimonen, T., Tiitinen, K., Hakulinen, J., Turunen, M. and Nieminen, H. 2017. Usage Data Analytics for Human-Machine Interactions with Flexible Manufacturing Systems: Opportunities and Challenges. In *Proceedings of 21st International Conference Information Visualisation - IV2017*. IEEE. pp. 427-434.

Varsaluoma and Väätäjä planned the study conducted in parallel with a data analytics tool development that was the main responsibility of the other authors. Varsaluoma performed the majority of the data collection including user interviews and online surveys and led the analysis of the results. Varsaluoma was in charge of writing the publication, with major contributions to method, results and discussion sections.

P6. Varsaluoma, J., Väätäjä, H., Heimonen, T., Tiitinen, K., Hakulinen, J., Turunen, M. and Nieminen, H. Guidelines for Development and Evaluation of Usage Data Analytics Tools for Human-Machine Interactions with Industrial Manufacturing Systems. Accepted for publication in *Proceedings of the 22th International Academic MindTrek Conference* (Academic MindTrek'18). ACM, New York.

Varsaluoma and Väätäjä planned the user studies and Varsaluoma conducted the majority of the data collection and analysis. Varsaluoma led the paper writing, contributing mainly to related work, method, results, and discussion sections.

1 Introduction

Positive user experience is a well-known target in the product development of different types of digital products. This doctoral thesis addresses this issue through three different approaches aiming at supporting product development teams in creating products with the desired user experience. This chapter presents the background, structure, and goals of this compound thesis.

1.1 Background and Motivation

Today, product development companies consider pleasurable user experience (UX) vital for commercial success (Kujala & Miron-Shatz, 2013). User experience refers to "*a person's perceptions and responses that result from the use or anticipated use of a product, system, or service*" (ISO 9241-210, 2010). UX is becoming an increasingly important business goal, especially in the field of software development (Kuusinen & Väänänen-Vainio-Mattila, 2012; Rohn, 2007). Products that not only satisfy, but also exceed the expectations of customers or provide exceptionally good user experience, can help differentiate from the competitors. For instance, Jordan (1998) proposed that positive experiences with products from one company could improve users' repurchase intentions with the same company.

Designing for pleasurable UX is challenging, since UX is considered to be *dynamic*, *context-dependent*, and *subjective* (Law et al. 2009). Assuming a user-centered design mindset and including actual users in the design process can be seen as fundamental approaches when designing for desired UX (Gulliksen et al. 2003; ISO 9241-210, 2010). When designing for experience, the intended experience can be set to the fore, as a starting point to inspire the design process (Hassenzahl, 2013). The intended experiences can be described as *user experience (UX) goals*, aiming at providing a shared view of the intended experience for the whole product development team (Kaasinen et al. 2015; Väätäjä et al. 2015; Lu & Roto, 2014). However, there is a lack of empirical research on how these UX goals should be defined and communicated to stakeholders in product development of digital products.

The typical product development life cycle involves phases from the initial investigation and planning to iterative design, evaluation, and implementation (Roto et al. 2014). After the product is launched on the market, user support services may be provided, and user feedback can be collected (Roto et al. 2014). Human-centered design (HCD) traditionally focuses on the process before the product is launched on the market, including activities, such as familiarization with the context of use, gathering requirements, prototyping, and evaluation (ISO 9241-210, 2010).

However, the post-launch phase can be important for understanding how products are used over a longer period of time. According to Kujala et al. (2011), positive long-term experiences make people continue to use a product and recommend it to others. The value of such a recommendation or word-of-mouth for product acceleration and expansion is well recognized (Libai et al. 2013). Therefore, product development companies should benefit from understanding how to design for positive user experiences and evaluate how these experience goals are manifested with their customers over time. However, there is a lack of empirical research regarding the usefulness of long-term UX evaluation results from the perspective of practitioners in product development companies. Furthermore, little empirical research is available from exploring how retrospective *long-term UX evaluation methods* could be utilized to support product development in industry.

In addition to long-term UX evaluations, *usage data logging* can provide information on how, for example, web applications are used and offer an interesting viewpoint for observing if habits in product usage change over time (Jain et al. 2010; Karapanos et al. 2012a). While usage data logging has been widely utilized in marketing and e-commerce (see e.g., Lopes & Roy, 2015), there is a lack of studies focusing on how to support the utilization of usage data logging in the context of manufacturing systems development. First, to motivate any investments in usage data logging in companies in the manufacturing industry, it is important to understand how product development practitioners perceive the benefits of usage data logging. Second, approaches for introducing usage data logging for companies in manufacturing industry should be investigated. One possible approach is a collaborative development of visual data analytics tools between researchers in academia and product development practitioners in companies.

Human-computer interaction (HCI) literature provides guidance on human-centered design approaches, including various UX evaluation methods (e.g., Vermeeren et al. 2010) that can support product development teams in different phases of the product development life cycle. Case studies and reported experiences from empirical research utilizing different methods and tools for user-centered design can provide valuable information for product development teams in the industry. However, as presented above, such empirical studies may be scarcely available in terms of specific methods employed or the industrial domains investigated.

1.2 Research Objectives and Research Questions

The overall objective of this thesis is to create new understanding to support product development of digital products towards improved user experience for end-users. The thesis focuses on three topics and their role in product development from the user experience (UX) design perspective: UX goals, long-term UX evaluations, and usage data logging. These topics were chosen due to the lack of earlier explorative research done in specific product development contexts. Few studies have explored the elicitation process of UX goals and how specific, long-term UX evaluation methods and evaluation findings can be utilized in product development. Furthermore, little research is available on product development practitioners' perceptions towards usage data logging especially in the manufacturing automation context and on how to support collaborative development of visual data analytics tools for logged usage data in this context.

While the research is mainly conducted among users of digital products, such as software, one of the case studies explores the evaluation of experiences with non-digital, practical work tools, while utilizing UX evaluation methods traditionally used with digital products. In addition to UX evaluations, the goal is to explore how the utilization of quantitative usage data logging can be supported in product development. Therefore, the thesis mainly focuses on two points in the product development life cycle: the beginning of the design process when design goals are being formed and later stages when the long-term user experience and usage of the developed product can be evaluated.

The research questions of this doctoral thesis are the following:

RQ1. How can user experience goals be defined and communicated among stakeholders in product development?

RQ2. How can long-term user experience evaluation support product development?

> **RQ2a.** What kinds of perceptions do product development practitioners have about the usefulness of long-term user experience evaluation?
>
> **RQ2b.** How can user experience evaluation methods and tools support the long-term user experience evaluation in product development?

RQ3. How can the utilization of usage data logging be supported in product development?

The first research question (**RQ1**) aims to understand the beginning of the process for designing for specific types of user experience. The question seeks to identify the sources of inspiration when defining user experience goals and explore how these goals are communicated to stakeholders participating in the product development process. This question is approached by summarizing reported experiences from nine experience design cases in different contexts, conducted independently from this thesis work by other researchers and designers.

The second research question (**RQ2**) explores how long-term UX evaluation can benefit product development and the challenges related to evaluating UX over time. The two sub-questions address specific aspects of the main question: what kinds of perceptions do product development practitioners have toward long-term UX evaluation (**RQ2a**) and how can UX evaluation methods and tools support long-term UX evaluation in product development (**RQ2b**)? **RQ2a** is approached through a long-term case study in a company, where practitioners assess the usefulness of the long-term UX evaluation results of their products during three evaluation studies. **RQ2b** is approached by synthesizing both researchers' and participants' experiences with UX evaluation methods during the case studies included in this doctoral thesis. A specific focus is given to retrospective long-term UX evaluation methods in remotely conducted studies.

The third research question (**RQ3**) explores how to support UX designers, developers, and other practitioners participating in the product development process in utilizing usage data logging from product development perspective. This question is answered from two viewpoints, as related to the conducted studies. First, in order to support product development teams in utilizing usage data logging, the teams' needs for usage data logging and the expected benefits in the specific context need to be understood. This includes expectations of the benefits of usage data logging for improving the UX of the developed product. Second, tools that can support easy inspection of the logged usage data need to be available. The way in which this question is approached in this research is to develop a visual data analytics tool in collaboration with practitioners from a company. This approach enables synthesizing guidance to support the development and evaluation of visual data analytics tools in specific contexts and may result in more variety of such tools in the future.

1.3 Contributions

The results of this research work are related to the three topics of this thesis: setting UX goals, evaluating long-term UX, and utilizing usage data logging. Table 1 presents the relation of the five conducted case studies (**I-V**, presented in sections 3.2.1 - 3.2.5), six publications (**P1-P6**), and research questions, and provides summaries of the main contributions of the publications. The main theoretical contribution of this thesis is the *Experience Goal Elicitation Process* that illustrates the iterative process for defining and disseminating UX goals (**RQ1**) especially during the early phases of experience design in product development. As a practical contribution related to the utilization UX goals (**RQ1**), this thesis suggests *a notation for describing UX goals*, *characteristics for a good UX goal,* and *instructions for defining and evaluating UX goals*. These contributions can support UX designers in the early stages of experience design process.

Other results of this research work can contribute especially to the evaluation activities during the product development life cycle, either during the iterative development of the product or after product launch, e.g., when collecting feedback for the next product version. The reported findings from the *perceived usefulness of long-term UX evaluation results* (**RQ2a**) and *summarized benefits and challenges of specific long-term UX evaluation methods* (**RQ2b**) can inform product development teams interested in utilizing similar methods in their product development projects. Finally, product development teams, especially in manufacturing automation, can benefit of the results regarding the *perceived benefits of usage data logging for product development purposes* in this domain, and the *set of questions to support discussions on the feasibility of usage data logging* (**RQ3**). Lastly, collaborative development of visual data analytics tools between academia and industry is suggested as a viable approach to support the utilization of usage data logging in companies. The proposed *guidelines to support the development of analytics tools for usage data logging* (**RQ3**) can provide guidance for analytics tool developers in academia and industry. The results are presented in Chapter 4 and the contributions are discussed in Sections 5.1 and 5.2.

Table 1. Case studies (see sections 3.2.1 - 3.2.5), publications, and their contributions to related research questions.

Case Study	Publication	Related research questions and contributions
I	P3	**RQ2, RQ2a, RQ2b**. Explores how practitioners in a company developing interactive sports products evaluate the usefulness of long-term UX evaluation and utilize evaluation results in practice.
II	P4	**RQ2b**. Explores the utilization of AttrakDiff and UX Curve methods in a survey study for long-term UX with a practical, non-digital product.
III	P1	**RQ1**. Presents the Experience Goal Elicitation Process and instructions for defining UX goals based on two survey studies.
IV	P2	**RQ1, RQ3**. Example case presenting a user study for defining UX goals to support developing motivational mobile learning services. Demonstrates benefits of usage data logging in addition to remote user experience evaluation.
V	P5	**RQ3**. Presents potential usage metrics to collect and discusses the opportunities of usage data logging from the viewpoints of stakeholders in an industrial R&D company. Provides a set of questions for stakeholders to discuss in companies interested in utilizing usage data logging to support their product development activities.
V	P6	**RQ3, RQ2b**. Provides guidelines to support usage data analytics tool development in collaboration with end-users from manufacturing automation industry. Describes a visual data analytics tool, UX-sensors, for logged usage data and the process for its development with end-users.

1.4 Structure of the Thesis

The rest of the thesis is organized as follows. Chapter 2 presents related work to provide the reader a background of the topics studied during this thesis work. These topics include the concept of user experience (UX), temporal aspects of UX, UX goals, long-term UX evaluation and usage data logging. Chapter 3 presents the research approach and research process with summaries of each case study. Chapter 4 presents the main results of the thesis in relation to each research question. Chapter 5 discusses the contributions in the light of previous research and in relation to the product development life cycle. Finally, the assessment of the research and implications for future research are presented, followed by chapter 6 with conclusions.

2 Related Research

This section provides an overview of the main topics of the thesis as discussed in the relevant literature. Section 2.1 provides an overview of the concept of user experience, with focus on the temporal aspects of experiences. Section 2.2 describes experience design from the product development viewpoint, with focus on the experience design process, UX goals, long-term UX evaluation, and evaluation methods and tools. Section 2.3 provides an overview of usage data logging in HCI research and previous research regarding visual data analytics tool development with end-users, with a specific focus on manufacturing automation context. Finally, Section 2.4 summarizes the identified gaps in current research, as related to the theme of the thesis and the research questions.

2.1 The Concept and Temporal Aspects of User Experience

Although "user experience" (UX) is one of the core concepts in the field of the Human-Computer Interaction (HCI), there is no universally accepted definition for UX (Lallemand et al. 2015). One of the most prevalent definitions, although ambiguous (Hassenzahl, 2008), for user experience is from ISO: "*a person's perceptions and responses that result from the use or anticipated use of a product, system or service*" (ISO 9241-210, 2010). While numerous definitions have been proposed for UX (Desmet & Hekkert, 2007; Hassenzahl & Tractinsky, 2006; Law et al. 2009), HCI experts in industry and academia seem to agree that UX results from the interaction between the *user*, the *system* and the *context* (Roto et al. 2011). For instance, Hassenzahl and Tractinsky (2006) define UX as "*a consequence of a user's internal state, the characteristics of the designed system and the context within which the interaction occurs*".

Based on a survey study by Law et al. (2009), UX is generally considered as a *dynamic*, *context-dependent*, and *subjective* concept. In addition, several researchers have highlighted the emotional aspects of UX (Forlizzi & Battarbee, 2004; Hassenzahl &Tractinsky, 2006; Isomursu et al. 2007; Mahlke, 2005) as well as the holistic and phenomenological nature of UX (Wright, P. & McCarthy, 2004; Swallow et al. 2005). Furthermore, the temporal aspects of UX have been discussed in several studies, suggesting that UX can change over time (Roto et al. 2011; Fenko et al. 2010; Karapanos et al. 2008; Karapanos et al. 2009).

The emotional, subjective, and temporal aspects of UX distinct it from the concept of usability (Lallemand et al. 2015), which can be defined as ""*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*" (ISO 9241-210, 2010). Usability and UX are both grounded in User-Centered Design (UCD), often used as a synonym for Human-Centered Design (HCD), which is ''*an approach to interactive systems development that aims to make systems usable and useful*'' (ISO 9241-210, 2010). HCD can be viewed as a philosophy as well as a design approach, as it places the user at the center of all design activities (Mahlke, 2008). However, "human-centered", in comparison to “user-centered”, can be viewed as more holistic approach, considering all stakeholder groups, for instance designers and salespersons, affected by the product, not just the target users. UX is often considered as an extension of usability, with focus on people’s feelings when using interactive systems (Law et al. 2015). Good usability can be seen as a requirement for positive UX (Hassenzahl, 2008; Lallemand et al. 2015) and usability factors have been included in UX frameworks as “pragmatic” (Hassenzahl, 2003) or “instrumental” (Mahlke, 2008) aspects of UX.

Perhaps the most referred UX framework in academic research (Law et al. 2015) is the Hassenzahl’s hedonic-pragmatic model (Hassenzahl, 2003, 2004). Hassenzahl’s model describes UX as a result of two product attributes: *pragmatic* and *hedonic* (Figure 1). Pragmatic attributes, including *usability* and *utility,* relate to user’s needs to achieve behavioral goals, i.e. *manipulate* the environment, while hedonic attributes, *identification*, *stimulation*, and *evocation*, are related to the user’s self (Hassenzahl, 2003). Hassenzahl (2004) suggests that pragmatic products essentially provide efficient and effective ways to achieve user’s behavioral goals. Moreover, a hedonic product can have challenging and novel character, therefore providing stimulating experience, or the product can support identification by enabling the user to express one’s personal values to other people. Finally, evocation refers to how products, such as souvenirs, can provoke memories of the past events and relationships (Hassenzahl, 2003).

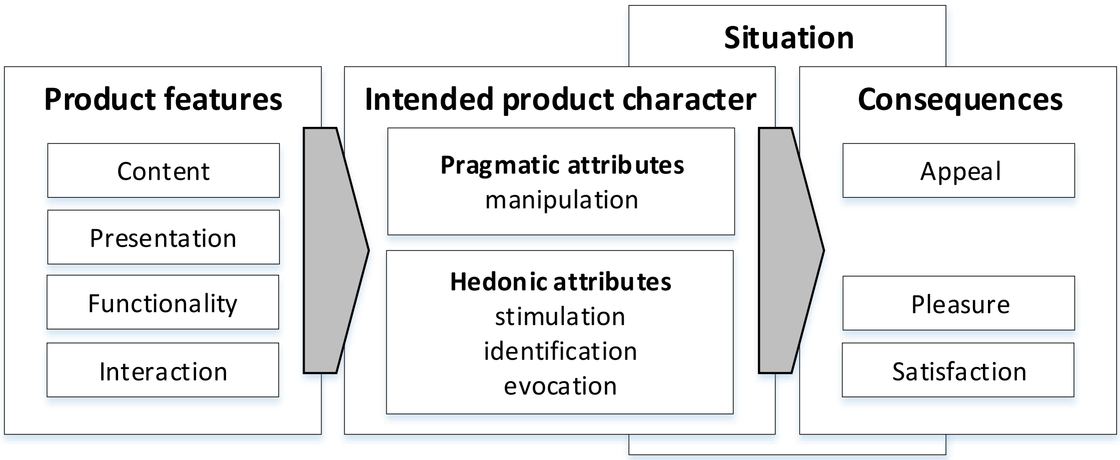

Figure 1. Hassenzahl’s UX model representing the key elements in the process of forming UX as consequences from interacting with the product in a specific usage situation (combination of two figures from Hassenzahl, 2003)

Mahlke & Thüring (2007) (see also Mahlke, 2008) present UX in the component-based model, where the influencing factors from the system, user and context/task affect the *perception of instrumental qualities* (e.g. usefulness, utility, and usability) and *non-instrumental qualities* (aesthetic, symbolic, and motivational aspects). These perceptions lead to *emotional user reactions* (e.g. subjective feelings, motor expressions, and physiological reactions). Together, these perceptions and emotional user reactions result in the *consequences of user experience*, such as overall judgments, choice between alternatives, and usage behavior. In comparison to

Hassenzahl's UX model in Figure 1, the Components of User Experience -model (CUE-model) contains a particular UX component for emotional reactions.

In this thesis, the concept of UX is understood as it is defined in Hassenzahl's UX model (Hassenzahl, 2003). However, here the temporal aspects of UX are emphasized, as it has been discussed in the following examples.

The **temporal aspects** of UX have been discussed by Roto et al. (2011) in the UX White Paper. The White Paper presents the time spans of UX, which divide UX into *anticipated UX*, *momentary UX*, *episodic UX*, and *cumulative UX* (Figure 2). The cumulative UX refers to series of usage and non-use episodes with a product. Cumulative UX can include longer periods, such as months or more, while episodic UX relates to a single usage episode and momentary UX to a change of feeling during one usage episode. *Anticipated UX* relates to expectations before the first usage episode, including also the imagined use of the system during specific moments or over time. In this thesis, the term *long-term UX* entails several usage episodes, similar to cumulative UX.

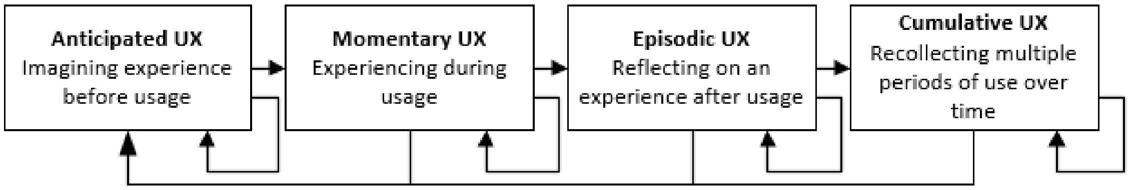

Figure 2. Time spans of UX and the user's internal process taking place in different time spans (adapted from Roto et al. 2011)

Several studies have shown how the significance of the different aspects of UX can change over time (Fenko et al. 2010; Karapanos et al. 2008; Karapanos et al. 2009). Karapanos et al. (2009) followed six individuals who purchased Apple iPhones during a five-week ethnographic study. Based on their results, they propose a framework (Figure 3) illustrating the temporality of experience, consisting of three main forces: *familiarity*, *functional dependency* and *emotional attachment.* These are considered as the main motivational forces supporting the transition between three experiential phases: *orientation*, *incorporation* and *identification*. The most appreciated product qualities can change in each phase. According to Karapanos et al. (2009), users' early experiences of excitement and frustration manifest in the *orientation* phase when users encounter novel features or learnability problems. Satisfying experiences related to product qualities that provided *stimulation* (e.g. *visual aesthetics* and *aesthetics in interaction*) and *learnability*, while dissatisfying experiences reflected learnability issues. As the product becomes more *incorporated* in users' lives, the experiences with it reflect how the product becomes more meaningful in various use situations. *Long-term usability* and *usefulness* were the most dominant product qualities regarding the satisfying and dissatisfying experiences. Finally, *identification* relates to how users form a personal relationship with the product, because of the increased incorporation of the product in their daily routines. Two perspectives for identification were found: *personal* and *social*. Users *personalized* their products and used them in *daily rituals*, while also reporting experiences related to social aspects of product ownership, such as *enabling self-*

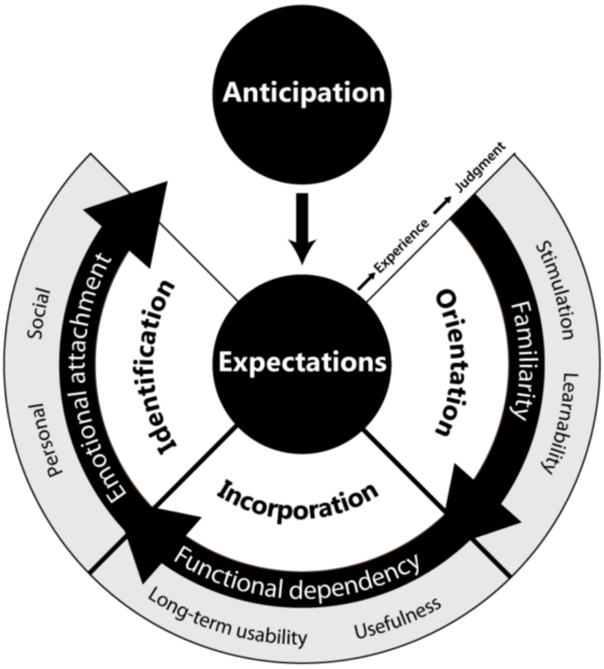

Figure 3. Temporality of experience by Karapanos et al. (2009, with permission).

expression (i.e. differentiating themselves from others) and *creating a sense of community* (i.e. being a part of a group with shared values) (Karapanos et al. 2009).

In general, there has been a movement in UX literature to focus more on the long-term UX (Karapanos et al. 2009), e.g. the role of memories (Norman, 2009), instead of experiences of early use with digital products. Kujala et al. (2011) have suggested that **the long-term UX is the reason why people keep using a product and recommend it for others over time**. Karapanos et al. (2009) have argued for the need of long-term studies by stating that **products are becoming more service-centered and the number of returned products due customers' remorse has increased**. Therefore, **it should be vital for companies to understand how to a) design for and b) evaluate that the product provides positive long-term experiences for their customers. This notion is one of the main motivations for the studies in this thesis that focus on the evaluation of user experience over longer usage periods**. In the following section, the relevant research work on UX design is presented, with foci on UX processes, UX goals, and the long-term evaluation perspective.

2.2 Perspectives of User Experience Design

This section presents related research on the design perspectives of UX that are relevant to the main goal of this thesis, i.e. improving UX in product development. First, the relevant experience design processes are presented, followed by the definition of and examples from utilizing user experience goals in experience design. The last two sections include related work on long-term UX research and long-term UX evaluation methods used in HCI.

2.2.1 Experience Design Processes

Experiences with interactive products are considered dynamic, subjective, and context-dependent (Hassenzahl & Tractinsky, 2006). Therefore, it is suggested that instead of designing *an experience* per se, designers can only *design for an experience*, with an aim to providing the intended experience for the user (Wright et al. 2003; Sanders & Dandavate, 1999). Designers can then observe and measure how successfully their intended experience was manifested when users interact with the product. This can include different physical, cognitive, emotional, and aesthetic aspects in experiencing a product. While designing for experiences is difficult, it is worthwhile as such approach can extend simple usability techniques and help to differentiate on the market (Forlizzi & Battarbee, 2004; Cagan & Vogel, 2001; Bloch, 1995).

In the context of HCI development, human-centered design (HCD), also called user-centered design (UCD), is the most well-known approach. Human-centered design process (ISO 9241-210, 2010) includes five process activities, starting with planning for HCD and proceeding in an iterative cycle that includes the following activities: specifying the context of use, specifying requirements, producing design solutions and evaluating designs against requirements (Figure 4). The cycle is repeated until the design meets the requirements. Active user participation is seen as a requirement for successful HCD (Gulliksen et al. 2003) and the lack of direct feedback from actual users has been associated with failure in software projects (Kujala, 2008).

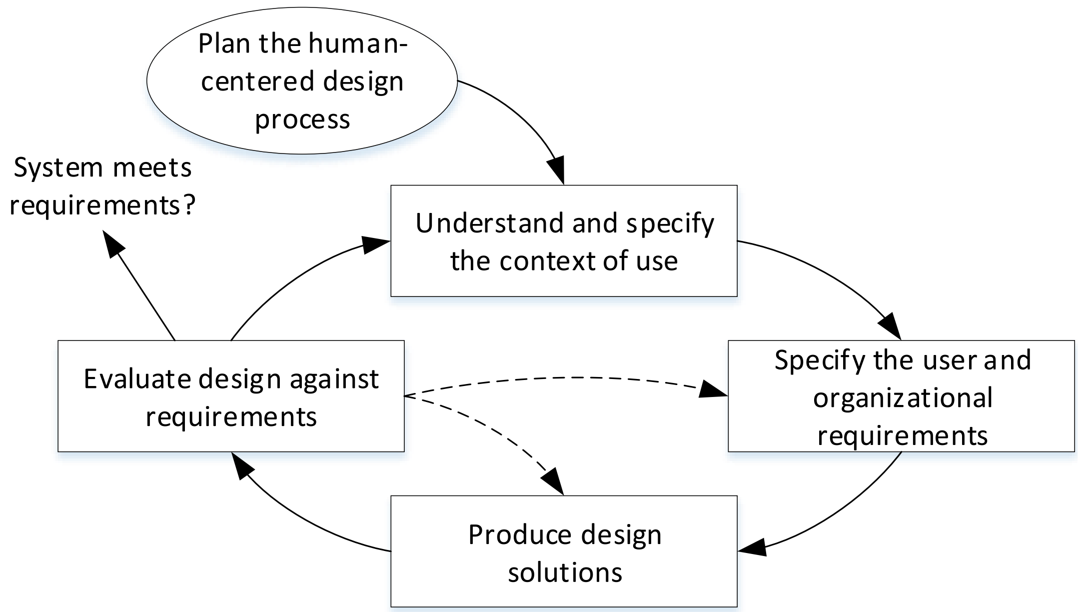

Figure 4. Human-centered design process (ISO 9241-210, 2010)

While the first version of the ISO standard from 1999 (ISO 13407, 1999) focused on usability, the updated version from 2010 (ISO 9241-210, 2010) includes user experience as one of the six key principles in HCD:

1. *The design is based upon an explicit understanding of users, tasks and environments.*
2. *Users are involved throughout design and development.*
3. *The design is driven and refined by user-centred evaluation.*
4. *The process is iterative.*
5. *The design addresses the whole user experience.*
6. *The design team includes multidisciplinary skills and perspectives.*

The ISO's HCD process does not instruct the use of specific methods in the different process activities. The choice of methods e.g. for UX evaluation should be based on the needs and requirements in the specific design project so that the goals of the evaluation situation can be met (Hartson & Pyla, 2012). By conducting research regarding the feasibility of specific UX evaluation methods in specific contexts, researchers can provide advice for UX designers planning their UX evaluation activities in similar contexts.

The interest towards UX has resulted in a number of novel design approaches and concepts, such as *emotional design* (Norman, 2004), *experience-centered design* (Wright & McCarthy, 2010), *experience design* and e*xperience-driven design* (Hassenzahl, 2010; Hekkert et al. 2003). According to Hassenzahl (2013, Section 3.4), Experience Design brings "*the resulting experience to the fore – to design the experience before the product*''. Similar to HCD, Experience Design emphasizes the user's perspective as the main reference point during the development process but especially highlights the quality of the experience as felt by the user (Lallemand et al. 2015). As Hassenzahl (2013, Section 3.6) states: "*Experience Design stands for technology, which suggests meaningful, engaging, valuable, and aesthetically pleasing experiences in itself*".

According to Hartson & Pyla (2012), UX design process entails four basic UX activities: analysis, design, implementation and evaluation. Hartson & Pyla (2012) propose a model for UX design process on a general level in the context of *interaction design*. Interaction design is "*the practice of designing interactive digital products, environments, systems, and services*" (Cooper et al. 2007). In the presented model, the *analysis* refers to activities that aim to understand users' work and requirements. Possible activities include familiarizing oneself with the context where the product will be used, extracting user requirements from the contextual data and synthesizing models that can inform design (Hartson & Pyla, 2012). *Design* refers to "*creating conceptual design and determining interaction behavior and look and feel*" (Hartson & Pyla, 2012). This involves design activities such as ideation, sketching, and brainstorming, but also redesign for the next product version. Design activities are realized in the *implementation* phase, which translates to prototyping. Prototypes are often built in parallel with design work and can vary from low fidelity (e.g. paper prototypes) to high fidelity prototypes (e.g. functional prototypes requiring programming). In the *evaluation* phase, the current design is evaluated to realize how well it meets user needs and requirements. The whole design process is iterative and each activity contains various possible sub-activities, such as utilizing specific evaluation methods in the evaluation phase. Furthermore, boundaries between the activities are not rigorous as there can be significant overlap between the different design activities (Hartson & Pyla, 2012).

“UX HOW?” booklet (Roto et al. 2014) from FIMECC UXUS research programme (Finnish Metals and Engineering Competence Cluster, User experience and usability in complex systems, 2010-2015) includes a process model relevant to the thesis, illustrating UX and product life cycle. The process model (Figure 5) summarizes UX activities during and after system development, especially related to UX goal setting and evaluation. The concept of UX goal is further discussed in the next section. The model is inspired by several studies done in the FIMECC UXUS programme (e.g. Karvonen et al. 2012b; Koskinen et al. 2013; Karvonen et al. 2014; Wahlström et al. 2014; Kaasinen et al. 2015; Kaasinen et al. 2017). This model is utilized in this thesis to anchor the main research contributions to steps in the product development life cycle (Section 5.2).

This thesis focuses on the UX activities prior the concept design (see Figure 5) and UX evaluation activities over time in the product development life cycle. The process model presented in Roto et al. (2014) is used as a framework to support the discussion about the results from the perspective of UX design process and the product development life cycle with emphasis to the UX activities studied in this thesis: UX goal elicitation, long-term UX evaluation and usage data logging. **The aim is to elaborate the beginning of UX design process regarding the elicitation of UX goals but also extend the scope to time after the product launch**, where product support activities such as user training, updates, and ongoing maintenance take place.

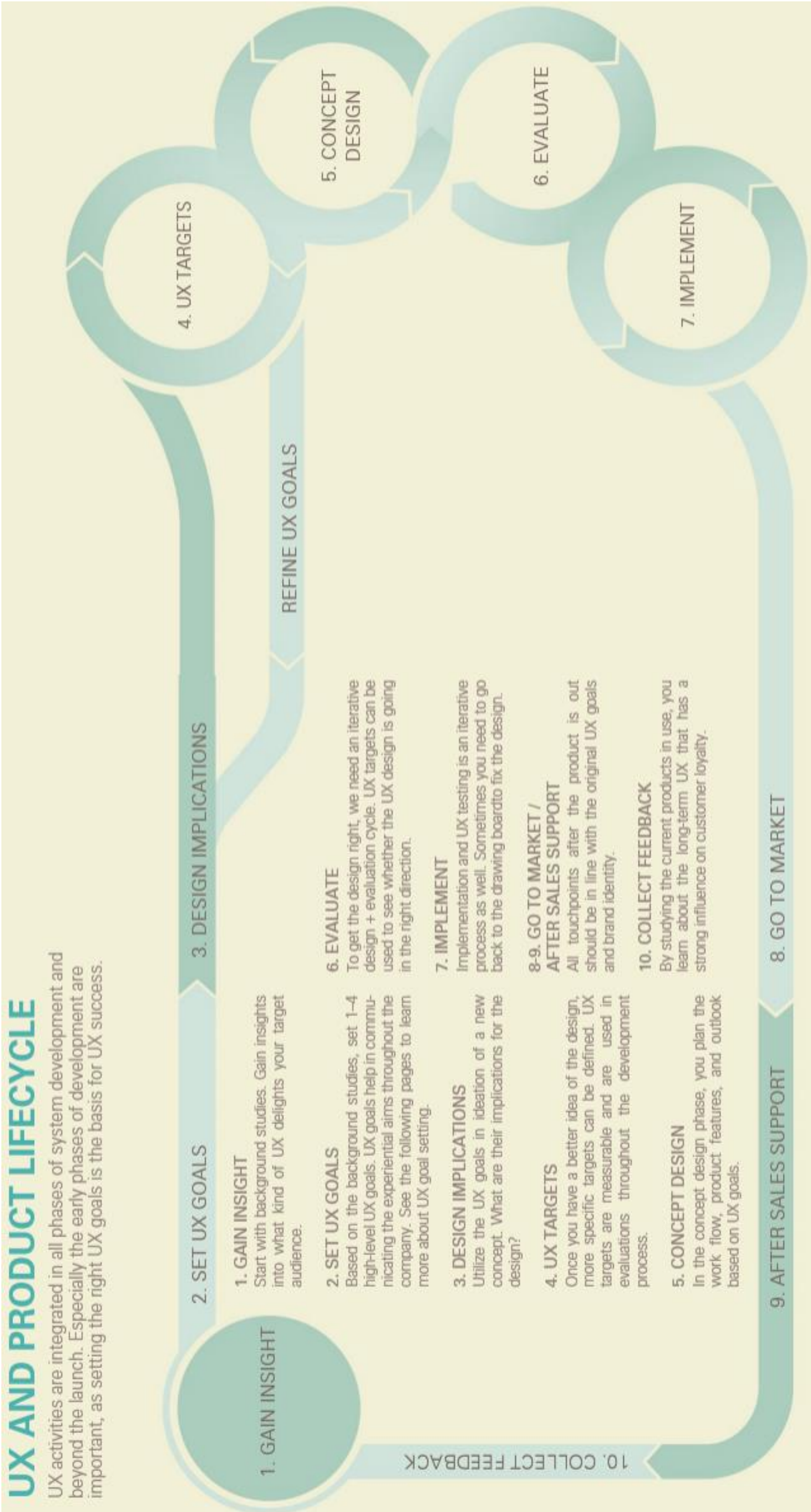

Figure 5. UX activities during a product life cycle process as based on research work in FIMECC UXUS research programme (Roto et al. 2014). Published with permission from the authors Eija Kaasinen, Hannu Karvonen and Joona Elo (visualization).

2.2.2 User Experience Goals

An important challenge at the beginning of an experience design process is to define what experience to aim for, while another challenge is to design something that should evoke such experience (Desmet & Schifferstein, 2011). The aimed experiences can be described as *experience goals* (Kaasinen et al. 2015) or *UX goals* (Väätäjä et al. 2015; Lu & Roto, 2014). In the context of this thesis, *experience goal* and *UX goal* are treated as synonyms, as in related literature (Kaasinen et al. 2015). UX goals refer to "*experiences that a designer intends the designed system to support for the end-users when they use the system in their activities*" (Väätäjä et al. 2015). Furthermore, UX goal states "*the intended momentary emotion or the emotional relationship/bond that a person has towards the designed product or service*" (Lu & Roto, 2014). Rogers et al. (2011) suggest that UX goals can include aspects such as enjoyable, exciting, challenging, fun, motivating, and creativity support. However, designers should recognize that there can be contradictions between UX goals and some UX goals are not even desirable, such as a fun and safe control system for an aircraft, where safety is the main concern.

Clearly defined UX goals that are shared with the entire product development team can help "*keeping UX in focus through the multidisciplinary product development and marketing process*" (Kaasinen et al. 2015). For UX goals to be useful in design and evaluation purposes, they need to be operationalized by mapping them to functional and non-functional requirements, and target experiences, including system qualities, feelings, and emotions (Väätäjä et al. 2015; Hassenzahl, 2010). According to Väätäjä et al. (2015), the operationalized UX goals become *UX targets*, when the UX goals are measurable and can be evaluated with users, for instance by comparing the users' subjective experiences (presented verbally) with the set UX goals.

According to Hartson & Pyla (2012), UX goals, *UX metrics* and *UX targets* guide the whole UX design life cycle, especially the prototyping and evaluation activities. They define *UX goals* as "*high-level objectives for interaction design, stated in terms of anticipated user experience*" (Hartson & Pyla, 2012). UX goals can be described as desired experiences that users will have when interacting with specific features of the design (ibid.). Hartson & Pyla (2017) suggest, that UX goals should be connected to a specific user group, describing user's tasks and background, as these can inform the setting of target levels for evaluating user performance. *UX measure* "*is a usage attribute that is assessed in evaluating a UX goal*" (ibid.). UX measures are *objective* when they can be directly measured by evaluators, or *subjective* when they are based on user's opinions. Examples of objective UX measures are initial performance, long-term performance and learnability, while subjective UX measures include first impressions (initial satisfaction) and long-term user satisfaction (ibid.). *UX metric* states what is being measured. Potential metrics include e.g. time on task, error rates, the percentage of task completed in a given time, or average scores on questionnaires (ibid.). A *UX target* is "*a quantitative statement of an aimed-at or hoped-for value for a UX metric*" (Hartson & Pyla, 2012) that indicates the level of success in the attained user experience. With the term *UX target*, Hartson & Pyla (2012) also refer to the group of information that describes a specific user group and a single UX goal with its UX measures, metrics and target level. Finally, Hartson & Pyla (2012) provide an example on setting UX targets for a ticket kiosk system. For a casual new user buying tickets, one UX goal suggested is "*fast and easy walk-up-and-use user experience, with absolutely no user training*". The UX measure for this goal is "*initial user performance*" and the chosen UX metric is the average number of errors during a task when buying a movie ticket, with less than one error (on average) as a target level. (Hartson & Pyla, 2012).

The first academic workshop to collect cases of UX goal utilization was held in a NordiCHI2012 conference (Väätäjä et al. 2012; Väätäjä et al. 2015). Nine case studies presented in the workshop by academics and practitioners represented different domains, including workplace (remote operation of cranes, a learning tool for forklift drivers), consumer applications (online bingo, designing for dogs), and education (teaching experience-driven design for university students). Three topics were discussed in the workshop regarding UX goals: 1) what constitutes a good UX goal, 2) how UX goals are identified, and 3) how UX goals affect design? A good UX goal was seen as something that *is measurable, clear and precise*, but *broad enough to allow space for design ideas*. Furthermore, UX goal should *guide the design*, *evoke design ideas,* and above all, support *communication*. UX goals were mainly identified through *user studies*, *literature* and *theory*, but could also be *given by a customer*. Furthermore, *brand*, *standards*, *ethical guidelines*, *benchmark study* and *common sense* were mentioned as sources for identifying UX goals. UX goals were seen as important as they can influence the project outcome by *focusing and guiding the design process* and *providing inspiration and vision*, but also *supporting communication to educate the organization about UX*. Finally, UX goals can change during the UX design process, as the chosen goals are made more precise, dropped out, or new goals (e.g. business goals) are identified. Since the results from the workshop represent a limited number of cases, more research is required, especially regarding the identifying and choosing of UX goals in real-life design cases. (Väätäjä et al. 2012; Väätäjä et al. 2015).

Kaasinen et al. (2015) studied the UX goal -setting in industrial environments and identified five approaches to gain insight and inspiration:

- **Brand**: UX Goals Derived from Company and Brand Image
- **Theory**: Deriving UX Goals from Scientific Understanding of Theory Human Beings
- **Empathy**: Inspiration from Designer's Empathic Understanding of Users' World
- **Technology**: UX Goals Identified Based on Possibilities and Challenges of a New Technology
- **Vision**: Inspiration from Investigating the Deep Reasons for Product Existence and Envisioning Renewal

Kaasinen et al. (2015) derived these approaches from four industrial design case studies with companies and supplemented their findings with literature study. These approaches were utilized in **Study III** (see section 3.2.3) as a background for studying the inspiration and identification of UX goals.

Karvonen et al. (2012a) suggest that as new design solutions are presented, they should be traceable back to the originally defined UX goals. In this way, the fulfillment of the aimed UX goals can be measured and evaluated in different phases of the design process, including when the product is evaluated with end-users (Karvonen et al. 2014). This evaluation for the claims of the original UX goals can include a combination of methods, such as user interviews, testing sessions, and UX questionnaires (Karvonen et al. 2014). However, as customers use the designed products over long-term, designers might need longitudinal studies to measure if their intended experiences were successful.

According to (Väätäjä et al. 2015), there is little research available in Human-Computer Interaction (HCI) literature regarding the process of defining, communicating and using UX goals as requirements in real-life design cases. **One of the goals of this thesis is to provide more empirical evidence from real-life design cases to increase the understanding about UX goal elicitation activities.**

2.2.3 Long-Term UX Research

Although there is no exact definition for longitudinal (or long-term) UX research in HCI literature, longitudinal research can be considered as something that looks beyond the initial user experience or learning experience (Jain et al. 2010). Researchers' interest towards the temporal aspects of usability and UX in HCI field has increased during the last decade. In CHI 2007, a special interest group was organized to discuss capturing longitudinal usability (Vaughan & Courage, 2007), continued by a panel presentation at CHI 2008 (Vaughan et al. 2008), a workshop at CHI 2009 and another SIG in CHI 2010 (Jain et al. 2010). In CHI 2012, HCI practitioners and researchers met in a workshop to discuss the theories, methods, and case studies of longitudinal HCI research (Karapanos et al. 2012a).

Karapanos et al. (2010) propose four methodological paradigms for longitudinal research in HCI aiming to understand changes in UX and behavior over time: *cross-sectional*, *within-subject repeated sampling*, *longitudinal* and *retrospective* (Figure 6). Cross-sectional studies include evaluations between different user groups with interest to specific background variables such as expertise in the product usage or length of product ownership. Within-subject repeated sampling designs study the same group of participants at two points in time, while longitudinal research includes more than two data gathering waves, therefore providing better insight into the form of change over time. Finally, Karapanos et al. (2010) propose retrospective designs as a lightweight alternative to laborious longitudinal studies. Retrospective approaches rely on supporting the elicitation of users' experiences from memory within a single measurement contact with the respondents, therefore decreasing the effort required from the respondents. Several methods for supporting such retrospective design are presented in the next section.

Earlier research suggests that conventional usability testing methods can be inadequate for revealing the problems that can frustrate more experienced users over time (Mendoza & Novick, 2005). Longitudinal studies try to understand what usability problems may persist over time and how the UX changes after the initial learning period with a product. This can be relevant, for example, 1) for managers in an online gaming company who are interested to know why their customers are not motivated to play their games after a while, or 2) product designers and their

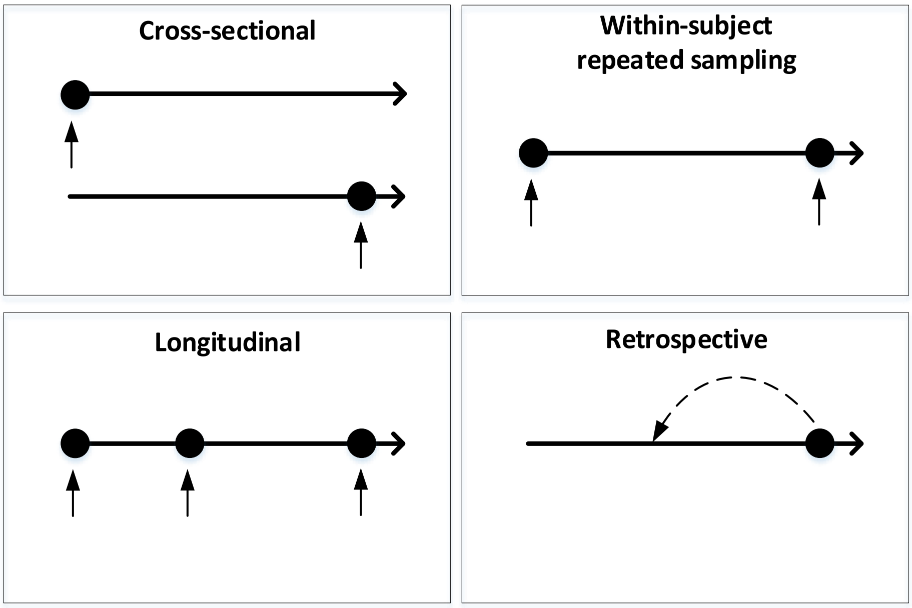

Figure 6. Longitudinal research paradigms in HCI (redrawn from Karapanos et al. 2010, horizontal arrowheads added).

customers who want to evaluate the learnability of a complex factory monitoring system with new employees. Longitudinal research "*is ideal for studying how and when users transition from novice to expert, as well as addressing issues such as abandonment or adoption rates, learnability, comfort with technology, productivity, and evolution of user perceptions*" (Jain et al. 2010). Despite these proposed applications, there is lack of empirical evidence on the actual benefits, challenges, and best practices of long-term UX evaluations in companies and of how product designers and other practitioners can utilize the acquired information in practice. Based on the candidate's current knowledge, empirical studies focusing on the *usefulness of longitudinal UX research results to work practice* are rare or non-existent in HCI literature.

Several open questions about methods, risks, benefits, and practices related to longitudinal research in HCI have been presented in CHI conference workshops, summarized by Jain et al. (2010). Traditional user research methods focus on the momentary or "first-time" experiences with products (Vermeeren et al. 2010) yielding mainly learnability and discovery problems instead of persisting usability concerns (Jain et al. 2010). Furthermore, studies where product usage is tracked over several days or weeks are rare because of their expenses and excessive participant fatigue (Karapanos et al. 2009). Although UX activities after the product launch could be beneficial for improving the current and future products, it is possible that often the focus of UX teams in companies has already changed on to other projects. However, according to Kujala et al. (2011), it is the long-term UX, not the details of each individual experience that makes people continue to use a product and recommend it to others. The value of such recommendation or word-of-mouth for product acceleration and expansion is well recognized (Libai et al. 2013). Furthermore, Karapanos et al. (2009) argue that 1) products are becoming more service-centered, where the revenue is increasingly coming from the continued service provision, and 2) the competition and legislation has resulted in longer warranties with more coverage, leading to an increase in complaints and product returns based not just on technical failures, but purely on the customer's remorse or failing to satisfy the customers' true needs (Ouden et al. 2006). Therefore, product development companies should benefit from understanding how to design for positive UX and evaluate how these aimed experiences are manifested with their customers over time.

Previous research has shown that the significance of the different aspects of UX can change over time (Fenko et al. 2010; Karapanos et al. 2008; Karapanos et al. 2009). For example, in the study by Karapanos et al. (2009) six people who had bought Apple iPhone were followed for one month. The researchers found that the importance of novelty and social meaning decreased after the beginning of use. However, different aspects of hedonic quality surfaced, such as having the product with you in daily rituals and important activities. More recently, Harbich & Hassenzahl (2016) conducted a 13-week longitudinal field study of UX with work-related interactive systems. They concluded that time is the best predictor for behaviors associated with UX in the work domain and for hedonic and pragmatic quality, verifying that UX changes over time. Furthermore, the change in UX over time differed substantially between participants and even increased over time. In another longitudinal study with 165 users of a proximity mobile payment service, Kujala et al. (2017) studied the role of expectations for subjective usability and emotional experiences after three and six weeks of product use. Their research revealed that if users have not familiarized themselves with the product and gained experiences with it in varied situations, the short-term UX measurements might reflect more users' expectations than actual experiences. However, after six weeks of product use, the effect of expectations decreased and the cumulative experiences during product use had more effect on evaluations (Kujala et al. 2017). This implies that **if companies focus only on the initial experiences with their products and services, they may get biased results**. Instead, long-term UX evaluations are also required for new products as these experiences can affect the evaluations and willingness to recommend the product to others (Kujala et al. 2017).

Longitudinal studies can vary in length, and case studies from three weeks to three years have been presented as longitudinal studies (Karapanos et al. 2012a). Von Wilamowitz-Moellendorff et al. (2006) propose three perspectives for HCI studies based on the time period the study covers. Usability tests are usually *micro* perspective studies, lasting from one to two hours. Longer-term studies include a *meso* perspective and *macro* perspective studies. Meso perspective studies vary from days to weeks, while the focus in macro perspective studies can vary from years to the whole product life cycle. Macro perspective studies in HCI are rare, but interest towards meso perspective studies has increased since 2006, as implied by the number of events, such as workshops, related to the topic (Karapanos et al. 2012a). The studies included in this thesis represent mainly meso perspective studies.

Longitudinal studies include two or more measurement points, e.g. observations, with the same users and are therefore useful for studying change over time (von Wilamowitz-Moellendorff et al. 2006). If the change over time is studied, some of the dimensions (e.g., users, tasks, products or measures) need to stay constant during the study period. However, in longitudinal studies where the same participants use a product over time, they might learn new ways to use the product, as they get more experienced with it. Therefore, as noted by Novick & Santanella (2012), participants between two measurement points in longitudinal surveys are not exactly the same. How meaningful this is for the validity of the longitudinal research results is an open question. For instance, if the longitudinal research starts from the moment of product purchase and understanding the learning process itself (e.g., how fast the participants will learn to use the product efficiently) is one of the goals of the study, then understanding the reasons why and measuring the change in users' skills over time is reasonable.

An interesting question is, how often and when should one conduct measurements in longitudinal (or repeated measurement) studies? Unfortunately, no common agreement is available in HCI literature on the most beneficial measurement times. According to Mitchell & James (2001), when change is studied, frequent enough measurements are more important than the exact timing of measurements. For further guidance, Ployhart & Ward (2011) advise considering "natural" measurement occasions for the studied phenomenon, interviewing and observing subject matter experts, and reviewing relevant literature from similar phenomena. To provide an example, in a longitudinal study by Kujala and Miron-Shatz (2013), 22 users of a new mobile phone model were followed. A web survey was used to measure users' 1) emotional reactions, 2) behavioral intentions, 3) perceived usability, 4) user experience, and 5) the most memorable experience episodes. A *Day Reconstruction Method* (DRM) questionnaire (Kahneman et al. 2004) was used during the first five days, after which the participants were surveyed on the sixth day, after 2.5 months, and after 5 months of product usage. Another study by Novick & Santanella (2012) used a cross-sectional approach to study what kind of frustration episodes novice and expert users reported with computers. In conclusion, they suggest that most of the differences take place within the first three to six months of product use. Therefore, studies that follow application use over a year were not considered beneficial. However, due to small sample size, more research is recommended to study the most relevant break-point for experience.

One common challenge in longitudinal studies is the participant drop-out during the study period (Jain et al. 2010). In order to keep the effort that is required from the participants in a reasonable level, researchers need to find a balance between the number of measurement points and the amount of work with a single measurement (Jain et al. 2010). In practice, there are several factors that can affect the required number and times of measurements. These may include, among others, available research resources, research interests and measured factors, the estimated length of the product's learning period, and stakeholders' expectations for actionable results. While the HCI research community has addressed the motivation and benefits to conduct longitudinal studies, the practical use of longitudinal research results in product development has received little attention. **Particularly, there is a lack of empirical research on how practitioners in companies utilize results from longitudinal research in their work and how they assess the *usefulness* of such information for product development purposes**. Therefore, more empirical research in different contexts would be an important addition to the body of knowledge of long-term UX and its evaluation methods in HCI.

2.2.4 Long-Term UX Evaluation Methods

HCI literature offers a vast number of available methods for UX evaluation. For example, Vermeeren et al. (2010) collected and analyzed 96 UX evaluation methods from academia and industry to evaluate how they are utilized in HCI field. 34 of the identified methods were reported being able to evaluate the long-term use of products. In another study, Rajeshkumar et al. (2013) created taxonomies for 89 UX evaluation methods. Evaluation methods can help designers choosing the best design, confirming that the design is on right track, or assessing that the designed prototype or the final product meets the UX targets (Vermeeren et al. 2010).

Jain et al. (2010) state that there are no specific methods required for longitudinal studies. Instead, they encourage combining quantitative and qualitative methods. In-situ methods that collect user feedback repeatedly over time, such as diaries (Bolger et al. 2003), *Experience Sampling Method* (ESM) (Csikszentmihalyi & Larson, 1987), and Day Reconstruction Method (DRM) (Kahneman et al. 2004) have been used in longitudinal studies. Retrospective methods such as *Change-Oriented analysis of the Relationship between Product and User* (CORPUS) (von Wilamowitz-Moellendorff et al. 2006), *iScale* (Karapanos et al. 2010, 2012b), *UX Curve* (Kujala et al. 2011), *DrawUX* (Varsaluoma & Kentta, 2012), and *MemoLine* (Sim et al. 2016) can be less taxing for participants when compared with repeated measurements. Retrospective methods are prone to biases as they rely on users' memories of experiences (Kahneman et al. 1993; Schacter, 1999). However, memories can still be important information for product development purposes, since the experiences that customers report to others and the customers' future behavior can be guided by these memories (Norman, 2009; Karapanos et al. 2012b). Karapanos et al. (2012b) suggest that retrospective techniques can be a viable option in studies where memories have higher importance than actuality. Next, the methods that were utilized in studies related to this thesis are introduced.

iScale. Grounded on the theories of the retrospective reconstruction of experiences and episodes from memory, Karapanos et al. (2010; 2012b) developed an online survey tool, iScale, to support respondents in recalling their experiences with a product over time, while minimizing the retrospective bias. With iScale, users are first asked to evaluate 1) the product's evaluated quality (chosen by the evaluator) just before purchasing it and 2) how their opinion has changed since then. Next, users are presented with a timeline from a moment of purchase to present time. Respondents continue by "sketching" a line that consists of linear segments that represent how respondents' perception of the evaluated quality changed over time. For each segment, the participant can add an experience report describing a cause of change in the evaluation. Thus, iScale provides graphs illustrating the remembered changes in experiences and experience narratives that can explain reasons for these changes. Karapanos et al. (2012b) conclude that in comparison with face-to-face interviewing techniques, a structured process for self-reporting, such as iScale, can survey large samples and therefore also "*inquire into rare experiences and atypical behaviors*." Karapanos et al. (2012b) showed that sketching the experience over time can increase the amount and the richness of the information recalled when compared to free recall, where no sketching is involved. However, as noted by Kujala et al. (2011), their study did not provide the interpersonal analysis of the graphs and their trend information, which could have provided information on how the overall evaluation is affected by the chronological order of experiences.

UX Curve. Aiming at the more cost-effective elicitation of longitudinal UX data, Kujala et al. (2011) created a pen-and-paper based method called UX Curve. In comparison to iScale, Kujala et al. (2011) note that UX Curve is designed to be used in face to face setting with the participant, while iScale is aimed more as an independently used self-reporting tool. UX Curve aims to support respondents in retrospectively reporting their experiences with a product. The method aims to support researchers in understanding the reasons why and how the user's experience may have changed over time. UX Curve includes a template presenting an empty two-dimensional graph area for drawing a curve and separate lines that are used for explaining the changes in the curve. The horizontal axis on the graph represents time from the moment of purchase until the current moment, while the vertical axis represents the intensity of the evaluated experiential aspect, such as ease of use. A vertical line divides the graph into positive upper and negative lower parts. In a validation study for the UX Curve method, 20 mobile phone users, with 3 to 12 months of usage experience, reported their experiences by drawing experience curves and describing possible changes in their relationship towards the phone (Kujala et al. 2011). Although curve drawing was considered challenging by some respondents, it was also found interesting and interactive, and all participants successfully drew their experience curves. The exact timing of the remembered experiences was considered difficult, suggesting that UX Curve rather provides the approximate reconstructions of the meaningful events. However, Kujala et al. (2011) argue that these events are important for designers as they can help in identifying issues that create positive experiences and affect customer loyalty. When comparing different curve types, Kujala et al. (2011) concluded that Attractiveness curve provided the largest number of reasons for explaining the change in user experience. Furthermore, the improving trend of the Attractiveness curve was related to the willingness of recommending the product to others, which is an important measure when estimating product growth (Reichheld, 2003).

AttrakDiff and AttrakDiff2. AttrakDiff (Hassenzahl et al. 2003) and AttrakDiff2 questionnaires (Hassenzahl, 2004) were designed for measuring the hedonic (identification and stimulation) and pragmatic UX attributes of a product or service. AttrakDiff2 consists of 21 semantic differentials (word-pairs) on a 7-point Likert scale. AttrakDiff is one of the most used questionnaires for measuring UX (Bargas-Avila & Hornbaek, 2011) and provides feedback on how users perceive the product at the current moment. However, for evaluating the change in UX over time, AttrakDiff can be used in repeated measurement and longitudinal studies to allow comparison between different measurement points over product usage. In addition to the UX attributes included in AttrakDiff, studies utilizing it have also been measuring aspects such as attractiveness (e.g. pleasantness or beauty) and overall goodness of a product (e.g. Hassenzahl, 2004; Kujala et al. 2013).

Although various methods and tools are available for evaluating long-term UX, **little empirical research exists where the usefulness of the results from long-term UX evaluations**

of products and services is studied from the perspective of practitioners in companies. In particular, **few empirical studies have focused on the utilization of retrospective "experience curve" methods such as iScale or UX Curve**. Finally, the utilization of long-term UX evaluation methods together with usage data logging has received little attention in the literature. In the following section, the previous literature in HCI is examined regarding usage data logging and the development of visual data analytics tools with users, with a specific focus on manufacturing automation context.

2.3 Usage Data in HCI

This section first briefly presents the concept of usage data logging from HCI research and UX design perspective. After this, a view is taken to the previous work that presents design approaches, guidelines and relevant case studies related to the development and evaluation of visual data analytics tools with end-users. Finally, the topic is approached from the manufacturing automation perspective, as related to the thesis' case study context, motivating the need for more empirical studies on this domain.

2.3.1 Usage Data Logging in HCI Research and UX Design

Usage data logging can be defined as *data logged from system use based on end-user interactions* (Väätäjä et al. 2015). This can include the features and functionalities of the system used by end-users including the associated metadata (e.g. time, data input, automation state) (ibid.). Data logging provides information on how web applications are used, for instance, and if users' habits in product usage change over time (Jain et al. 2010; Karapanos et al. 2012a). While usage data logging has been utilized in other fields such as marketing and e-commerce (e.g. Lopes & Roy, 2015), in this thesis the focus is on human-computer interaction research and UX design. Usage data logging can provide data to address multiple UX measurement needs in a product development organization, such as identifying 1) changes in user behavior, 2) how device functions are used, and 3) how different user groups access features (Väätäjä et al. 2015; Ketola et al. 2009).

In usability studies, the automatic capturing of user interface (UI) events (e.g. mouse movements and keyboard presses with respect to the application state) has long been considered a fruitful source in understanding product usage (Hilbert & Redmiles, 2000). In their survey regarding "*computer-aided techniques for extracting usability related information from UI events*", Hilbert & Redmiles (2000) provided a categorization of different techniques arranged as a hierarchy. Table 2 provides a summary of their categorization and descriptions of different techniques.

Grimes et al. (2007) summarize several benefits and limitations common to usage data logging in their study for analyzing query logs for search engines. First, data logging does not disturb the user and therefore provides unbiased observational data. Query logs can include a diversity of

Table 2. Hierarchy for computer-aided techniques for extracting usability related information from UI events. Summarized from Hilbert & Redmiles (2000).

Synchronization and searching	Techniques that allow the synchronization or cross-indexing of the UI events with e.g. coded observation logs or video recordings.
Transformation	Transforming event streams through selection, abstraction and recording to support human and automated analysis.
• Selection	Selecting or filtering events or sequences of interest from the "noise".
• Abstraction	Synthesizing new events based on single events or additional contextual information. For instance, single events where input field is edited, a new value provided and user's focus shifted to another UI component could be synthesized as "value provided" events.
• Recording	Producing new event streams from the results of the selection and abstraction tasks. Enables the analysis of the selected and abstracted events with similar manual and automated analysis techniques as performed on raw event streams.
Analysis	Techniques for analyzing event stream data.
• Counts and summary statistic	Utilizing counts and summary statistics to understand user behavior e.g. feature use counts, error frequencies, help system usage.
• Sequence detection	Techniques that allow the detection of *target* sequences within *source* sequences of events. For example, the investigator's goal could be to generate a list of matching event sequences for further inspection or automatically recognize sequences that violate the normal usage of the system.
• Sequence comparison	Techniques for comparing *source* sequences against *target* sequences regarding their similarity. In general, the aim is to identify potential usability issues by comparing "ideal" or expected sequences against actual sequences by users.
• Sequence characterization	Techniques that aim at constructing abstract models for characterizing or summarizing interesting sequential features of the *source* sequences. Examples of techniques include a) generating process models with probabilities associated with transitions (Guzdial, 1993) and b) construction of models characterizing the grammatical structure of verbal interactions among participants in design meetings (Olson et al. 1994).
Visualization	Techniques to present the resulting data after transformations and analyses in forms that allow humans to utilize their innate capabilities for interpreting visual representations. These techniques can support investigators in linking the analysis results back to UI features of the studied system.
Integrated support	Environments that include a collection of different transformation, analysis and visualization techniques provide integrated evaluation support. Some can also include a built-in support for managing artifacts for the specific domain, such as subjects, evaluations, tasks, data and analysis results.

tasks from millions of users, some of which can be even impossible to duplicate in any other data source. However, the log data does not tell what the user meant by the query and if the user was satisfied with the results. While behavioral data from logging can support product designers in understanding *what* the user does with the system, it does not tell *why* the user has made the specific choices (Grimes et al. 2007). Therefore, triangulating logged usage data with subjective data collected with methods such as user observations, interviews or surveys can provide a more holistic understanding of user behavior over time (Shneiderman & Plaisant, 2006).

One advantage of data logging is that it measures users "in the wild", therefore allowing controlled experimentation such as *A/B tests* (Kohavi et al. 2007), where users are randomly exposed to one of the two variants, for example, of the same website. However, logging only captures the use of a specific system, meaning that if the user uses other systems or interacts with other people, this is not evident from the data (Grimes et al. 2007). Grimes et al. (2007) continue that logs can include noise that has to be filtered. For example, query logs may include noise from robots, spam, data outages, and recording errors. Grimes et al. (2007) summarize that logs are mainly beneficial for analyses where a large amount of data is required and for testing the impact of changes.

In research that focuses on mobile device usage in a natural setting, usage data logging has often been combined with other qualitative data collection methods, such as experience sampling (Larson & Csikszentmihalyi, 1983), electronic user diaries, and interviews. Such studies can aim, for example, at a better understanding of mobile usage behavior and interaction patterns (e.g. Ferreira et al. 2014) or to support design and development efforts of different mobile services (e.g. Liu et al. 2010). Recently, Bhavnani et al. (2017) introduced a retrospective methodology where the visualizations of the logged activity were presented to the participant in an interview setting. In addition to logging user interaction activities, the sensing, processing, and storing capabilities of today's mobile devices can support researchers in collecting contextual data regarding the user's location, movement, or basic physical activities (e.g. standing, sitting, running) (Froehlich et al. 2007; Martin et al. 2013). Combining such contextual data with objective user interaction data and subjective data from qualitative methods can provide researchers a more holistic view of user's experiences with mobile technology in a real-life setting.

Game development is a field where understanding users' experiences over time can be beneficial, for example, when the difficulty of the game is balanced to be suitable with the user's learning curve (Hullett et al. 2012). In the recent years, the utilization of game metrics has become an active topic for supporting the data-driven design and development of games (Hullet et al. 2012; Drachen et al. 2013). *Game metrics* are "*quantitative measures of something related to games*", including variables, features or calculated values, such as a number of hits in a shooter game (Drachen et al. 2013). Due to the increased complexity of today's games, the collection of long-term metrics is the only viable mean to understand how players interact with the game over time and to provide a sufficient coverage of the various gameplay states (Hullett et al. 2012). In their article, Hullet et al. (2012) present an analysis of approximately three years of log data collected from an auto racing game *Project Gotham Racing 4* after its release on Xbox 360 in 2007. Based on patterns identified within the data, they identified differences between regular and infrequent players in how they approach the game. Also, several gameplay options were infrequently used by the players. Based on the findings, recommendations for future development were made, including a removal of the rarely utilized options and a more structured introduction to new players to keep them engaged (Hullet et al. 2012).

In summary, interest towards usage data logging has been increasing in HCI and its related fields as the development of data logging technology and fast network connections allow easier access to data that are automatically collected from users. However, **various challenges hinder the utilization of log data in UX design, such as lack of contextual information when**

interpreting the log data, the sheer amount of data generated and skills required for data analysis. Developing user-friendly data analytics and visualization tools can support product developers and designers in interpreting log data collected from users. In the following section, related work is summarized on the development and evaluation of usage data analytic and visualization tools together with users. The topic is addressed from the viewpoint of developing industrial manufacturing automation systems, as the case study (**Study V**, see section 3.2.5) in this thesis was conducted in this context.

2.3.2 Development and Evaluation of Visual Data Analytics Tools with Users

This section provides a summary of the related work that discusses the development of visual data analytics tools together with users. The collaborative development of visual data analytics tools is an approach utilized in **Study V** (section 3.2.5) and one of the contributions from this study aims to support similar development activities. However, the aim here is not to provide a list of currently available data analytic and visualization tools and approaches, but to review the previous work done in presenting design approaches, guidelines and relevant case studies where visual data analytics tools were developed together with users. Finally, previous studies done in manufacturing automation context are highlighted, as related to research done for this thesis.

In this work, the *visual data analytics tool* refers to software applications and web services that offer a variety of data analytics and data visualization features for inspecting log data, with the aim to provide users with insights regarding the inspected data and what it represents. According to categorization done by Hilbert & Redmiles (2000), visual data analytics tools that support the inspection of logged usage data belong to the techniques of *integrated support*, offering a variety of transformation, analysis and visualization techniques (see Table 2).

There are various proposed approaches and guidelines for supporting the development and evaluation of visual data analytics tools with users. Carpendale (2008) presents a good overview of different evaluation approaches and methods for visualizations, while Munzner (2009) provides advice on when to choose between different methods. The choice of an evaluation approach depends on the goals of the study. For example, quantitative laboratory experiments focus on precision, while sample surveys can aim for more generalizable results (Carpendale, 2008). Utilization of qualitative evaluation methods such as the observations or interviews of users, especially during field studies in the real use context, can support evaluators in obtaining a richer understanding of different factors that may influence the development and usage of visual data analytics tools (Carpendale, 2008; Patton, 2001). Carpendale (2008) encouraged that **more studies evaluating visualizations should utilize qualitative methods**.

The Multi-dimensional In-depth Long-term Case Study (MILC) (Shneiderman & Plaisant, 2006) is an approach to evaluate visualization tools with both qualitative and quantitative methods, originating from the studies of creativity support tools (Shneiderman et al. 2006). MILC combines field studies with participant observation, interviews, surveys and automated logging of user activity. Shneiderman & Plaisant (2006) suggest that MILCs can be beneficial for studying the efficacy of novel visualization tools regarding their strengths and for iterating the tool with end-users while providing evidence to warrant further development. The MILC approach and its derivatives (Perer & Shneiderman, 2008) have been utilized to support the development and evaluation of visualization tools for event sequence analysis (Wongsuphasawat et al. 2011) and electronic medical records analysis (Stolper et al. 2014). MILC has been identified as a relevant approach to evaluating how domain experts utilize visual analytics over time (Gotz & Stavropoulos, 2014). Wongsuphasawat and colleagues (2011) found that periodic meetings with a domain expert supported the generation of insights and allowed additional questions and guidance for tool development. Stolper et al. (2014) also demonstrated benefits in utilizing case study approach when documenting insights that were generated during the long-term use of a visual data analytics tool.

Lam et al. (2012) stress that approach to evaluating visualizations should be based on evaluation goals and questions rather than methods. They approach the topic by presenting seven types of evaluation scenarios for visual data analytics tools, based on the literature review of 850 papers from the information visualization domain. The scenarios are categorized into those that can support the understanding of data analysis processes and those that can help the evaluation of the visualizations themselves. One of the presented scenarios is the user experience evaluation of visualizations, where the proposed goal is to understand "*what do my target users think of the visualization?*" (Lam et al. 2012). In addition, Lam et al. (2012) propose the following questions to be considered in UX evaluations of visualization tools:

- What features are seen as useful?
- What features are missing?
- How can features be reworked to improve the supported work processes?
- Are there limitations of the current system which would hinder its adoption?
- Is the tool understandable and can it be learned?

Sedlmair et al. (2012) present a nine-stage framework for conducting design studies and practical guidance for designing visualization systems in collaboration with domain experts. Alongside the framework, they describe 32 design study pitfalls to guide the whole process from learning and designing to reporting design studies. The framework is based on their own experiences and literature review in the fields of human-computer interaction (HCI) and social science. Recently, Crisan et al. (2016) extended the framework by providing practical guidelines that consider also external constraints that can affect visualization design and evaluation. By taking into account the external constraints, regulatory and organizational, visualization researchers and practitioners should be able to improve their visualization solutions regarding their utility and validity, while also improving the likelihood that collaboration with industrial partners is successful.

Several researchers have reported experiences from conducting visualization evaluation in specific work contexts, such as large company setting (Sedlmair et al. 2011), bioinformatics (Saraiya et al. 2006) and game development (Medler et al. 2011). Sedlmair et al. (2011) listed challenges and provided recommendations for evaluating data analysis processes and visualization tools in a large company setting, based on their experiences from a variety of studies in this context. Saraiya et al. (2006) conducted a long-term study with bioinformaticians to evaluate how they use visualizations to gain insights into the data. They emphasize two reasons for conducting long-term evaluation studies in real context instead of short-term laboratory experiments: 1) to recognize the users' natural motivation to do data analysis and 2) the evaluation of the significance of insights. Saraiya et al. (2006) conclude that longitudinal studies make it possible to inspect the long-term insight generation process and identify any long-term usability problems with data visualization tools. Medler et al. (2011) developed a visual game analytics tool *Data Cracker* in collaboration with a game development team who were the target users of the tool. Medler et al. (2011) argue for developing visual data analytics tools in parallel with product development. They propose that visual prototypes are used when discussing how the tool could be beneficial for the development team. Tool designers should keep in mind the broad audience of end-users and encourage users to utilize the tool also after the product release, to provide useful feedback on the development of future products.

To the candidate's best knowledge, few scientific publications exist where visual data analytics tools have been utilized especially for analyzing usage data with industrial manufacturing or related industrial systems. Holzmann et al. (2014) studied how user interaction data from a touch screen based robot controller can be the acquired and visualized, in order to provide cost-efficient solutions for evaluating the usability of handheld terminals in the automation industry. The goal was to support developers by inspecting how users interact with the user interface and by identifying possible issues with the users' workflow, such as navigation problems or unused functions. Based on interviews with a programmer and two project managers from automation industry enterprises, Holzmann et al. (2014) identified *navigation path analysis* and *usage intensity* as the most important topics for data logging in this context. In another study on automation industry context, Grossauer et al. (2015) developed a prototype for visualizing navigation flows through an application. After applying the visualization tool to multiple datasets, they suggest that developers of such tools should provide users 1) a wide variety of filters and 2) views that show the whole navigation data and allow the inspection of individual sequences.

In summary, there is little public research available on the subject of usage data logging in manufacturing automation context, especially regarding how the developers and designers of automated manufacturing systems utilize logged usage data. First, **more information is needed about the expectations, benefits and challenges related to usage data logging in manufacturing automation and related industrial contexts**. Second, **in order to support the utilization of logged usage data with visual data analytics tools that provide positive user experience, guidance to support the development of analytics tools in manufacturing automation context is required**.

2.4 Research Gap

In the following, the gaps in current research in the field of HCI and UX design are summarized, as related to the themes of this doctoral thesis and the research questions.

First, few empirical studies have been published regarding the utilization of UX goals in real-life experience design cases (Väätäjä et al. 2015; Roto et al. 2017). Defining and utilizing UX goals as a part of the product design process can be challenging (Roto et al. 2017). Therefore, empirical research from real-life design cases in different contexts is required to understand how UX goals can be utilized in product development. Such research can also provide guidance for defining and disseminating UX goals in product development, supporting designers, product developers and other stakeholders in industry and academia. This notion motivates the first research question **RQ1. How can user experience goals be defined and communicated among stakeholders in product development?**

Second, long-term UX has been suggested as a key aspect why people keep using and recommending products to others (Kujala et al. 2011). Therefore, ways that can support companies in designing for positive UX over time or evaluating long-term UX with their customers can be considered as valuable topics for research. This motivated the second research question, focusing on the evaluation aspect: **RQ2. How can long-term user experience evaluation support product development?** However, little is known of company practitioners' attitudes towards the usefulness of long-term UX evaluations and applicability of the results of such studies on real-life product development. A need for more empirical research on this topic motivates the first sub-question: **RQ2a. What kinds of perceptions do product development practitioners have about the usefulness of long-term user experience evaluation?** Furthermore, while various methods are available for evaluating experiences with products over time, few empirical studies in particular discuss the applicability of and experiences from utilizing retrospective methods for evaluating users' experiences over time, such as iScale (Karapanos et al. 2010, 2012b) or UX Curve (Kujala et al. 2011). In order to inspect the utilization of UX evaluation methods in long-term studies, the second sub-question is presented: **RQ2b. How can user experience evaluation methods and tools support the long-term user experience evaluation in product development?**

Third, the usage data logging and utilization of visual data analytics tools can be seen as a beneficial approach in several domains for designers and developers to acquire insights on how end-users utilize the developed system. The third research question addresses this topic: **RQ3. How can the utilization of usage data logging be supported in product development?** Currently, little research is available on this subject in the manufacturing automation context. First, there is a need to learn more about the expectations that product developers, managers, and other stakeholders have regarding the benefits and challenges for usage data logging in the manufacturing automation and related industrial contexts. Second, while visual data analytics tools can support the utilization of logged usage data for product development purposes, guidance for tool developers is required to create tools that can provide positive user experience. Such guidance should be based on empirical research in real-life design cases and there is lack of such studies done in the manufacturing automation context.

The research related to the three topics in this thesis, UX goals, long-term UX evaluations and usage data logging, contributes to different phases of the product development process. Guidance regarding the elicitation of UX goals can support development teams at the beginning of the product development process, where understanding of the context and requirements gathering for the designed system takes place. Evaluating the fulfillment of UX goals might require a high-fidelity prototype or even a finalized product that is used over an extended period. In these cases, the long-term UX evaluation methods may prove to be beneficial. The findings regarding a) the perceptions towards the usefulness of long-term UX evaluations, b) the feasibility of specific long-term UX evaluation methods, and c) benefits related to usage data logging in specific contexts, can inform other development teams in choosing suitable approaches and methods when planning the evaluation of their products. Finally, reporting experiences and providing guidance for the development of visual data analytics tools in collaboration with users can support other tool developers working in similar contexts. Over time, this should result in better analytics tools that can support tool users, such as data analysts, UX designers or product developers, in inspecting collected usage data from their systems. In summary, this work can provide support for development teams at the beginning of the development process when the utilization of UX goals, but also a need for long-term evaluations (related to the chosen UX goals) and usage data logging, should be considered. On the other hand, development teams interested in evaluating the long-term UX or logging the usage of their products, be they prototypes or products already on the market, could benefit from this research work by learning from the findings reported in the included studies.

3 Research Design and Methods

This chapter presents the overall research approach in Section 3.1. Section 3.2 provides a description of the research process, research projects where the case studies were conducted and the summary of each case study, including the research contexts, methods and participants. Finally, Section 3.3 presents how the research ethics were considered during the included studies.

3.1 Research Approach

This thesis belongs to the field of human-computer interaction (HCI) that is "*a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them*" (Hewett et al. 2014). The main focus is on HCI's sub-field user experience (UX). **The aim is to understand how product development of digital products, from the perspective of user experience design, can be supported with the utilization of UX goals, long-term UX evaluation, and usage data logging.** The used research approach is *explorative* and utilizes the *embedded single case study research approach* (Yin, 2003) to study the research questions. Case study is a design of inquiry often utilized in HCI and many other fields. Case studies focus on a specific period of time and activities, during which researchers collect information using a variety of data-collection methods (Stake,1995; Yin, 2003). Embedded case studies contain more than one sub-unit of analysis (Yin, 2003). The embedded single case study approach was chosen since during the research period for the thesis the candidate was working in several collaborative research projects between academia and industry, each project involving a different type of context, people and product to study, depending on the collaborating company. **Study I** was a case study summarizing findings from three UX evaluation studies with products from one company. **Study II**, **IV** and **V** were each a case study in the context of one company and its product. **Study III** summarized findings from nine reported experience design cases that had been conducted earlier, independently of this thesis research. This variety in the research projects and their overall goals motivated the candidate to focus on the three topics of this thesis work: UX goals, long-term UX evaluations, and usage data logging. The units of analysis for each case study are described in the section 3.2. (**Table 3**) and in its sub-sections 3.2.1 – 3.2.5.

In those case studies that evaluated product usage, the data reflected real use situations in living or working environments that the study participants reported or that the researchers observed. These studies included interactive digital products aimed for the consumer market, education, and industrial work-environment, but also one non-digital, practical work tool for consumers. Respondents reported their experiences with the products in real use context, which is considered as an essential factor to affect how people experience the use of products and services (Law et al. 2009). This paradigm shift from controlled laboratory experiments to research "in the wild" has been highlighted in the field of HCI during the recent years (Chamberlain et al. 2012).

The research approach was mainly *qualitative*, however, studies that included data logging and scale-based questionnaire questions, provided also *quantitative* data. Qualitative methods included interviews, user observation, and user surveys with open-ended questions. Quantitative data were collected with scale-based survey questions and automatic data logging from the usage of the evaluated systems. The specific focus of this thesis was in the methods supporting retrospective long-term UX evaluation, including iScale (Karapanos et al. 2010; 2012b) and UX Curve (Kujala et al. 2011). In HCI, qualitative research focuses on understanding the qualities of technology and how people use technology in their lives; how they think and feel about it (Adams et al. 2008). Qualitative research aims to answer “how” and “why” questions (Taylor et al. 2016) and is therefore well suited for explorative research that aims at understanding users’ experiences with technology. Finally, the conducted evaluation studies were long-term in nature, including several data collection points over time or retrospective evaluations of experiences.

In parallel with **Study I** and **II**, the candidate initiated a research through design process (e.g. Zimmerman, 2007), resulting in a prototype of an online survey tool DrawUX (Varsaluoma and Kentta, 2012). The design process was inspired by the experiences with UX Curve (Kujala et al. 2011) and iScale (Karapanos et al. 2012b) tools. The DrawUX tool is presented in section 4.2.2.5.

In the following section, the overall research process and methods utilized in each study are described in more detail.

3.2 Research Process and the Case Studies

The research process consists of five case studies, resulting in six publications that are included in the thesis. All case studies were conducted when the candidate worked as a researcher at the Tampere University of Technology, at the unit of Human-Centered Technology. Data collection was conducted during the years 2011-2016, as part of three research projects done in collaboration with industry, as described below. The collaborating product development companies provided the study context, as their products were studied in the projects. An exception was **Study III** where participants were academics who reported experiences from their own design projects although these also included work done in collaboration with industry. In total, 130 individuals participated in the studies, including 124 survey respondents and 6 respondents who participated in both interviews and surveys. Including all measurement points over long-term studies, data from 185 survey responses and 20 interviews were analyzed. Log data was collected and analyzed from 61 participants. Next, the summaries of the research projects are provided to give an overview of the context for the thesis work, followed by the descriptions of each case study included to the thesis.

Research work began in research project **Delightful Long-Term User Experience: Creating Customer Loyalty (DELUX, 2011-2013)**. The aim of the project was to investigate long-term user experience in order to understand how to improve customer satisfaction and loyalty in prolonged product use. The project was funded by the Finnish Funding Agency for Innovation (Tekes) and industrial partners Nokia, Fiskars, Suunto and PAF. The research process for this thesis work began with a focus on long-term UX and aimed to understand the benefits of long-term UX evaluation from product development practitioners' perspective (**Study I**) and explore how different methods support long-term UX evaluation (**Study I** and **II**), especially in remote survey studies. During the project, a parallel design process was initiated for an online survey tool DrawUX (Varsaluoma and Kentta, 2012) (see section 4.2.2.5. for details).

After the DELUX project, the thesis work continued in **Cross-Cultural Design for Mobile Learning -project (CCD MobiLe, 2013-2015)**. CCD MobiLe aimed at developing cross-cultural design practices, tools and guidelines for mobile learning solution development. Research in the project focused on the affect of culture on UX of multimodal and collaborative mobile learning services. The project was part of the TIVIT Education Services Programme funded by Tekes and was conducted in collaboration with companies DIGILE and Nokia (bought by Microsoft near the end of the project). **Study IV** was conducted in mobile mathematics learning context, and while the studied motivational aspects were not in the scope of this thesis, the study offered an interesting case for defining UX goals based on user research and a chance to explore the benefits of usage data logging in this context.

The research work continued in **FIMECC (Finnish Metals and Engineering Competence Cluster) research programme Usability and User Experience in Complex Systems (UXUS, 2010-2015)**. The candidate worked in the UXUS programme during the years 2014-2015, first in parallel with the CCD MobiLe project. UXUS aimed at increasing the competitiveness of Finnish metal and engineering industry by introducing UX thinking for company operations. This included creation of new interaction concepts and innovative practices in developing user and customer experience. In addition to academic partners, companies that participated in the programme included Valmet, Valmet Automation, Rocla, Fastems, KONE, Konecranes, Rolls-Royce and SSAB. In UXUS, the candidate's research work related to the elicitation of UX goals (**Study III**) and supporting the collaborative development of visual data analytics tools for usage data logging (**Study V**). In **Study III** the scope of the thesis was extended to exploring the beginning of UX design process. **Study III** was carried out as a part of an academic conference workshop with a focus on understanding how UX goals are utilized in product design. In **Study V**, the focus was mainly on usage data logging and long-term UX evaluations, this time in manufacturing automation context. While **Study IV** had provided empirical results from the benefits of usage data logging over long-term product usage, with **Study V** the scope of the thesis was further extended towards the quantitative measurements of product usage in long-term evaluation studies.

Table 3 summarizes the relations of research projects, case studies, publications, research questions, main contributions, and units of analysis for each case study. The research questions in this doctoral thesis are:

RQ1. How can user experience goals be defined and communicated among stakeholders in product development?

RQ2. How can long-term user experience evaluation support product development?

RQ2a. What kinds of perceptions do product development practitioners have about the usefulness of long-term user experience evaluation?

RQ2b. How can user experience evaluation methods and tools support the long-term user experience evaluation in product development?

RQ3. How can the utilization of usage data logging be supported in product development?

Table 3. Projects, case studies (C.S.), publications (Pub.), their contributions to related research questions, and units of analysis in each case study.

Project	C.S.	Pub.	Related research questions and contributions	Units of analysis
DELUX	I	P3	**RQ2, RQ2a, RQ2b**. Explores how practitioners in a company developing interactive sports products evaluate the usefulness of long-term UX evaluation and utilize evaluation results in practice.	Practitioners participating in product development
DELUX	II	P4	**RQ2b**. Explores the utilization of AttrakDiff and UX Curve methods in a survey study for long-term UX with a practical, non-digital product.	1) End-users, 2) UX evaluation methods: AttrakDiff, UX Curve
UXUS	III	P1	**RQ1**. Presents the Experience Goal Elicitation Process and instructions for defining UX goals based on two survey studies.	1) Design cases utilizing UX goals, 2) Designers participating in the design cases
CCD MobiLe	IV	P2	**RQ1, RQ3**. Case study presenting a user study for defining UX goals to support developing motivational mobile learning services. Demonstrates benefits of usage data logging in addition to remote user experience evaluation.	End-users of the mobile learning service
UXUS	V	P5	**RQ3**. Presents potential usage metrics to collect and discusses the opportunities of usage data logging from the viewpoints of stakeholders in an industrial R&D company. Provides a set of questions for stakeholders to discuss in companies interested in utilizing usage data logging to support their product development activities.	Company employees participating in the project.
UXUS	V	P6	**RQ3, RQ2b**. Provides guidelines to support usage data analytics tool development in collaboration with end-users from manufacturing automation industry. Describes a visual data analytics tool, UX-sensors, for logged usage data and the process for its development with end-users.	1) Company employees participating in the project, 2) Developers of the UX-sensors tool

The following sections summarize each case study, including the research methods, participants, research process and data analysis. More details are available in the included publications.

3.2.1 Case Study I: Exploring How Product Development Practitioners Perceive the Usefulness of Long-Term UX Evaluation

The first case study **aimed to gain insights on how product development practitioners perceive the usefulness of long-term UX evaluations of their products**. With little earlier empirical evidence available regarding the perceived usefulness of long-term UX evaluation in literature, this explorative study aimed to provide results to fill this gap. As a part of DELUX research project (Section 3.2), researchers at the Tampere University of Technology conducted three long-term UX evaluation studies in collaboration with a Scandinavian company developing interactive digital sports equipment. Summary of these evaluation studies is provided in Table 4, with more detailed descriptions available in **P3**. As the results from the long-term UX evaluations were reported to the company personnel, the candidate carried out surveys where the employees assessed the usefulness of the results of the evaluation studies of their products. Six personnel surveys, including one follow-up survey regarding the evaluation study "Sports Watch B" (SWb) part 2 results, were conducted between October 2011 – January 2014. (**P3**)

Table 4. Summary of the three long-term UX evaluation studies reported to the company personnel in Study I**. Results from the evaluation studies DC and SWb were reported in two parts. (**P3**, p. 81)**

Evaluation study	Number of respondents	Measured product usage period	UX evaluation methods
Diving computer (DC)	Part 1: 33 Part 2: 21	Part 1: Varied from 1 to 5 months Part 2: 6 months	Retrospective: Web survey with Attrakdiff + iScale
Sports watch A (SWa)	25	From 3 to 6 months	Longitudinal/Retrospective: Weekly web survey
Sports watch B (SWb)	Part 1: 111 Part 2: 104	Part 1: From 1-2 to 4-5 months Part 2: From 1-2 to 7-8 months	Longitudinal/Retrospective: Monthly web survey with At-trakdiff

Participants. 30 company employees responded at least in one of the five personnel surveys. In total, 52 responses were collected. The respondents were divided into four categories based on their title and work tasks: 1) Manager, high-level, 2) Manager, 3) Designer/UX Specialist, and 4) Other (e.g. developer). Each personnel survey received responses from one to six respondents from each category, except the "other". Five respondents answered the follow-up survey: two high-level managers, two managers and a quality assurance person. (**P3**)

Procedure and methods. The results of the long-term UX evaluation studies were presented to the company personnel in live meetings by a research team from the Tampere University of

Technology. A contact person from the company provided open invitations to stakeholders prior to each presentation. Each presentation summarized data collected from customers depending on the used survey questions (see **P3** for details). Some examples of the reported results include: 1) background information, e.g. previous experiences from the brand and similar products, 2) expectations before purchase (in retrospect), 3) evaluations of product attractiveness, 4) satisfaction with the product and why, 5) the importance of the product to oneself and why, 6) the willingness to recommend the product to friends and why, 7) AttrakDiff evaluations, 8) iScale curve shapes, 9) positive and negative experiences with quotes, and 10) the design implications based on findings. Session participants were asked to provide feedback by answering a paper questionnaire either during or after each presentation. After each presentation, the presentation and other relevant data files were sent to the contact person to disseminate them in the company. An email invitation to answer the follow-up web survey was sent to all those who participated at least one of the case study SWb's results sessions. Before answering the follow-up survey, respondents were asked to look through the SWb reports. In the surveys, it was inquired how the respondents perceived 1) the results' interestingness, 2) relevancy to one's work, 3) context for practical use, 4) usefulness and novelty value, 5) intention to utilize the results in practice, and 6) the usefulness of the results in practice, when evaluated in retrospect. The rationale for choosing the survey questions is presented in **P3**. (**P3**)

Data analysis. Responses to open-ended questions were analyzed with conventional content analysis (Hsieh & Shannon, 2005). Categorization of the results was done by linking together similar items identified in the responses. Because of the small number of responses to the follow-up survey, no statistical tests were conducted to inspect the possible relationships of the results between the follow-up survey and the previous surveys. (**P3**)

3.2.2 Case Study II: Exploring the Utilization of UX Curve Method and AttrakDiff Questionnaire in a Remote Long-Term Study of a Non-Digital Product

The second case study explored the retrospective reporting of experiences with a non-digital product, non-powered pruning shears, using the AttrakDiff questionnaire (Hassenzahl et al. 2003; Hassenzahl, 2004) and the UX Curve method (Kujala et al. 2011). This case study was conducted at the same time as data collection was ongoing with **Study I**, but focused on the UX Curve method, as it had not been utilized in **Study I**. Although the evaluated tool was non-digital, experiences from the utilization of the UX evaluation methods in a remote study, especially the UX Curve as a novel method, were considered relevant for this thesis. **The aim of this case study was to explore how suitable the used evaluation methods are when used in a remote survey to evaluate experiences with a non-digital, practical work tool after three months of usage.** Suitability was evaluated based on how successfully users could provide feedback with the method and the quality of the feedback. The paper survey was carried out between June – October 2012 and is reported in **P4**.

Participants. Participants were students and personnel from Finnish vocational institutes providing education for horticultural studies. 29 (22 female) persons participated the study (one was dropped out from the original 30). 22 were students and 11 were school personnel or in working life (four were both studying and in working life). Age varied between 20 to 62 years (M=39.9 years). Only one of the respondents had used similar pruning shears before (2-5 times) but all respondents knew the brand. (**P4**)

Procedure and methods. The online survey tool Webropol was used for conducting a screening questionnaire. The chosen participants received the studied pruning shears in the mail and were instructed to use them as if they had bought them for themselves. After three months, a paper questionnaire was posted to the participants, including a UX Curve drawing task, the AttrakDiff questionnaire and questions regarding the willingness to recommend the brand or the product. With UX Curve, participants were instructed to draw an experience curve and add comments depicting aspects related to product pleasantness over three months. (**P4**)

Data analysis. UX Curve forms were scanned to transfer the experience curves into a digital format. Similar to Kujala et al. (2011), curves were categorized for improving, stable and deteriorating. Each curve's start and end point were also categorized as positive/neutral/negative depending on the point being above or below the middle line of the vertical scale that assessed the product pleasantness. Conventional content analysis (Hsieh & Shannon, 2005) was used to analyze the experience narratives that explained changes in the product pleasantness. This process included coding items from the narratives into appropriate categories. The initial framework for the categories was derived from literature. Pragmatic product aspect categories included effectiveness, efficiency, satisfaction (ISO/IEC 9241-11, 1998), and utility (Grudin, 1992). Hedonic product aspects categories included stimulation, identification, evocation, and beauty (Hassenzahl, 2004; Hassenzahl, 2003). One of the authors conducted the first round of categorization, after which the categorization was iterated together with another researcher until consensus was reached. (**P4**)

3.2.3 Case Study III: Online Survey of Design Cases for Developing Experience Goal Elicitation Process and Instructions for Defining User Experience Goals

The utilization of UX goals in product design was one of the central themes in the research work done in the UXUS research project and provided a natural research topic for the thesis work. The third case study extended the focus of the thesis by exploring the elicitation and communication of user experience goals at the beginning of the experience design process. **The aim was to understand how UX goals are defined, what are the characteristics of a good UX goal and how should UX goals be communicated to stakeholders**. The study was carried out between October 2014 – June 2015, following an academic workshop in the NordiCHI2014 conference. The study is reported in **P1**.

Participants. The respondents were researchers from academic institutions and had submitted their case study papers to an experience design related conference workshop. Nine responses were included in the analysis, including people from Sweden, Finland, Germany, and United Kingdom. Five of the nine answered the follow-up survey. Other five responses to the follow-up survey came from three HCI researchers from the Tampere University of Technology and from two of the workshop organizers. (**P1**)

Procedure and methods. The study consisted of four phases: 1) a survey for experience design cases, 2) the analysis of the results and creation of model prototypes and instructions, 3) a follow-up survey, and 4) the iteration of the most promising model and the instructions. An online survey was used to inquire about the experience design case studies: what the targeted experiences were, how they were defined and with whom? Possible insights and inspirations for setting UX goals were inquired based on the five approaches suggested by Kaasinen et al. (2015). Open-ended questions and sentence completion technique (Soley and Smith, 2008) were utilized. Approximately after eight months, a follow-up questionnaire was carried out to gather feedback to further iterate the model prototypes and instructions that were created based on the first survey. (**P1**)

Data analysis. Responses to the open questions were analyzed and categorized by two researchers following conventional content analysis (Hsieh and Shannon, 2005). Based on the results, previous literature, and feedback from HCI researchers from the Tampere University of Technology, a set of instructions for defining and evaluating UX goals and three versions of Experience Goal Elicitation Process model were created. In total, nine different case study descriptions were included in the analysis and as input to the model creation. The reported design cases represented areas such as industry, entertainment, education, well-being, healthcare, marketing, and informatics. Based on the follow-up survey results, the models and instructions were iterated, resulting in the current model and updated instructions. (**P1**)

3.2.4 Case Study IV: Online Survey and Usage Data Logging to Study Motivational Aspects and for Defining User Experience Goals for Mobile Mathematics Learning System

The fourth case study was conducted in the CCD MobiLe research project and **aimed at exploring the motivational aspects of South African learners in mobile mathematics learning context**. Although the motivational aspects are not in the scope of this thesis, **the study provided an interesting case for defining UX goals based on user research and literature**. Furthermore, **experiences from usage data logging during the study inspired the formulation of the third research question RQ3**, further expanding the scope of the thesis. The study was carried out between June 2014 – April 2015 and is reported in **P2**.

Participants. Participants were Microsoft Math service users who had registered to the web service. 53 responses (22 female, 42%) were used in the analysis. Respondents' age was between 14 and 42 years (M = 18.3, SD = 3.87), with one respondent younger than 16 years and three older than 19 years. 45 (85%) were in a primary school or college, 5 (9%) were higher education students, one in working life, and 2 (4%) in other life situation. 68% rated math in general to be easy or very easy, suggesting that the sample was slightly biased towards learners skilled in math. (**P2**)

Procedure and methods. A two-part questionnaire was integrated into the service platform and advertised in the news section of the service. Respondents were instructed to answer both questionnaires in order, after which they could participate in a lucky draw of five free airtime coupons for mobile data. In addition to respondents' background, the questions included topics such as user experience, motivation, behavioral intentions, and use context (see **P2** for question details). The survey was open for one month in December 2014. Log data from the service usage that were used in the analysis included the number of completed separate mathematical quizzes by each participant. (**P2**)

Data analysis. Responses to qualitative questions were initially coded and categorized by one researcher and iterated with another researcher, after which the second round of coding was conducted, resulting in the final categorization of the responses. Quantitative data were analyzed with SPSS software, using appropriate non-parametric tests. Log data were analyzed with SPSS for making sub-group analysis and inspecting the actual usage of the service. Initial UX goals were derived from the analysis results and literature by one of the researchers and then iterated together with another researcher to form a set of three proposed UX goals. (**P2**)

3.2.5 Case Study V: Exploring Perceptions Towards Usage Data Logging and Supporting Visual Data Analytics Tool Development in Industrial Manufacturing Automation Context

The fifth case study was carried out during a collaborative research project with a manufacturing automation supplier company (UXUS project). **The aim was to support the utilization of usage data logging in R&D related activities by developing a visual data analytics tool in close collaboration with the practitioners from the company**. This case study was included to the thesis because it further explored the theme of usage data logging from **Study IV** and offered an interesting case of long-term UX evaluation of a product in development, i.e. the data analytics tool. During the long-term evaluation period of the analytics tool between May – November 2015, repeated interviews and questionnaires were carried out to understand what kind of perceptions the practitioners had towards usage data logging, how practitioners' current work practices related to data logging and what kind of needs they had that usage data logging could support. Tool evaluations with the practitioners were carried out to provide feedback for the developers of the usage data analytics tool. The results summarize the perceptions towards usage data logging in manufacturing automation context and provide guidelines to support the development of visual data analytics tools for logged usage data in manufacturing automation context. (**P5** & **P6)**

Participants. Six employees from the company participated the long-term study, but one developer left the study midway. The respondents represented roles with tasks such as 1) management of research and innovation development, 2) technical customer support, 3) product management, and 4) software development. Respondent details are presented in **P5**.

Procedure and methods. The study process was inspired by the Multi-dimensional In-depth Long-term Case Study (MILC) approach as described by Shneiderman and Plaisant (2006). Over the six-month study period, feedback was collected from the updated versions of the developed tool with several methods, including 1) user observations with interviews and 2) online surveys with the practitioners from the company, and 3) heuristic evaluations with external evaluators. Log data were collected to follow the analytics tool usage over its development period. User observations with interviews provided mainly qualitative data regarding the usability of the tool and practitioners' perceptions towards usage data logging, while also supporting the understanding of practitioners' current work tasks and needs related to data logging. Questionnaires provided quantitative data concerning specific factors related to the tool's user experience over time. Two researchers planned and conducted the data collection during the first two iterations of the tool, after which one researcher carried out the data collection during the following three iterations. (**P5** & **P6)**

Data analysis. Qualitative data were analyzed with conventional content analysis (Hsieh and Shannon, 2005) by coding comments regarding different perceptions towards usage data logging into representative categories. The first set of the guidelines to support usage data logging tool development was created by one of the researchers and then discussed and iterated with the whole research team. (**P5** & **P6)**

3.3 Research Ethics

This study followed the ethical principles of research as advised by the Finnish research ethics authority TENK (Finnish Advisory Board on Research Integrity). TENK's ethical principles for research cover three key areas: 1) respecting the autonomy of research subjects, 2) avoiding harm and 3) privacy and data protection (TENK, 2012). Prior to the data collection, all study participants were informed that the study is voluntary, how the data is anonymized, how the data is stored, and how the results are used. In studies including automated data logging, participants had been informed that their log data is collected from their activities for research purposes. In general, data logging poses several ethical and legal questions that need to be considered. For instance, in **Study V** it was found that there were various privacy, security and intellectual property issues regarding the access to customers' log data that the collaborating company had to consider before the developed usage data analytics tool could be utilized in more extensive manner.

The conducted studies involved no mental, financial or social harm to the participants, which was of special consideration in **Study IV** with minors. All data were stored in a password protected server and the results were reported in a form that does not allow the identification of individuals.

4 Results

This section presents answers to the research questions based on the relevant results of the conducted studies and publications included in the thesis. Section 4.1 presents the results regarding **RQ1** on how user experience goals can be defined and communicated among stakeholders in product development. **P1** and **P2** provide results from case studies that are elaborated to answer **RQ1**. Section 4.2 summarizes the results related to how long-term UX evaluation can support product development (**RQ2**), what kinds of perceptions do product development practitioners have about the usefulness of long-term UX evaluation (**RQ2a**) and how can UX evaluation methods and tools support long-term evaluation in product development (**RQ2b**). Results from **P3**, **P4**, and **P6** are related to these questions. In Section 4.3, ways to support the utilization of usage data in product development are presented. First, the perceived benefits of usage data logging in product development (**RQ3**) are presented based on results from **P2**, **P5**, and **P6**. Second, the guidelines for developing visual data analytics tools for usage data logging are summarized based on findings from **P6**.

4.1 How Can User Experience Goals Be Defined and Communicated?

The first research question (**RQ1**) is "**How can user experience goals be defined and communicated among stakeholders in product development?**" **P1** (**Study III**) summarizes the findings from nine reported experience design case studies, regarding the defining and communication of UX goals. **P2** (**Study IV**) serves as an example case where UX goals are defined based on the previous research and a user survey in a mobile mathematics learning context. Furthermore, **P1** presents the model of an Experience Goal Elicitation Process and instructions to support UX designers and researchers in defining and evaluating UX goals. In **P1**, the term *experience goal* was used as a synonym for *UX goal*, as in previous work by Kaasinen et al. (2015). Therefore, the term *experience goal* also appears in some of the **P1** results. The reported UX goal examples and the Experience Goal Elicitation Process can inform UX designers and researchers who are defining UX goals in collaboration with other stakeholders who participate in the experience design process.

4.1.1 Case-Specific User Experience Goals

This section presents a set of UX goals based on the results from **Study III** (**P1**) and **IV** (**P2**) as **examples of UX goals defined in various design and product development cases**. Furthermore, **a notation for describing UX goals is proposed.**

Table 5 presents the reported UX goals from **P1** and three UX goals from **P2.** The UX goals from **P1** were reported by the survey participants, while the three UX goals from **P2** were formed by the authors (see **P1** and **P2** for details). The reported cases represent areas such as industry, entertainment, education, well-being, healthcare, marketing and informatics. From Table 5, it is evident that the majority of the UX goals were unique to the design cases. However, design cases done in a similar context or utilizing specific technology had related UX goals. Examples of similar UX goals include the sense/feel of control in two industrial cases and the experience of curiosity in two entertainment-related cases utilizing augmented reality.

Table 5. Examples of the user experience goals from case studies in different contexts (P1**,** P2**).**

Pub.	Case context and description	Case specific user experience goals
P1	Industry: Developing paper machine quality control system	1) Learnability, 2) Awareness, 3) Feel of control, 4) Success
P1	Industry: Concept design approach Innoleap for industrial work activity. Example of a ship command bridge design case	1) Being one with the ship and the sea, 2) Feeling of community, 3) Feeling of efficiency, 4) Feeling of trust towards peers, 5) Sense of control
P1	Entertainment: Considering quality of experience in a location-based augmented reality horror adventure	1) Overall experience of curiosity, tension and 'black-humour' horror, 2) Feeling of presence, 3) Speculative play, 4) Support trajectories as journeys through hybrid spaces
P1	Entertainment/Education: Use of mobile augmented reality and outdoor education principles to create something for families visiting a museum	1) Arouse curiosity, 2) Focus on natural and cultural landscape, 3) Communicate author's life and authorship, 4) Support outdoors education, 5) Sustainable experience over time
P1	Education/Well-being: Designing technology to combat (cyber)bullying in classrooms	1) No-blame strategy: not blaming bullies, 2) Positivity, 3) Kind authority, not strict or punishing, 4) Dialogue
P1	Health care/Well-being: Enhancing patient agency in spinal cord injury (SCI) rehabilitation	1) Patient-centredness, 2) Ease of use, 3) Ownership, 4) Network navigation, 5) Projection
P1	Informatics: Enhancing archival UIs with UX techniques	1) Bring user experience of archives closer to modern day web
P2	Education: Designing motivating mobile learning services for developing countries	1) Autonomy, 2) Competence, 3) Efficiency

The exact definitions of the reported UX goals in **P1** were not within the scope of the survey. However, based on the topics of chosen UX goals in Table 5, three of the reported cases in **P1** and the case in **P2** contained usability-related UX goals, including "learnability", "ease of use" and "efficiency". The usability quality could also be described from an experiential aspect, e.g. "feeling of efficiency". There were also goals that did not seem to be directly related to experiential aspects, such as "support outdoors education" or "dialogue". More elaborative descriptions of these goals could have explained how they were related to experiences with the product. Generally, it seems that the set of chosen UX goals in one design case can include both experiential and more practical goals.

To supplement the UX goals described in Table 5, the full definitions of three UX goals from **P2** are provided here as an example. These definitions are based on the fundamental psychological needs discussed by Sheldon et al. (2001) and the identified motivational factors to use the mobile mathematics learning service that was studied in **P2**.

- *"Autonomy* – to feel freedom by choosing when and how to study using the mobile learning service. The *autonomy* goal could be supported with good accessibility and by enabling inexpensive usage. Content should support informal learning for example with clear examples, step-by-step instructions and theory reading." (**P2**)
- *"Competence* – to feel successful by achieving new goals in the learning process. The *competence* goal could be achieved for example by offering sufficient challenge in tasks and emphasizing the personal progress with playful elements such as collecting points, competing with fellow learners and rewarding from achievements." (**P2**)
- *"Efficiency* – to feel that no time is wasted in the use of the mobile learning service. The *efficiency* goal is supported by instant feedback from learning exercises, easy to use user interface and short loading times." (**P2**)

The above notation that was used in **P2** for describing UX goals includes: **1) a short name for the UX goal that is easy to remember, 2) the detailed description of the goal, including a feeling or experience the user should have when using the system, and 3) design implications to support achieving the UX goal**. It is suggested here, that using a predefined notation for describing UX goals can support their clear communication among the product development team. Furthermore, examples of design implications should help concretize the, sometimes abstract, UX goals for the whole product development team. However, in addition to written UX goals, various other ways can also support the communication of the UX goals with the product development team and related interest groups over the product development period, as presented in the following results sections.

4.1.2 Defining and Communicating User Experience Goals

What is a good UX goal? This was inquired in **P1**, where the survey respondents stated **characteristics for a good UX goal.** The following sentence summarizes the respondents' responses:

> *A good user experience (UX) goal is clearly expressed so that all stakeholders understand it in the same way and precise enough so that it can guide design work. The definition should include emotion or the feeling users have while interacting with the product or service. A UX goal should be related to the context of use by coming directly from end-users or be grounded in previous research. It should be possible to evaluate the proposed UX goals to assess how well they are achieved. Finally, while each UX goal should be realistically achievable, ambitious UX goals can provide inspiration and drive creativity.* (**P1**)

Interestingly, one of the respondents in **P1** commented that UX goal does not have to be realistically achievable, as it can still act as a target for design. In this sense, UX goals could act as a source of inspiration, setting a target to strive for although it may not be necessarily achievable with the given resources. Finally, providing practical design implications to support the design process in reaching the proposed UX goals could be advisable, as presented together with the definitions of UX goals from **P2** at the end of the previous section. Although suitable examples from previous research or user studies may not always be available at the beginning of the experience design process of new products or services, such examples could be included later during the design process, as UX designers get more familiar with the context and users. Practical examples may also help in forming a common understanding of the UX goals among stakeholders who participate in the development process.

In **P1**, the respondents were asked **where they got insight and inspiration to define what experience to aim for**? The question was formulated based on the suggested five approaches to gain insight and inspiration for UX goal setting by Kaasinen et al. (2015): brand, theory, empathy, technology, and vision. In the reported case studies in **P1**, "empathy" (7/9 responses) and "visioning" (5/9) were the most common sources. All three industrial cases mentioned "brand" and "vision". This suggests that the industrial cases had specific business purposes, in contrast to more research-driven cases. Respondents in **P1** also reported five "Other" sources of insight and inspiration. These were 1) "the environment", 2) "co-design cued by site visits", 3) "the values of an author to whom the museum is based on", 4) "previous published pilot work related to the user group", and 5) "rules and functionalities created by the paper generation". In the previous study, Kaasinen et al. (2015) included co-design in the "empathy" approach, according to how co-design was originally described by Sanders and Dandavate (1999). Based on **P1** results, respondents combine multiple sources of insights and inspiration when defining UX goals. Indeed, three or more inspirational sources were utilized in seven out of nine design cases (**P1**). In **P2**, the insight for the proposed UX goals emerged from theory (Sheldon et al. 2001) and empathy, based on the user survey results.

Who participates in defining user experience goals? The results from design cases in **P1** show that Researchers (5/9 cases) and Designers (4/9: UX designers, graphic designers, game designers) participated most often in the definition process of the UX goals. Other groups included Developers (3/9), Topic experts or specialists (3/9: an expert panel of educators, outdoor educator specialists, consultants), Management/employees/clients (3/9: museum management and employees, and the spinal injury unit director), and Students (1/9). Interestingly, users participated in the definition of the UX goals only in three cases (education/well-being and two industrial cases). Although the "designer's empathic understanding of the users' world" was the most frequently mentioned source for insight and inspiration (7/9 cases), it seems that users themselves participate in the definition of the UX goals less often. This was also the case with UX goals presented in **P2**, which were based on survey data collected from users, in addition to theories from literature. However, when predefined UX goals such as those presented in **P2** are utilized in

UX design of similar products or services, consulting potential end-users could provide important feedback for UX designers regarding the feasibility of the predefined UX goals for that specific design case.

How can user experience goals be communicated among stakeholders? Table 6 summarizes the results from **P1**, indicating that the most common ways to verbally discuss UX goals were brainstorming sessions, workshops and meetings (3/9 responses). Other ways to communicate the UX goals included written reports and documentation (2/9) and sketch-level scenarios (2/9). For example, in the entertainment/education case in a museum context, "around forty design concepts were sketched down and presented to stakeholders". In the spinal cord injury rehabilitation case, user observations during an ethnographic study were utilized in the creation of "a generalized timeline of a patient's journey through the spinal injury unit". In summary, **P1** survey results suggest that UX goals are communicated between stakeholders at least in writing, verbally and via artifacts, including personas and sketches. (**P1**)

Table 6. Reported approaches to communicate user experience goals (P1, p. 329).

Case context	Approaches to communicate user experience goals
Industry	Documents, Verbal presentations
Industry	Scenarios, Sketching
Entertainment	Audiovisual material, Free play
Entertainment/Education	Bodystorming, Brainstorming, Mood boards, Personas, Scenarios, Sketching
Education/Well-being	Reports, Academic publications
Health care/Well-being	Workshops, Verbally in meetings, Ad-hoc interaction, Patient journey timeline

4.1.3 Process and Instructions for Defining and Evaluating User Experience Goals

As a theoretical contribution to answer **RQ1** and as a summarization of findings from **P1,** the Experience Goal Elicitation Process is presented in Figure 7. The cloud shape on the left presents the potential sources of insight and inspiration, as described by Kaasinen et al. (2015). These sources provide insights and inspiration to the iterative process where stakeholders process the information with different approaches, such as brainstorming or co-design (examples come from the studied cases in **P1**). The sources and approaches overlap since the approaches can also act as "approaches to building sources for insight and inspiration", as stated by one of the respondents in **P1**. The iterative planning process should first result in a tentative list of UX goals, which can be described verbally or through other means, such as sketches or mood boards (to learn about mood boards see e.g. Mcdonagh & Storer, 2015). These tentative UX goals should then be prioritized. Based on the most important goals, a selection of the target UX goals should be made. As the design process proceeds, different means can be utilized to communicate the UX goals for

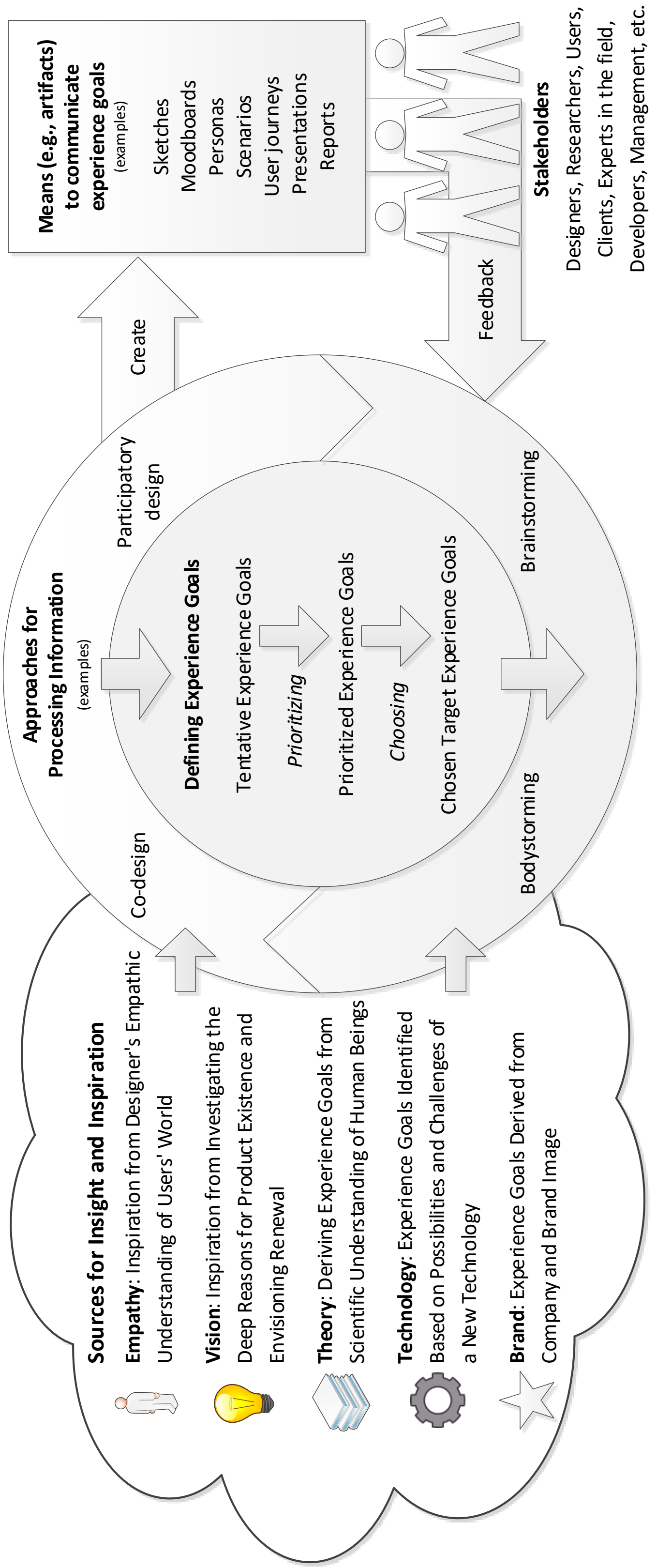

Figure 7. The Experience Goal Elicitation Process model (P1, p. 330). Sources for insight and inspiration from Kaasinen et al. (2015).

all stakeholders, so that common understanding of the targeted experiences can be achieved. Artifacts, such as sketches, personas, and use scenarios, can carry the message and remind all stakeholders of the targeted experiences.

The whole experience design process should be guided by the UX goals defined at the very beginning of the design process. However, in the spirit of the iterative process, the targeted UX goals can be further refined based on feedback from stakeholders and as more data is gathered from the sources of insight over time. (**P1**)

Finally, as a practical contribution from **P1** to answer **RQ1**, instructions for defining and evaluating UX goals are provided in Table 7. These instructions are derived from the results in **P1** and previous literature (Väätäjä et al. 2012; Karvonen et al. 2012a; Karvonen et al. 2014) (see **P1** for details) and are not intended to include all possible steps in the experience design process. Instead, the aim is that UX designers in industry and UX researchers in academia could use these instructions as a general checklist to ensure that in discussions with other stakeholders such as managers, developers, customers or user representatives, the different aspects of good UX goals are considered.

Table 7. Instructions to support designers when defining and evaluating UX goals. (adapted from P1**, p. 331)**

Instructions for defining and evaluating UX goals	
Describe, prioritize & choose	1. Use/choose methods and means to describe UX goals so that all stakeholders can create a shared and similar understanding. 2. Consider possible user requirements connected with the UX goals. You can also describe emotions or feelings the user is aimed to experience. 3. Describe goals precisely enough to make them actionable for designers in the design process. Describe also the reasoning behind the goals (why) as designers need to select the proper means of conveying (how) the experience (what). 4. Prioritize the UX goals to aim for and choose goals that can realistically be achieved (or at least targeted).
Communi-cate & iterate	5. Plan what means (e.g. artifacts) to use to communicate the UX goals for stake-holders. 6. Iterate the goals as you learn more throughout the design process. Revise what de-liverables to use if you find better ways of communicating.
Measure & evaluate	7. If experience is measured, operationalize the UX goals and select appropriate (qualitative) metrics for evaluation. 8. Plan how to trace the later design solutions back to UX goals so that it is possible to evaluate the fulfilment of the goals in different phases of the design work.

4.2 How Can Long-Term User Experience Evaluation Support Product Development?

This section addresses the **RQ2**, "**How can long-term user experience evaluation support product development?**" by answering two sub-questions: "**What kinds of perceptions do**

product development practitioners have about the usefulness of long-term user experience evaluation?" (**RQ2a**) in Section 4.2.1. and "**How can user experience evaluation methods and tools support the long-term user experience evaluation in product development?**" (**RQ2b**) in Section 4.2.2.

4.2.1 Perceived Usefulness of Long-Term User Experience Evaluation

The first sub-question for **RQ2** is "**What kinds of perceptions do product development practitioners have about the usefulness of long-term user experience evaluation?**" (**RQ2a**). Results from **Study I** (**P3**) provide answers to this question based on three evaluation studies with a company developing interactive digital sports products. Longitudinal and retrospective user studies were carried out with the company's customers regarding user experiences with the studied sports products over time (described in Section 3.2.1).The aim was to understand how the practitioners in the company perceived the usefulness of the long-term UX evaluations and the reported results. In five personnel surveys, the perceived "usefulness" of the long-term research results were assessed from different aspects, including 1) the interestingness of the results, 2) relevancy to one's work, 3) context for practical use, 4) usefulness and novelty value, 5) intention to utilize the results in practice, and 6) the usefulness of the results in practice, when evaluated in retrospect. Next, the most relevant results regarding the perceived usefulness of the results from the long-term studies in **Study I** are summarized. For more detailed results, see **P3**.

What kind of long-term UX information was considered interesting? Respondents were keen to *compare the reported results with their own expectations*, as one designer commented: "Expected results. Good that initial 'hunch' of UX predicted results were aligned with actual results". *Positive and negative experiences*, with *quotes from the users* were found interesting, as well as *changes in UX (if any) over the measurement period*. Other topics of interest included 1) *user satisfaction with the product and overall UX*, 2) *UX and problems with specific components*, and 3) *factors affecting UX*. Interestingly, two managers (one high-level) in different sessions commented that the findings were already known to them, mainly from other sources. (**P3**)

What kind of information in the long-term UX evaluation results was considered relevant to the practitioners' own work? This was inquired in the first four results report sessions. Three of the most repeated responses were: 1) *all results/user feedback and anything related to UI or UX were relevant*, 2) *the results show where to focus when prioritizing development activities*, and 3) *positive or negative user experiences and user feedback were relevant*. In three different sessions, one respondent with Design/UX position answered "*how the product is taken into use"* and one manager responded "*Long-term UX*". It is also noteworthy that three different respondents found results confirming their expectations or current understanding of the topic, which were based on results from other sources. This suggests that the rigorous research approach was considered useful in validating findings from other sources such as customer care although some of the results of the long-term studies could be considered "outdated" when they were reported. (**P3**)

Where the information that was considered relevant could be used? The most often mentioned use for the results related to *proposing, conceptualizing or defining new products*. As one manager commented: “Good points for defining product specifications and requirements before actual development project.” Respondents also suggested that the relevant findings could be used to *update and develop existing products further* and *in future product development by fixing the identified issues*. Four responses related to *comparing the results with findings from other sources* and *guiding further user research*. The rest of the findings with two responses each included the *design of new products*, *UI design*, *UX design*, *marketing communication*, *understanding customers better*, *product and service integration*, and *system testing*. (**P3**)

The perceived usefulness and novelty. Perceptions of the usefulness and novelty of each results section (see Section 3.2.1) were measured also with two Likert scale questions, using a scale from 1 (Not at all useful/novel) to 5 (Very useful/novel). These questions were asked only during the final reporting session of the third evaluation study (SWb part 2, see Section 3.2.1 for description) to gather quantitative data to supplement the open comments from nine respondents. The presented results sections included, among others, 1) AttrakDiff measurements over time; 2) quantitative analysis results, such as how emotions related to the willingness to recommend, satisfaction with the product, and AttrakDiff measurements; and 3) summary of positive and negative experiences that users reported over time. All results sections are listed in **P3** (Table 3). In general, the usefulness of the results was rated higher than their novelty value. Except for one results section, all usefulness mean values were between 3.9 and 4.1, while mean novelty values were between 3.1 and 3.4, above the scale average (3). The only exception were the quantitative analysis results showing how, for instance, emotions related to AttrakDiff measurements. These were considered the least useful (M=2.9), but also the most novel (M=4.0). Based on the open comments, the time allocated to the presentation and the presentation content were not enough to properly communicate the findings from the quantitative analysis, as one designer commented: “This part was a bit difficult to comprehend and would have needed a little bit more practical explanation. Still, the content was interesting”. Overall, the novelty values had slightly higher standard deviations than the usefulness values, indicating that perceiving the results from the reported long-term UX study as useful was more common among the respondents than finding them novel. (**P3**)

The intention to utilize the results in practice. All the five personnel surveys included the question: “How likely will you utilize the presented results in your own work? (Not at all likely 1-7 Very likely)”. All the mean scores were above average, suggesting that majority had intention of using some of the results in their own work. Evaluation study SWa had the lowest mean score (4.89) and the highest standard deviation (1.45). Two managers who gave the lowest scores (3 and 4) commented that a) the product was “not in the core of my responsibilities” and b) similar feedback had been received from other sources. (**P3**)

The actual utilization of the long-term evaluation results. Nine months after the SWb part 2 results presentation, the participants were asked if they had actually utilized any of the SWb part 1 or part 2 results in their work. Responses to the follow-up survey were received from two high-level managers, two managers and one quality assurance person. One manager and the quality assurance person had not utilized the results. The manager stated that there had not been enough time or a suitable project where the UX evaluation results could have been utilized. The quality assurance person replied that the SWb part 1 presentation already gave all the information he needed and that he was not even aware of the SWb part 2 results that were provided to the company nine months earlier. The three other participants, manager and two high-level managers, had been looking for information about: 1) "consumer long-term usage" and "using the findings for future work" (Manager); 2) "who buys the product and why", "who are our users and what they are experiencing", and "what kind of products we should make" (Manager, high-level); and 3) "enhancement ideas", "feature priorities", "usability pros and cons", and "as motivational feedback to development team to help them understand how important different UX aspects are" (Manager, high- level). (**P3**)

What kinds of perceptions do product development practitioners have about the usefulness of long-term user experience evaluation? Finally, the findings from **P3** are summarized to answer **RQ2a**. The three evaluation studies were conducted in one company, but included different products and employees in different roles. However, the findings, discussed below, should be general enough to apply in other R&D companies developing digital products. These findings can provide value for practitioners, e.g. UX designers, working in similar companies as well as academics, e.g. UX researchers, working in collaboration with these companies.

The findings suggest that long-term UX evaluation studies can provide results that are perceived as interesting, relevant, and useful by practitioners, i.e. managers and designers/UX specialists, developing digital sports products. To summarize, the respondents found the long-term UX evaluation results relevant in **1) comparing the results with previous knowledge, 2) understanding the change in UX over time, 3) focusing future work, 4) conceptualizing and developing future products, and 5) updating current software products**. (**P3**)

Most practitioners in the studied company had high intentions of utilizing the results of the long-term evaluation studies in their own work. However, **lack of time** and **opportunities to utilize the information**, or simply **not being aware of the available information** can be reasons for not utilizing the results in practice. Furthermore, the novelty value of the long-term UX evaluation results can be diminished by **similar findings from other sources**, such as feedback received by customer care. Despite the findings being "outdated" due the long study period that the long-term evaluations require, it seems that **the results of carefully planned long-term studies do have value for practitioners by confirming their own expectations and subjective understanding of the topic**. (**P3**)

Long-term UX evaluation results seem to have the most value in proposing, planning, and developing future products. However, companies developing software products that can be updated after their release on the market may also benefit from findings from long-term studies. **As software updates generally alter the product in some way, these can be interesting measurement points in long-term UX studies**. Finally, based on a comment by UX designer from the collaborating company, studies such as the six-month SWb evaluation study (see Table 4) would have been difficult for the company with their internal resources. This suggests that lightweight approaches are needed for long-term UX evaluation studies as practitioners need to carefully plan how to use their available resources.

4.2.2 User Experience Evaluation Methods and Tools Supporting Long-Term User Experience Evaluation

The second sub-question to **RQ2** is **RQ2b**: **How can user experience evaluation methods and tools support the long-term user experience evaluation?** The question is answered by summarizing the experiences from the UX evaluation methods and tools that were used during the long-term UX evaluation studies included in this thesis. Therefore, the aim is not to provide an overview of all possible evaluation methods and tools for long-term UX evaluation but to share the insights and experiences from utilizing the specific methods and tools during this dissertation process. The conducted long-term studies include the viewpoints of product development practitioners in a company (**P3**), the experiences during an iterative development of a software product (**P6**), and studying the applicability of UX evaluation methods for a practical, non-digital work tool (**P4**).

Table 8 summarizes the products, respondents and the used methods or tools from those case studies where long-term UX was evaluated. More detailed descriptions of the used methods and tools are presented in sections 2.2.4, 3.2.1, 3.2.2, 3.2.5 and in publications **P3**, **P4** and **P6**.

Table 8. Summary of the studies where UX was evaluated over time.

Case Study	Pub.	Evaluated product	Respondents	Evaluation approach and used methods/tools
I	P3	Diving computer (DC)	Customers who had purchased the products themselves	DC: Retrospective: Online survey with AttrakDiff + iScale survey tool.
		Sports watch A (SWa)		SWa: Longitudinal / Retrospective: Weekly online survey
		Sports watch B (SWb)		SWb: Longitudinal / Retrospective: Monthly online survey with AttrakDiff
II	P4	Non-powered pruning shears	Horticulturalist students and school personnel who received the product	Retrospective: Paper-based survey with AttrakDiff2 + UX Curve method.
V	P6	Usage data analytics and visualization tool UX-sensors	Employees of a company developing flexible manufacturing systems	Longitudinal / Retrospective: User observations and interviews + Online surveys after each iteration

In order to answer **RQ2b**, Table 9 summarizes the benefits and challenges that were identified with each method regarding its feasibility in long-term UX studies. The table combines experiences from altogether six researchers, including the candidate, working in three case studies, where long-term UX evaluation methods and tools were utilized. The methods and tools included: user observations and interviews, surveys on paper and online, the AttrakDiff questionnaire (Hassenzahl et al. 2003; Hassenzahl, 2004), and experience curve drawing tasks in remote studies with the iScale survey tool (Karapanos et al. 2012a) and the UX Curve method (Kujala et al. 2011). The results have novelty value in reporting the first time when the iScale tool and the UX

Table 9. Identified benefits and challenges with the utilized methods and tools during long-term UX evaluation studies. Based on Studies I, II and V.

Method or tool	Benefits	Challenges
User observations and interviews	- Rich source of qualitative data with audio and video recordings - Opportunity to ask clarifying questions when required - Possibility to refer to previous discussions in repeated sessions	- Repeated sessions can be taxing for participants and researchers - Arranging suitable meeting times for several participants - Time consuming analysis of qualitative data
Surveys	- Flexible, provides both qualitative and quantitative data - Several UX surveys readily available in literature - Participants can answer on their own time - *Paper surveys* do not require familiarity with computers - *Online surveys* are generally faster and cheaper than paper surveys, as results are readily in digital format for analysis	- Not possible to ask clarifying questions - Reminders may be required during repeated measurement studies. Email reminders can go missing in cluttered mailbox. - With *paper surveys*, printing and posting questionnaire forms and transferring responses to digital format for analysis can be time consuming - *Online surveys* require some familiarity with computers from evaluators and respondents
Attrakdiff questionnaire (Hassenzahl et al. 2003; Hassenzahl, 2004)	- Measures various product characters in terms of UX - Allows comparisons of measurement points over time	- Some word pairs may not be fitting for all product types (e.g. non-powered hand tools) - Does not provide reasoning for the ratings - No translation available in Finnish
Experience curve drawing task	- Provides both qualitative and quantitative data regarding the experiences over time in one evaluation session - Provides visualization and reasoning of the changes in user's experience over time	- Instructing "curve drawing" tasks for respondents in remote studies - Only one UX quality attribute measured per curve - Analyzing, visualizing and comparing the resulting experience curves requires additional tools and can be challenging, especially with many responses
iScale survey tool (Karapanos et al. 2012)	- Results in digital format - Grounded on theories of the retrospective reconstruction of experiences from memory	- May require some familiarity with computers
UX Curve Method (Kujala et al. 2011)	- Pen-and-paper style answering does not require familiarity with computers	- Interpreting unclear curve drawings or handwritten experience narratives - Transferring curves and comments to digital format takes time and requires additional tools

Curve method are used as a part of remote survey studies. Finally, a description is provided of the DrawUX tool (Varsaluoma and Kentta, 2012), an online survey tool which development was inspired by the experiences with iScale and UX Curve in remote studies. The following sections present the summaries of the experiences regarding each method's potential in supporting long-term UX evaluation in product development.

4.2.2.1 User Observation and Interview

In **Study V** (**P6**), observation and interview sessions were conducted with company employees who participated in the collaborative development project of a usage data analytics and visualization tool *UX sensors*. The analytics tool was developed in an iterative manner and feedback was collected from the practitioners after each update. Observing and interviewing participants over the one-hour sessions provided useful feedback for the developers of the UX sensors tool regarding the usability issues and suggestions for new features that could be considered in the future iterations of the tool. In the following sessions, it was possible to inquire if the updated tool met the users' requirements and refer to the discussions with the participants from earlier meetings. Although arranging the sessions with the busy company employees and the analysis of the observation and interview data required time, these meetings with the users were the most fruitful approach from all that were used in understanding how the practitioners evaluated the UX sensors tool over time.

4.2.2.2 Surveys

In **Study II** (**P4**) a remote study was conducted by using a **paper-based survey** that included a UX Curve drawing task for the participants. Although this was a single measurement study, the preparing and sending the paper forms and transferring the results to the digital format required significant time from the researchers. In this sense, it can be argued that paper-based surveys are not ideal for remote repeated measurement studies when compared with online surveys. However, paper surveys can be more suitable for participants who have no access to or have little experience with computers.

Online surveys to study long-term UX were used in **Study I** and **V**. While the first evaluation study (DC) in **Study I** was a single measurement study, all the other online surveys were repetitive measurement studies (see Table 8). Based on the experiences, online surveys were an easy and fast way to collect both qualitative and quantitative feedback repeatedly from users over time. As the structure and questions remained mainly similar to each measurement, the first questionnaire could be copied to prepare additional questionnaires. When necessary, SMS messages were used as reminders in addition to the emails that included links to the online questionnaires to get the participants to response approximately at the same time. Sometimes participants missed the emails and had to be contacted with alternative methods, such as a phone call, to check if they were still participating in the long-term studies.

4.2.2.3 AttrakDiff Questionnaire

AttrakDiff (Hassenzahl et al. 2003) and AttrakDiff2 questionnaire (Hassenzahl, 2004) were utilized in **Study II** and in two case studies during **Study I**. AttrakDiff offers an off-the-shelf questionnaire to measure hedonic and pragmatic UX attributes of a product or service. AttrakDiff consists of 21 semantic differentials (word-pairs) on a 7-point Likert scale. As a quantitative method, in repeated measurement studies it provides an easy way to compare the changes in the evaluated UX quality attributes over time.

During **Study II**, the researchers participated in a workshop with Finnish HCI-experts to translate the AttrakDiff word-pairs into Finnish, so that they could be utilized also in the planned study. This was expected to be necessary, as the respondents were not likely to be familiar with all the English terms used in Attrakdiff. In **Study II**, AttrakDiff was found to be a viable method when evaluating the user experience of a non-digital, practical product, such as pruning shears. However, during the analysis of the results, it was learned that some word-pairs (*lame-exciting* and *easy-challenging*) had to be removed from the hedonic quality stimulation results based on Cronbach's alpha. Furthermore, a word-pair *Takes me distant from people - Brings me closer to people* resulted in some amused comments from respondents when such practical tool was evaluated. To conclude, when using AttrakDiff for evaluating experiences with non-digital, practical products, discretion should be used when analyzing the results, as all word-pairs may not be suitable for the evaluation of such products.

Finally, AttrakDiff only answers the question how the product is experienced at the moment of evaluation and does not provide reasons for the ratings. Therefore, for product development purposes AttrakDiff should be accompanied with qualitative feedback methods, such as open-ended survey questions or user interviews, in order to get a more comprehensive understanding of the ratings or changes in the experience over time. For example, in a large online survey with hundreds of respondents, AttrakDiff questionnaire could reveal users who should be contacted and inquired for additional details regarding their experiences.

4.2.2.4 Experience Curve Drawing Task with iScale and UX Curve

The **iScale survey tool** (Karapanos et al. 2012b) was utilized in the first evaluation study (DC, see Table 8) of **Study I**, to retrospectively measure the customers' experience over time and to collect experience narratives from the experiences with a diving computer. This also acted as a pilot study to utilize iScale as a part of a remote survey. After replying to an online questionnaire, the respondents used a link to access a separate iScale website. In iScale, users first rate their opinion on the evaluated product at the moment of purchase and then graph linear segments on a timeline in a serial order that represents their experiences from the moment of purchase until the present. For each segment, the respondent can add an experience narrative as an open comment to describe his or her experiences regarding the possible change in the measured attribute. In **Study I**, "product attractiveness" was used as an experience attribute on the y-axis (More attractive – Less attractive).

During the remote study with iScale, some challenges were faced. Although specific on-screen instructions had been added for the iScale tool based on a pilot study, these seemed inadequate when the tool was used remotely without a researcher to instruct its use for the respondents. First, during analysis, it was noticed that 69% of the timelines finished with the end-point at the center of the y-axis, suggesting that the present experience had not changed from the moment of purchase. However, it seemed that respondents had not noticed that before sending their response, they had to connect their line to the end-point and then move the end-point vertically on the y-axis to depict their present experience with the product. Second, while respondents who had successfully added experience narratives to their timelines had done it well and explained their experiences with care, two-thirds of the respondents had either failed or decided not to add experience narratives. In the end, the missing experience narratives were inquired via email by sending images of the plotted timeline figures to the respondents. These timeline figures were created with MS Excel based on the exported timeline data from iScale. Third, although respondents could rate their experiences at the moment of purchase differently, in iScale the visual timeline always started from the center of the y-axis. This made the visual comparisons of the timelines between respondents difficult, as the starting points and therefore the scale on y-axis had different meaning based on the rated experience at the moment of purchase. Therefore, when reporting the results, it might be more sensible to report the timelines separately from each other. If timelines from different respondents are presented together, it should be emphasized that although all timelines start from the same spot on the y-axis, the measured experience attribute on the y-axis is in relation to how each respondent rated their experience at the moment of purchase although this may not be evident from the visualization.

In conclusion, it is proposed that clear user instructions and carefully controlled steps in the reporting process are required before digital graphing tools for the reconstruction of experiences, such as iScale, are suitable for remote studies. However, if the tool is used alongside face-to-face interviews, for instance, where its use can be instructed, the current version of iScale can prove more useful. Indeed, the results that iScale provided, including “positive and negative comments”, “quotes from users”, “long-term UX of product” and “changes in UX over time”, all were in the top three of the topics of interest with product development practitioners in **Study I**.

The UX Curve method (Kujala et al. 2011) was used as a part of a paper-based survey in **Study II** (**P4**), where respondents reported their experiences with pruning shears after three months of usage. To the best of the candidate’s knowledge, **P4** is the first report of a study using UX Curve in a remote setting. The participants were asked to first draw and then add written comments on an “experience curve”, depicting their evaluation of the product’s pleasantness over time. At the end of the survey, respondents could provide open feedback about the curve drawing task. Based on the feedback, 19/29 (66%) of the respondents found the task easy to do or clear. One respondent found the task fun and two appreciated the free style of reporting experiences. The main challenge that 9/29 (31%) of the respondents mentioned was to remember all their experiences or their correct order after three months. (**P4**)

The current findings add to the previous study by Kujala et al. (2011) with mobile phone users, where user satisfaction and product recommendation were related to the shape of the Attractiveness curve. The current study with pruning shears (**P4**) suggests that **the shape (improving/deteriorating) of the Pleasantness curve (i.e. “how good the product is or feels”) can relate not only to the willingness to recommend the product, but also how users evaluate**

the overall goodness and the hedonic quality identification (i.e. how the product is enabling user to express one's personal values to other people) of the product. (**P4**)

In comparison to using the iScale tool in remote studies, pen-and-paper style answering with the UX Curve method seems to be easier for the respondents to comprehend. However, interpreting hand-written comments and time needed for transferring the responses and the drawn curves into the digital format for analysis and reporting is a significant downside of UX Curve. Although originally intended as a single measurement retrospective method, it would be interesting to study if UX Curve could be utilized also as a repeated measurement method, where users add to the curve daily or weekly when using a product. However, from the product development perspective, a digitalized approach to experience curve drawing seems more practical than a paper-based approach, as having the data in digital format can result in considerable time savings when collecting, analyzing, and reporting the results, especially with larger sample sizes.

4.2.2.5 DrawUX Survey Tool

Experiences with the iScale survey tool (**Study I**) and the UX Curve method (**Study II**) during the DELUX project (see Section 3.2) resulted in the planning and development of DrawUX[1] (Figure 8), an online survey tool (Varsaluoma and Kentta, 2012). DrawUX supports reporting experiences with both retrospective and repetitive measurements approach in the long-term studies of products and services by combining basic online survey questions with experience curve drawing tasks inspired by UX Curve (Kujala et al. 2011) and iScale (Karapanos et al. 2012b). Although currently there are no scientific publications of long-term UX evaluation studies that utilize the DrawUX tool, it is briefly presented here as it represents a practical contribution originating from **Study I** and **II**. Furthermore, DrawUX relates to **RQ2b** as an online tool that can potentially support product development practitioners and researchers in the long-term UX evaluation studies of products and services. The candidate was responsible for the design

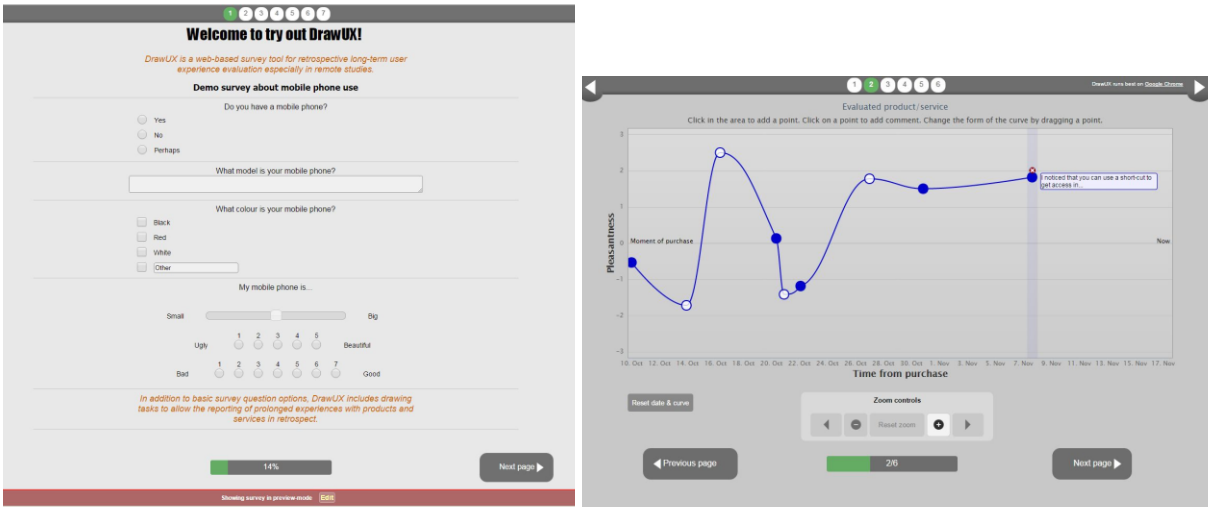

Figure 8. Screenshots from the DrawUX survey tool.

[1] DrawUX prototype version: https://drawux.cs.tut.fi/ (access date 2018 September 12)

of the tool and supervised its development that was conducted by research assistants working in the same research projects as the candidate.

The following features were designed to make the DrawUX survey tool suitable for long-term UX evaluation of products and services for practitioners, such as UX designers working in product development companies, as well as UX researchers in academia interested in long-term UX studies:

- Allows single-measurement (retrospective) and repeated measurements (diary-type) surveys. In repeated measurement surveys the respondent can answer the selected questions repeatedly to e.g. evaluate a product or service over time.
- Allows the combinations of different question types and experience curve drawing tasks, for example for collecting background data and experience curves with a single survey.
- When preparing the survey, the researcher can modify the experience curve drawing tasks, including the rating scales, all visible texts and the number of required experience narratives per curve. This allows e.g. the translations of the drawing tasks for respondents from different language areas and prevents respondents from submitting curves without experience narratives.
- Data Viewer for viewing all response data, including selections of experience curves and their comments, calculating an average curve, filtering the curve data and exporting curve figures (or all response data) for further analysis or reporting purposes.
- Translation tool for adding translations for open-ended questions and experience narratives e.g. for exporting the experience curve figures with translated narratives.

Thus far, a pilot study with a collaborating company where a group of customers used DrawUX to report their experiences with a web-based service resulted in positive feedback. However, more research is required to evaluate the feasibility of the tool in actual product development work. Still, the **DrawUX survey tool can be seen as a practical contribution potentially supporting UX evaluation work**, having novelty value in its features that support a) respondents in reporting their experiences over time in remote studies and b) evaluators in analyzing and reporting the results from the experience curve drawing tasks.

In the following section, ways how to support the utilization of usage data logging in product development are discussed. The results include the identified and expected benefits of usage data logging and guidelines to support the collaborative development of visual data analytics tools with users in manufacturing automation context.

4.3 Supporting the Utilization of Usage Data Logging in Product Development

The third research question is **RQ3: How can the utilization of usage data logging be supported in product development?** The answer to **RQ3** is based on the findings and experiences concerning usage data logging in **Studies IV** and **V** as reported in **P2**, **P5**, and **P6**. The service and products which usage was logged in these studies included a) a mobile mathematics learning service (**P2**), b) flexible manufacturing systems (**P5** and **P6**), and c) a visual data analytics tool for logged usage data (**P6**). In the following sections, the findings are presented regarding the expected and identified benefits of usage data logging for product development purposes. After this, a set of questions is provided to inspire product development teams when they consider the utilization of usage data logging to support their activities (**P5**). Finally, a summary is provided of the guidelines from **P6** to support the collaborative development of visual data analytics tools for usage data logging, derived from a case study in the manufacturing automation context.

4.3.1 Perceived Benefits of Usage Data Logging for Product Development

Table 10 summarizes the main benefits that the company practitioners in **Study V** expected usage data logging to provide in future, supplemented with the authors' experiences from **Study IV** and **V**. Some of the main benefits of usage data logging are that it **shows what users really do with the product** and it **is not dependent on the users' recollections of their actions**. For example, in **P4**, it was noticed that some users who stated that they were very interested in using the mobile learning service had not actually used the service that much. Furthermore, as reported in **P5**, the customer support team of a flexible manufacturing systems supplier can get a better understanding of the chain of events as logged usage data shows what the user has done during complex error events, while the user may have difficulties in remembering all the actions in detail.

Table 10. Summary of the identified or expected benefits of usage data logging for product development based on Study IV (P4**) and Study V (**P5 **and** P6**).**

Study	Identified or expected benefits of usage data logging
IV & V	Shows what users really do, instead of what they tell they do. Can be used to verify users' retrospective reports of product usage.
IV & V	Automatically done usage data logging during long-term studies does not disturb users in a way other data collection methods would.
V	Analysis of usage data over months or years can reveal e.g. interesting usage patterns or user activity that can justify or guide additional, more qualitative user studies.
V	Supports continuous user interface development e.g. by informing developers on how the product is used after major software updates or by revealing inefficient ways of using the system, suggesting a need for UI changes or user training.
V	Provides new business opportunities. Examples from the manufacturing automation domain: improve customer support services (e.g. fault diagnosis), new customer training offers, additional value from customer reports, and evidence for accidents.

Usage data logging happens automatically in the background and therefore **does not disturb the user** in a way as replying to surveys or participating in interviews would. **Usage data logging can offer a substantial amount of data from months or even years of system usage**. However, as log data only tells *what* the user has done but not *why*, the **detection of interesting patterns**

or usage situations in the data could direct and justify the allocation of resources for conducting additional studies with a more qualitative approach. For example, in **P6** it was possible to identify participants who were less active in utilizing the developed data analytics tool during its collaborative development period, therefore prompting us to inquire reasons for this during interview sessions. Furthermore, in **P5**, developers saw **major system updates as fruitful instances for collecting log data to see if users take advantage of any new features**. Seeing how different features are used and what usage patterns emerge may raise more questions that require interviews or observations to understand reasons for users' behavior.

In **P5**, **new business opportunities** that usage data logging could provide are reported. As discussed above, the logging of usage data was expected to support **continuous user interface development** and **improve the quality of customer support services** in the context of flexible manufacturing systems. Other potential benefits included a) **opportunities for customer training offers** based on identified inefficient use or repeated error situations in the customer's site, b) **customer reports** that provide additional value by periodically reporting the summarizations of system usage, and c) **evidence for accidents** in cases where the liability of the damage is uncertain. (**P5**)

In summary, to answer **RQ3**, several benefits that usage data logging could provide for product development purposes have been presented above. However, for product development purposes, it is suggested that usage data logging is utilized together with other, qualitative data collection methods for acquiring a more holistic understanding of the product use, especially regarding the reasons for users' activity and users' experiences with the product or service. Lastly, the focus in **RQ3** has been on the possible benefits of usage data logging. However, resources and skills are required for tasks such as setting up the logging capabilities, acquiring needed permissions for data logging, and analyzing the log data before usage data logging can show its potential in supporting product development.

As a practical contribution to support development teams to consider the benefits and requirements for usage data logging, in **P5** it is proposed that stakeholders in manufacturing automation companies consider the following questions if they are interested in usage data logging (**P5**, p. 433):

- **Possibilities**: What type of usage data and related data from the system and context can be logged?
- **Goals**: What do we want to learn from these data? What value can these data provide and for whom?
- **Data analysis**: Who has the requisite skills and context knowledge to analyze the log data?
- **Data access and security**: How can the log data be accessed? Who owns the data? How will the data be transferred and stored? How do we ensure security?
- **Tools and data wrangling**: Which data analytics and visualization tools are suitable for the needs of different stakeholders? How much data wrangling (i.e. transferring

and mapping the "raw" data to suitable form) is required to import the raw log data to these tools?

Although based on a case study in manufacturing automation context, the proposed questions should support practitioners working also in other contexts, where the logging of end-user interactions is possible. Examples of such product contexts would be various web-based services, mobile systems or industrial systems. Discussing these topics with stakeholders representing different roles in the company, such as management, product development, design, customer support and marketing, should ensure that everyone's needs are understood, resulting in a more versatile utilization of logged usage data. In cases where customers or end-users might find value in logged usage data, their representatives should also be included in the discussions. With a systematic process in place that supports effortless logging and then importing usage data to an easy to use analytics tool, it is more likely that stakeholders with less experience with analytics would also utilize usage data. For instance, as was learned in **P5**, the key usage metrics such as "the use frequency of features" and "user actions before and after specific events" interested most respondents, while more complex usage data, such as event sequences, could benefit those who were already more familiar with data logging and analytics. (**P5**)

Developing visual data analytics tools for logged usage data in close collaboration with end-users, e.g. representatives of a development team in a company, is an approach that can potentially result in a tool that provides positive user experience and supports the goals of practitioners working in different roles. In the following section, guidelines are provided to support such collaborative development process.

4.3.2 Guidelines for Collaborative Development of Visual Data Analytics Tools

In relation to **RQ3**, it is argued that one approach to support the utilization of usage data logging for product development purposes is the collaborative development of visual data analytics tools with end-users, such as development team members in a product development company. Next, nine guidelines are summarized from **Study V** (**P6**) to support the development and evaluation of visual data analytics tools for logged usage data. **P6** includes more elaborate descriptions of the following guidelines.

Guidelines for developing and evaluating visual data analytics tools for logged usage data:

1. **Gather an Interdisciplinary Team to Support the Development Process.** Employees in different roles will likely have different requirements and aims for usage data logging. For example, developers and customer support personnel can be interested in details related to specific error situations, while manager-level personnel could be interested in general usage data statistics. Including stakeholders from marketing, sales and user training, as well as customers' representatives, could provide even more insights into the possibilities and challenges of usage data logging. Participating in the development of the

data analytics tool can benefit the collaborating company also by improving the basic data literacy skills of the employees (Medler et al. 2011). (**P6**)

2. **Ensure Early Access to Real Logged Usage Data.** Customer data are usually confidential, and especially if data are gathered from different countries, requires familiarization with the local rules regarding privacy, security and intellectual property issues related to data logging and usage (Manyika et al. 2011). When access to real usage data is negotiated in advance, the developed tool can be tested more efficiently and any disputes over the data access are minimized in the future. In our case (**P6**), we learned that resources should be allocated to building trust and showing the value that the customer can get from sharing the logged usage data with the supplier. This could include reports on how customer's different teams use the system over time and suggestions for additional user training. One option is to use synthetic data that at least allows the inspection of the functionality of the analytics tool and concrete discussions with stakeholders while tool developers are waiting for access to real data (Crisan et al. 2016). However, lack of interesting data can affect stakeholders' motivation to explore the tool on their own and participate in the evaluation activities. (**P6**)
3. **Identify Other Data Types That Can Support Usage Data Analytics.** During the requirements gathering process, analytics tool developers should identify what other contextual data could support users in analyzing logged usage data and consider whether this data can be visualized with the same tool. For example, developers may need to view human-machine interaction events and events generated by different digital system services on the same timeline to support the sourcing of error events. (**P6**)
4. **Allocate Resources to Explore the Log Data Structure Prior to Data Wrangling.** This is especially important if different teams are responsible for the logging services and analytics tool development, as in our case (**P6**). Understanding the structure and meaning of the log data and mapping it to meaningful labels and functionality in the visualization can take considerable effort and require close collaboration with the developer(s) familiar with the logging procedure. We also recommend mapping out potential 'edge cases' (e.g., log file types or log entries that differ in formatting from others) to avoid unnecessary troubleshooting. Moreover, subsequent changes to logging services should be made in a way that does not change the log format, to avoid additional work on data wrangling. If changes are unavoidable, care should be taken to work with the analytics tool developers to limit the scope of required changes. (**P6**)
5. **Establish Coverage of Logging and Compatibility with the Visualization Tool.** In addition to understanding what the log files contain (guideline 4), it is important to establish which log files are required to fully address the design requirements. In our case (**P6**), log data was utilized from one part of the flexible manufacturing system, which was not enough to implement all the planned features for the visualization tool. For example, it was discussed that teleservice log files should be incorporated into the data visualization tool, but these files were not made available during the study. Furthermore, older systems

may log data differently, meaning that not all desired log data may be available from all systems and in the same format. Therefore, support for all different kinds of log events can be difficult to implement and it should be decided what kind of logging the developed visualization tool should primarily support. (**P6**)

6. **Combine Expert Evaluation and Field Study Methods to Include Different Viewpoints.** We learned that heuristic evaluations (Forsell & Johansson, 2010) by external HCI experts supplement user observations and interviews in the early stages of the iterative development process by detecting additional usability problems related to user interface and data visualizations. While it can be challenging to find HCI experts who are also familiar with the specific domain, such as manufacturing automation, hiring students with HCI or visualization background can be a viable option (Tory & Möller, 2005). Stakeholders from the company could also act as evaluators, but it may be challenging to motivate them to invest time in learning the evaluation process and conducting the evaluations. (**P6**)
7. **Collect Log Data of the Analytics Tool to Follow Its Usage.** It can be convenient if tool developers can use the tool itself to analyze log data collected from its usage. The log data can reveal how actively the participants use the analytics tool, without a need to disturb company employees with questions regarding the tool usage. This information can be used to motivate the participants and plan interventions if needed. Logged usage data from the tool can also provide additional information on how different features are used over time, especially outside observation sessions. Finally, log data can provide information about how the tool is used after the collaborative development period, revealing its applicability over time. (**P6**)
8. **Provide Support for Users with Varying Analytics Skills.** Interactive data analytics tools should support users who are less familiar with programming and analytics (Heer & Kandel, 2012). Help texts and instructions for novice users can be especially useful for users who did not actively participate in the collaborative development process of the visual data analytics tool. Presenting the generally most interesting data first in the UI is recommended. In our case, this meant the frequencies of used features and error events, which were the first tabs in the main data browsing and analysis view of the UX-sensors tool. (**P6**)
9. **Support the Sharing of Insights.** The key principles of developing creativity support tools include support for collaboration and open interchange (Shneiderman et al. 2006). While stakeholders have their own channels for communication, visualization tool developers can support the sharing of insights during group discussions and by implementing features into the tool that support information sharing. For example, we allowed users to add notes to the usage data timeline, with the aim that others inspecting the same data could view these comments (**P6**). Easy ways to exporting data tables or visualization images from the data analytics tool can also support users in sharing their findings with others. (**P6**)

These guidelines are derived from a case study where the visual data analytics tool was developed by an external team of researchers from academia, and are aimed at other designers and developers of visual data analytics tools. However, these guidelines could also be used by internal teams working in companies towards developing analytics tools for supporting the utilization of logged usage data. Furthermore, they can be useful background data to inform discussions with stakeholders, for example when contemplating the questions presented in the previous section concerning the utilization of usage data logging.

Finally, it is claimed that providing guidance based on real-life design cases done in collaboration with product development teams in the industry is an important way to support other tool developers in creating visual data analytics tools. Furthermore, such collaborative process can **help development teams in utilizing usage data logging to support their own work, including activities related to product development**.

5 Discussion

This chapter first summarizes the theoretical and practical contributions for each research question of the thesis in contrast to earlier scientific research. After this, the implications of this research are anchored on a process model for UX and the product development life cycle. Next, the reliability and validity of the research are assessed. Lastly, the limitations and suggestions for future research are presented.

5.1 Revisiting the Research Questions and Contributions

This section presents the research questions, summarizes the contributions of the research for each question in theory and practice, and discusses the contributions in light of previous research.

5.1.1 RQ1. How can user experience goals be defined and communicated among stakeholders in product development?

Based on results in publications **P1** and **P2**, both practical and theoretical contributions are proposed to answer **RQ1: How can user experience goals be defined and communicated among stakeholders in product development?** The practical contributions include 1) a summary of UX goals to provide examples from design cases in different contexts, 2) a suggested notation for verbally describing UX goals, 3) a description of characteristics for a good UX goal, and 4) instructions to support designers when defining and evaluating UX goals. The main theoretical contribution is the Experience Goal Elicitation Process model, focusing on how UX goals can be defined and disseminated during product design. These findings are in line with previous work by Väätäjä et al. (2012; 2015) and Kaasinen et al. (2015) in that empathic understanding of the users' world (e.g., user studies) seems to be the most potent source for inspiration when defining UX goals. However, it seems that although understanding users should be essential for designers to become familiar with the design context, users may not always be active collaborators when UX goals are defined. There can be several possible reasons for this, for example, 1) user participation in UX goal definition is not always feasible; 2) designers may already possess substantial experience from similar experience design cases; or 3) the experience design may seek possibilities for experiences rather than solve existing problems or evident needs that users may have (**P1**).

Empirical research from real-life design cases in different contexts is required to understand how UX goals can be utilized in product development. However, few empirical studies have been published regarding the utilization of UX goals in real-life experience design cases (Väätäjä et al. 2015; Roto et al. 2017). This research contributes to the body of knowledge on this topic. The presented results regarding the aspects of good UX goal reflect the findings from Väätäjä et al. (2012; 2015). UX goals should be clearly expressed and precise enough to guide the design work but at the same time broad enough to leave room for inspiration. Recently, reported experiences from four experience design cases within industry emphasize challenges in utilizing UX goals in practice (Roto et al. 2017). In particular, in the definition of UX goals, it is difficult to find the appropriate level of abstraction, as high-level goals may remain too vague and very specific goals can hinder ideation (Roto et al. 2017). In part, the guidelines proposed in **P1** should support the definition process of UX goals. In the findings, describing the feeling or emotion related to the UX goal as a part of the description of the UX goal was emphasized. In earlier research, Lu & Roto (2014) have also suggested that the momentary emotion or emotional relationship with the product should be stated in the UX goals. Another proposed approach is to include descriptions of design implications that can support the achievement of each UX goal. These design implications, added perhaps later in the design process, could further concretize the UX goals for the development team and stakeholders. UX designers can utilize different methods common in HCI design practice to carry the meaning of the intended experience, such as visual representations in the form of mood boards, scenarios, or sketches. While Table 6 lists various methods that can be used to communicate the UX goals, details of the stakeholder communication, such as how well each communication method worked, when to use which method, how and with whom, were not in the scope of the study (**P1**), but pose interesting topics for further research. In practice, it seems that traditional brainstorming sessions, workshops and meetings are the most preferred ways for communicating UX goals (**P1**).

Emerging from experiences during a research project in the automotive industry, one proposed approach that can be used to tie UX goals to practical product development and help maintaining focus on user's experiences is utilizing UX milestones (Kremer et al. 2014). The proposed process uses UX milestones as checkpoints during a journey from an abstract UX framework towards more detailed descriptions of targeted UX, utilizing stories and storyboards to communicate the experience. The process results in prototypes and final proofs of concept that are evaluated with target users, regarding the aspects of target user experiences. The contributions of this thesis could be utilized to supplement the process by Kremer et al. (2014), for instance by providing more guidance on the process of defining and describing UX goals at the first milestones. Furthermore, while clear milestones bring clarity to the UX design process, the findings in this study (see e.g. Table 7) emphasize that iteration of the UX goals is important as the product development team becomes more knowledgeable of the use context and users and receives feedback from stakeholders. This might require taking steps back in the design process and in the case of UX milestones, re-checking previously reached milestones.

The results from this research work can provide **guidance for defining and disseminating UX goals** in product development, **supporting especially UX designers,** but also other stakeholders, such as **product developers and product managers in achieving a common understanding of the UX goals during the product development project.**

5.1.2 RQ2. How can long-term user experience evaluation support product development?

The contributions to answer **RQ2: How can long-term user experience evaluation support product development?** are discussed next in the context of two sub-questions regarding a) the product development practitioners' perceptions toward the usefulness of long-term UX evaluations (**RQ2a**) and b) the feasibility of specific UX evaluation methods and tools for long-term UX evaluations in product development (**RQ2b**).

5.1.2.1 RQ2a. What kinds of perceptions do product development practitioners have about the usefulness of long-term user experience evaluation?

In HCI literature, several researchers have discussed the motivation and benefits of conducting longitudinal studies (e.g., Karapanos et al. 2009; Jain et al. 2010; Kujala et al. 2011). However, there is a lack of previous empirical research on how practitioners in companies, e.g. managers, designers and developers, utilize results from such longitudinal research in their work and how they assess the usefulness of such information, e.g., how interesting or novel the information is, for product development purposes. The practical contribution from **P3** includes a summary of three evaluation studies where practitioners' perceptions toward the usefulness of long-term UX evaluation results were reported. In summary, long-term UX evaluation results were considered relevant for 1) comparing the results with previous knowledge, 2) understanding the change in UX over time, 3) focusing future work, 4) conceptualizing and developing future products, and 5) updating current software products.

Challenges that were identified related to the time and resources that longitudinal studies demand, especially if they include several measurement points. Busy UX experts in companies may rarely have the opportunity to conduct such studies. One possible option for companies interested in long-term evaluation studies is to collaborate with academic research partners, assuming that suitable funding instruments are available to enable such studies. In addition, there may already be sources that can provide some long-term data, such as user feedback collected by customer care, even if it was collected with less rigor than in UX evaluation studies. Overlapping findings from customer care and systematic long-term UX evaluation studies can diminish the value of the evaluation results, but on the other hand, also validate the current understanding of how users experience the product. It could be useful for UX designers to look into information held by customer care and sales personnel before planning long-term UX studies. This way, UX designers could understand better what information is lacking and what is already known based on direct feedback from customers.

These findings **contribute to the current understanding in HCI research on how product development teams in companies utilize long-term UX evaluation results**. In practice, these findings can **inform UX designers and other UX experts working in the industry when planning long-term evaluations**, e.g., when justifying their claims for investing resources for conducting evaluation studies to support product development.

5.1.2.2 RQ2b. How can user experience evaluation methods and tools support the long-term user experience evaluation in product development?

As a methodological contribution to support product development work with long-term UX evaluations of products, a summary was provided on the experiences from utilizing specific evaluation tools and methods (**P3**, **P4**, **P6**). The methods and tools included the following: user observations and interviews, surveys on paper and online, the AttrakDiff questionnaire (Hassenzahl et al. 2003; Hassenzahl, 2004), the iScale survey tool (Karapanos et al. 2012a), and the UX Curve method (Kujala et al. 2011). The reported studies are novel, especially in two aspects: 1) the AttrakDiff questionnaire and UX Curve method were utilized in the evaluation of UX with pruning shears, a non-digital, practical product, and 2) the UX Curve method and iScale survey tool were utilized as a part of the remote survey. The identified benefits and challenges related to the utilization of each method in practice are summarized in Section 4.2.2. Next, some of the key findings are discussed.

First, when using AttrakDiff for evaluating experiences with practical, non-digital products, some scale purification may be required. AttrakDiff uses multi-item scales to measure hedonic and pragmatic UX attributes of products. In **Study II**, the Cronbach's alpha test suggested removal of two word-pairs (*lame–exciting* and *easy–challenging*) from items measuring the hedonic quality stimulation. Furthermore, feedback from respondents suggested that the word-pair *Takes me distant from people–Brings me closer to people* is not relevant when evaluating pruning shears. Pilot testing surveys with end-users or domain experts could be one approach to provide UX designers feedback to help to judge if specific items in AttrakDiff should be removed from the scale already before the actual data collection. Especially in studies with multiple measurement points, an abridged version of AttrakDiff might be required to reduce participant fatigue. For instance, Hassenzahl & Monk (2010) utilized an eight-item version of the AttrakDiff2 questionnaire when evaluating multiple websites. From product development perspective, AttrakDiff is readily available, fast and easy to apply questionnaire that provides information about how users currently perceive the product. However, AttrakDiff alone does not answer the question *why* or *what* should be done to improve specific UX attributes.

Second, the findings with the UX Curve method (**P4**) in a remotely done evaluation study of pruning shears (**Study II**) showed that the majority of participants could report their experiences over time with ease when relying on written instructions. However, nearly one-third of the respondents had challenges in remembering the order of their experiences after three months, highlighting a common challenge with retrospective evaluations, which rely on remembering experiences (Kahneman et al. 1993; Schacter, 1999). Kujala et al. (2011) had similar findings in a validation study of the UX Curve, suggesting that UX Curve rather provides the approximate reconstructions of the meaningful events. While the paper version of the experience curve drawing task was easy to comprehend by participants, transferring the results to digital format was laborious. Perhaps in future, this process could be supported with digital tools. For instance, a mobile application could be developed to automatically scan the curves from camera images. However, such automation may be troublesome to implement, as the style of hand-drawn curves can vary greatly and the curve drawing area may include any number of written experience narratives or even hand-drawn images. An interesting finding was revealed with the UX Curve as related to the shape (improving/deteriorating) of the pleasantness curve (i.e. “how good the product is or feels”). The pleasantness curve seems to be related to how willing the users are to recommend the product, how they evaluate the overall goodness, and the hedonic quality identification of the product. These findings add to the previous study by Kujala et al. (2011) with mobile phone users, where user satisfaction and product recommendation were found to be related to the shape of the attractiveness curve. These results provide further evidence to strengthen the reliability of the experience curve drawing tasks as an evaluation method by suggesting what can be concluded from the trends of the curves, in addition to the qualitative data in the form of written experience narratives. For UX practitioners, UX Curve can provide interesting trend information of how the overall UX or specific UX attributes have evolved over time. Furthermore, as users need to provide justification for their improving or deteriorating experiences, these written comments are likely in a form where they can be utilized to inform UX design work. However, one should be aware that the recalled events might not be accurate, as stated above. Furthermore, when utilized in remote studies, the amount of information that participants provide can vary greatly (from 1 to 7 experience narratives per participant in **P4**) and it is not possible to ask clarifying questions if the user’s descriptions are not detailed enough to be useful in design work.

Third, in **Study I,** the iScale survey tool was utilized as a part of a remote evaluation study where users’ experiences with a diving computer were evaluated retrospectively. However, the version of iScale that was used in the evaluations had usability issues that affected some of the resulting experience timelines that respondents plotted with the tool. Also, the resulting timelines by different respondents were not directly comparable with each other. This was because the starting point in the middle of the y-axis could have different meanings depending on how a respondent rated the experience at the moment of purchase, which was asked on a separate Likert scale prior to plotting the timeline. This made it challenging to report the results by comparing experience timelines from several respondents. In a previous study by Karapanos et al. (2012b), iScale was seen as a potential self-reporting tool, in contrast to labor-intensive face-to-face interviews that require skilled interviewers (Groves et al. 2009). However, the findings from this research suggest that the identified usability issues should be considered before the iScale tool can be recommended as a self-reporting tool for remote studies. If the tool is used face-to-face with respondents, where its use can easily be instructed, iScale should be able to show its potential in supporting the reconstruction of experiences from memory and providing rich qualitative data in the form of experience narratives from users to inform product development. Apart from the usability issues, iScale has mainly the same challenges and benefits as UX Curve method. One clear benefit is that the results are readily in digital format. In **Study I**, the successfully reported experience narratives were carefully written and provided interesting findings for the practitioners regarding the changes in UX over time (**P3**, Figure 2). Finally, when comparing iScale with free-hand graphing, Karapanos et al. (2012b) showed that sketching the experience over time with iScale can increase the amount and the richness of the information recalled.

In **Study V** (**P6**), the repeated quantitative measurements of UX aspects were considered as a general “alarm” if something was going wrong in the design of the visual data analytics tool. However, in practice, the qualitative data in the form of user comments and notes from user observation provided concrete findings on what to improve in the tool’s next iteration. Quantitative measures could prove to be more beneficial in studies with significantly more participants, where collection and analysis of qualitative data would require too many resources. In such cases, qualitative data could be gathered from selected representatives from the user groups, while other participants could provide quantitative data to offer designers an overall view of the current UX of the developed product.

Finally, the possibilities of **combining evaluation methods for long-term studies** are discussed. Table 9 summarized experiences of altogether six researchers using various methods and tools for long-term UX evaluations in three case studies. Based on these experiences, it is recommended that in order to gain more holistic view of the product use, UX designers should **combine different data collection approaches for long-term UX studies**. For instance, in **Study V** (**P5** and **P6**), the Multi-dimensional In-depth Long-term Case Study (MILC) approach (Shneiderman and Plaisant, 2006) was applied. MILC combines field studies with participant observation, interviews, surveys and automated logging of user activity. However, it was found that including heuristic evaluations (Forsell & Johansson, 2010) done by external evaluators to MILC approach can reveal important usability issues with the visual data analytics tool UI (**P6**, see also Section 4.3.2., guideline 6).

Due to challenges that were identified during the remote studies with the current versions of the iScale tool (Karapanos et al. 2012b) and the UX Curve method (Kujala et al. 2011), using these tools can be recommended primarily for face-to-face situations with participants. The “experience curve drawing” tasks could be combined with user interviews where the experience curves might inspire more elaborate discussions of the past experiences. To improve the feasibility of “experience curve drawing” tools for supporting product development activities, they should be easy and efficient to use also when collecting feedback from a large number of customers around the world. In future, improving the current tools or introducing new tools such as DrawUX (Varsaluoma and Kentta, 2012) may provide some solutions to the identified challenges with experience curve drawing tasks in remote studies.

However, realizing that resources for UX practitioners in companies can often be too limited for conducting long-term observation or interview studies, other approaches can be viable options for collecting data to understand UX over time. For example, when inspecting the UX of non-powered hand tools remotely, such as in **P4**, a short survey right after the first use experiences to capture the first impressions, and a retrospective measurement (e.g. with an extensive survey or experience curve drawing task) after 2-3 months might be adequate, as the use frequencies of such hand tools can vary greatly. Alternatively, while not utilized in these case studies, a lightweight diary method could be used so that users could report their experiences whenever they use the product during the evaluation period. Such diary methods can vary from simple notes making on a paper to online tools supporting repetitive reporting, such as the DrawUX survey tool (presented in section 4.2.2.5). Online survey tools with the AttrakDiff questionnaire (Hassenzahl, 2004) combined with open-ended questions can also offer interesting comparison points over time in repeated measurement studies. In case of digital products, such as online software systems, the possibility of logging user actions with the system can offer additional information to supplement subjective UX evaluation results, as discussed in the following section.

While there are numerous available UX evaluation methods (see e.g., Vermeeren et al. 2010) and instructions for their use in HCI literature, **empirical research regarding the utilization of experience curve drawing tasks, with methods such as iScale or UX Curve, are still rare.** These explorative studies provide **new empirical evidence regarding the utilization of these methods for long-term evaluation studies in different contexts**. These findings can **inform UX researchers in academia and UX practitioners in the industry when considering which methods to use in their long-term evaluation studies.** Furthermore, as the results of the studies were reported to the collaborating companies in each case study, these studies have supported the product development processes in these companies by providing them with knowledge and experiences from UX evaluation methods and conducting long-term UX evaluations with these methods. For example, experiences from **Study I** (**P3**) and **II** (**P4)** were disseminated to the collaborating companies as a part of a Long-Term UX Evaluation Toolbox, a document combining information from literature and experiences regarding the methods utilized during the project. Finally, the **DrawUX survey tool was presented as a practical contribution that can potentially support UX evaluation work**, having novelty value in its features that support a) respondents in reporting their experiences over time in remote studies and b) evaluators in analyzing and reporting the results from the experience curve drawing tasks. Online evaluation tools supporting self-reporting, such as DrawUX, might help product development teams, especially UX designers, in conducting remote and face-to-face long-term UX evaluation studies of their products and reporting the findings to relevant stakeholders.

5.1.3 RQ3. How can the utilization of usage data logging be supported in product development?

RQ3 is "How can the utilization of usage data logging be supported in product development?" RQ3 was approached from two viewpoints: 1) inspecting the requirements and expected benefits of usage data logging in a specific product development context and 2) supporting the utilization of usage data logging through collaborative development of visual data analytics tools for logged usage data. The main contribution lies in the study's context, the development of manufacturing systems, as little previous research work is available where visual data analytics tools have been utilized for analyzing logged usage data with industrial manufacturing or related industrial systems.

Some of the identified or expected benefits of usage data logging, as presented in Section 4.3.1, could well apply outside the studied contexts that included developing mobile learning services and systems for manufacturing automation industry. Examples of such benefits were that a) logged usage data can be used to verify users' own recollections of product usage, b) data logging does not interrupt normal product usage, c) log data can reveal interesting usage patterns and possibly justify a need for further qualitative inspection, and d) log data can inform developers on how the product is used after major software updates, therefore supporting product development decisions. Indeed, Grimes et al. (2007) also emphasized that data logging does not disturb the user and therefore provides unbiased observational data when studying query logs for search engines. In another example from game development field, Hullet et al. (2012) showed how analysis of long-term logged usage data resulted in recommendations for changes in the user interface of the studied auto racing game. These earlier results from other industrial domains support the generalizability of the identified and expected benefits of usage data logging from this research work. However, the expected business opportunities from usage data logging were more specific to the domain of manufacturing automation. These opportunities included improving customer support services (e.g., fault diagnosis), new customer training offers, providing additional value from customer reports and obtaining evidence for accidents, for example, regarding the liability of the damage. Overall, these results **contribute to the current understanding of how product development practitioners, especially in the manufacturing industry, perceive the benefits of usage data logging.** Further research should explore how successfully log data can meet these expectations and provide additional value for practitioners in product development and other stakeholders, including customers and end-users.

From a product development perspective, logged usage data could be a useful channel especially for development teams that have little opportunities in observing how their products are used in the field. For instance, in the case of supplier companies in the manufacturing industry, customers can be located around the world, making site visits costly. Remotely collected log data can provide an overall view of how the product is used and how individual use patterns emerge, as long as data collection and analysis on individual level is agreed with the customer. However, as noted by Grimes et al. (2007), log data does not explain *why* the user has made specific choices. Therefore, it seems that while usage data logging might support the generation of UX insights related to practical usability aspects, such as efficiency or effectiveness, insights related to emotional aspects of product use would require other evaluation methods. As results from **Study V** suggest, data logging could be useful in identifying and justifying situations where a more qualitative approach, such as user observation or interview, is necessary to understand reasons for user actions. However, there are ethical and legal issues that need to be considered in this approach, such as the anonymity of the log data and if designers can contact the specific person experiencing problems with the system, based on his or her logged behavior. Often the case can be that log data is anonymized and there is no way to contact specific users to learn more about their experiences with the system. Long-term studies, where participants agree that their product usage is logged over time, could be one potential approach where logged usage data could be utilized together with more qualitative UX evaluation methods. Long-term participation might also decrease the possible Hawthorne effect, i.e. the change in behavior due the feeling of being observed, in comparison to short-term UX evaluations.

According to the findings from **Study IV** and **V**, when utilizing self-reporting UX evaluation methods, such as surveys, usage data logging can provide a more realistic view on how the evaluated product is used. Therefore, it seems that exploring logged usage data can be beneficial in both longitudinal (**Study V**) and retrospective (**Study IV**) UX evaluation studies, assuming that log data is available from the evaluation period. An interesting approach could be to study the memories of experiences with retrospective curve drawing tools such as iScale, UX Curve or DrawUX, and compare this data with logged usage data to identify use frequencies, patterns in use and their relation to the memories of experiences. In retrospective studies, one useful approach is to explore the visualized log data together with users, as it can support the recalling of events. Bhavnani et al. (2017) utilized log data visualizations in retrospective interviews regarding mobile phone usage and found that log data provided helpful cues for the participants to recall details from app usage. In their conclusion, Bhavnani et al. emphasize the need for ethical considerations of finding the balance between what activities are logged and what details of the log data are presented to the participants, while not making participants uncomfortable with the logging. This would be an interesting question to study also in the work context, such as with users of flexible manufacturing systems (**Study V**). What level of detail in the log data users are comfortably willing to share and discuss with designers for product development purposes? Supposedly, in work context users might perceive such data logging more negatively, feeling that their personal work performance and skills are evaluated, and that if they use the system in a "wrong way", it may affect their reputation in the workplace. However, further investigation is required in this topic.

In the results Section 4.3.1, as a practical contribution, a set of questions is provided to inspire discussion among development team members to consider the feasibility of usage data logging in their work context. The proposed topics include 1) possibilities and goals for usage data logging, 2) skills required for data analysis, 3) data access and secure handling of the data, 4) tools for data analytics and visualization, and 5) transferring the "raw" data to a suitable form for data analytics tools (i.e., requirements for data wrangling). Although not an exhaustive list of all topics related to usage data logging, **these questions can be used by development teams to inspire discussions on the key aspects related to the utilization of usage data logging**. For instance, the ethical and legal considerations regarding usage data logging should be discussed in the first (possibilities and goals) and third (data access) points, as they may greatly restrict what data can be logged and utilized e.g. for product development purposes.

Guidelines to support the development of visual data analytics tools for logged usage data are presented in Section 4.3.2. While these guidelines were derived from a case study in the domain of flexible manufacturing systems (**P6**), some of the guidelines reflect experiences reported in studies from other domains, such as game development (Medler et al. 2011). For instance, Medler et al. (2011) also emphasized that gathering an interdisciplinary team can greatly support the development of analytics tools. Furthermore, Sedlmair et al. (2012) also proposed that accessing real logged usage data is one major design study pitfall. The proposed guidelines also agree with earlier research, which has suggested that data analytics tools should a) support users who are less familiar with analytics (Heer & Kandel, 2012) and b) provide support for collaboration (Shneiderman et al. 2006). It is possible that the proposed guidelines could be applied outside the manufacturing industry to support the development of visual data analytics tools, such as in the development of different web-based services or industrial systems. Furthermore, while in **Study V** an external team of researchers from academia developed the data analytics tool, the proposed guidelines could inform also internal company teams developing analytics tools for logged usage data.

Although there can be various challenges in the collaborative development of data analytics tools between researchers from academia and practitioners from industry, as summarized in the guidelines from **P6**, we argue that this can be a viable approach to supporting product development teams in utilizing logged usage data. Participation in the development and evaluation of analytics tools can benefit the collaborating company by improving the basic data literacy skills of the employees (Medler et al. 2011). Medler et al. (2011) argue for developing visual data analytics tools in parallel with product development. This is advisable whenever possible, as it can decrease the need for data wrangling when the data logging services in the system and transferring the data to the analytics tool can be designed in parallel. Indeed, one of the challenges in **Study V** was that the data analytics tool was developed by an external research team after the data logging services had been implemented. This related to the fourth guideline: "Allocate Resources to Explore the Log Data Structure Prior to Data Wrangling" (section 4.3.2)**.** In our case, analytics tool developers faced challenges in transferring the log data to a suitable format for the analytics tool and had to work in close collaboration with the system developers who had a deeper understanding of the logging process of the manufacturing system.

To conclude, the **proposed guidance for developing visual data analytics tools for usage data logging contributes to filling the gap in explorative research done in manufacturing automation domain**. By providing support for the development of visual data analytics tools for usage data logging, product development teams should have better tools at their disposal for utilizing usage data logging to support their work. Finally, from UX design and evaluation perspective, usage data logging is considered to be most valuable approach in supporting other UX evaluation methods, that can provide more qualitative data e.g. to explain reasons for users' actions and experiences with the product.

5.2 The Role of the Research Contributions in the Product Development Life Cycle

This section discusses how the main research contributions of this thesis relate to the product development life cycle and can support product development activities. In Figure 9, the main research contributions are anchored to the steps in UX and the product development life cycle model by Roto et al. (2014) presented in Figure 5. In the adapted product life cycle model in Figure 9, one can see that the evaluation data gathered after product launch, as suggested by the findings in **P3**, can be used to update the current product (in case of, e.g., software products) or to inform the design process of the next product version. The research contributions of this thesis mainly relate to investigation (steps 1, 2, 4) and evaluation activities (steps 6 and 10) during the product development life cycle as illustrated in Figure 9.

First, this research provided examples of UX goals, a notation for verbally describing UX goals, a description of characteristics for a good UX goal, and instructions for defining and evaluating UX goals. All these can be utilized in the very first stages of UX goal definition (steps 1–4). These steps include various *investigative activities* when the development team aims at understanding the current tasks (e.g., familiarization with the context by interviewing users), and

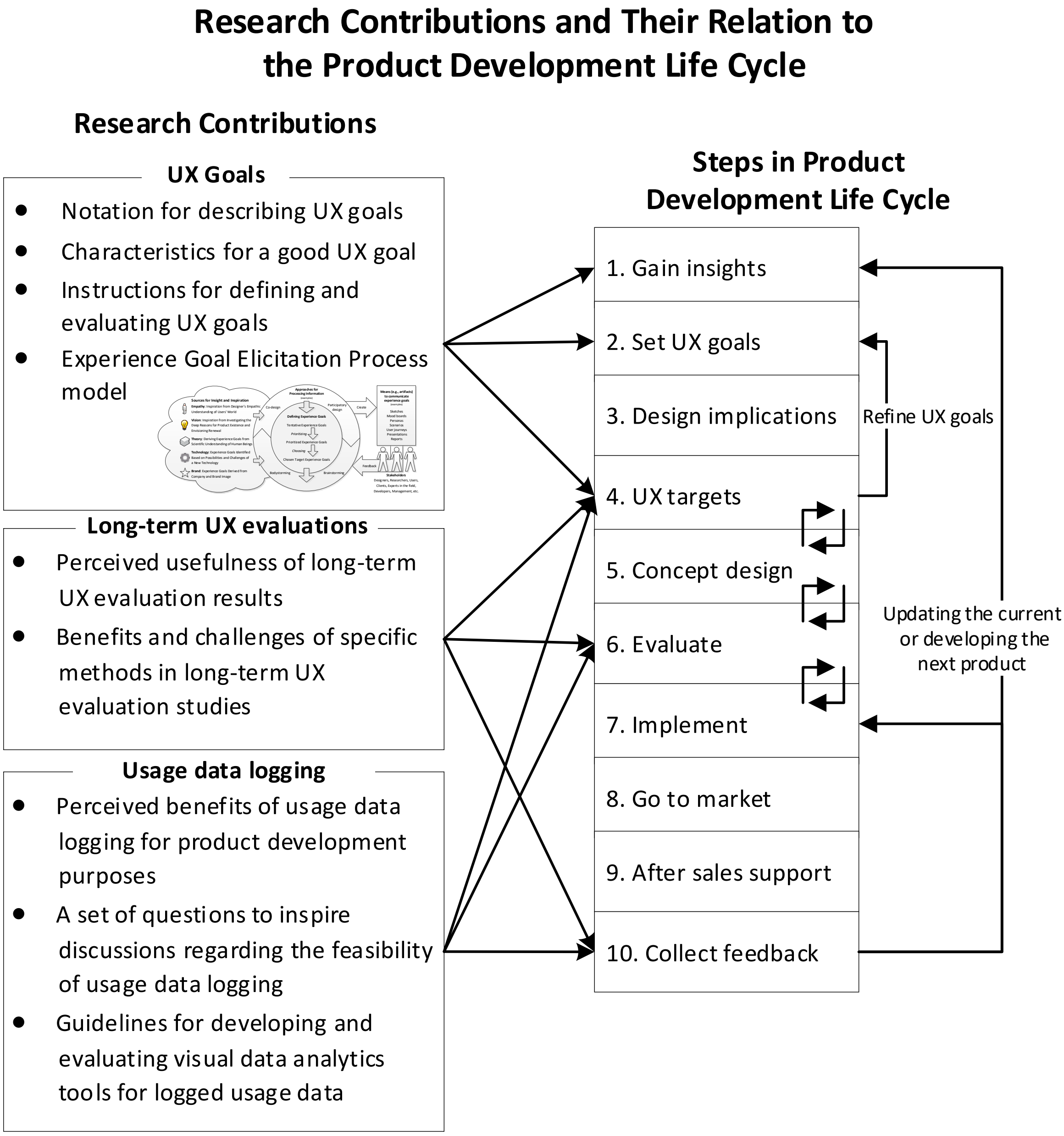

Figure 9. The relation of the research contributions to the steps in the product development life cycle. The Product Development Life Cycle on the right from Roto et al. 2014, adapted with permission (see Figure 6 for original).

defining UX goals and other requirements for the developed system (Roto et al. 2017). The Experience Goal Elicitation Process (Figure 7) illustrates the process for setting UX goals. It expands the model by Roto et al. (2014) by including examples of means (e.g., artifacts) for communicating UX goals with the development team and relevant stakeholders, whose feedback can further contribute to the UX goal definition process. UX goals should be operationalized as measurable UX targets to assess how well they have been achieved during the later phases of the development life cycle. These evaluations can happen during the iterative design and implementation phases (steps 4-7), or when collecting feedback from the product on the market (step 10).

When the measurement of UX goals, i.e., UX targets, is considered, methods for their evaluation should also be decided. *Evaluation activities* include the evaluation of the tool in relation to UX targets and evaluation of the feasibility of initial UX goals (Roto et al. 2017). In our studies, evaluation activities included evaluations made during the development process (step 6) and when collecting feedback after product launch (step 10). This research provided empirical findings regarding the usefulness of long-term UX evaluation results and the identified benefits and challenges of different long-term UX evaluation methods. These findings can inform UX designers and researchers during investigative activities when they consider which methods to use for measuring UX targets (step 4), or later on during the actual evaluation activities (steps 6 and 10). Evaluating the fulfillment of UX goals might require a high-fidelity prototype or even a finalized product that is used over an extended period. In these cases, long-term UX evaluation methods may prove to be beneficial. Finally, the developed DrawUX survey tool can be used in evaluation activities during the product development (step 6) or for evaluating products already on the market (step 10).

If feasible, the possibility of usage data logging should be considered in the early phases of product development. However, sometimes usage data logging is implemented to a product already on the market, such as in **Study V**. The set of questions for companies considering the utilization of usage data logging (**P5**, p. 433) might prove useful in these early phases. If a decision is made to utilize usage data logging, e.g., based on the expected benefits of it (see Table 10), and there is a need to start a development of new or integration process of some available visual data analytics tool, the proposed guidelines (**P6**) can provide support for this process. It is advisable to develop analytics tools at the same time when the logging capabilities of the developed system are implemented, so that the data wrangling activities (i.e., mapping and transferring the "raw" data to suitable form for an analytics tool) and visualizations to the data analytics tool can be designed and discussed at the same time. If such a tool for analyzing logged usage data is implemented, it can be utilized in evaluation activities during or after product development (steps 6 and 10) to further inform product development activities, such as continuous UI development (**P5**).

Tools and methods for long-term UX evaluations, such as those utilized in this thesis, are only some of the many methods available for evaluating how well UX targets are achieved. To some extent, usage data logging could also provide information to assess if certain UX targets, especially pragmatic ones, are reached (step 4). However, considering the subjective nature of user experience, other methods, such as surveys measuring specific aspects of UX, or qualitative approaches, such as user interviews, might be more suitable for assessing UX targets.

The conducted studies in this thesis did not relate to all steps in the product development life cycle, but included products already on the market (e.g., digital sports equipment, pruning shears, flexible manufacturing systems) or in development (visual data analytics tool). However, these findings can provide guidance especially for UX designers, product developers, and product managers working in academia or industry, when **a) investigating the feasibility of UX goals, long-term UX evaluations, or usage data logging to support their product development activities; b) conducting long-term UX evaluations of their products;** or **c) participating in development of visual data analytics tools for logged usage data**.

5.3 Assessment of the Research

In this section, the quality and limitations of this research are assessed regarding the reliability and validity of the included studies. This research was *explorative* in nature and primarily utilized *qualitative* data gathered during five *case studies*. The research questions reflect the practical approach in the studies, most of which were conducted in close collaboration with practitioners from companies, including real end-users and products on the market.

Reliability refers to the question whether the results are repeatable, i.e., if subsequent researchers could arrive at the same insights by following the same steps again (Dubois & Gibbert, 2010; Denzin & Lincoln, 1994). Limitations in terms of reliability lie in the descriptions of the conducted studies. Although the studies have been documented by the researchers who conducted them, the publications may have limitations in some of their descriptions. For example, the exact results of the evaluation studies in **Study I** were not reported due reasons of space and confidentiality issues. In general, the reliability was enhanced by carefully describing the study participants, utilized methods, and procedures for data collection and analysis in each publication. For instance, when presenting the guidelines for developing and evaluating visual data analytics tools for logged usage data, references are provided to the empirical findings during the study or from related literature to justify the proposed guidelines.

Validity addresses how consistent the conclusions from the research are. The rigor of field research is commonly assessed with construct validity, internal validity, external validity, and ecological validity (Dubois & Gibbert, 2010).

Construct validity refers to the quality of the study in investigating what it claims to investigate, i.e., how successfully the research procedure leads to accurate observations of reality (Dubois & Gibbert, 2010; Denzin & Lincoln, 1994). *Triangulation*, i.e., studying the phenomena from different views by utilizing different data collection methods and data sources (Denzin & Lincoln, 1994), can improve the construct validity in case studies (Dubois & Gibbert, 2010). The **RQ1** ("How can user experience goals be defined and communicated among stakeholders in product development?") was explored in two studies, **Study III** and **IV**. **Study III** included survey responses from researchers representing nine different design cases, i.e. data sources, while **Study IV** contributed especially to the provided examples of UX goals, based on 53 survey responses and previous literature. Alternative data collection methods, such as interviews, could have improved the construct validity in **Study III** by providing a more comprehensive view of the design cases. Regarding **RQ2a** ("What kinds of perceptions do product development practitioners have about the usefulness of long-term user experience evaluation?"); only one company was involved in **Study I**, where the usefulness of long-term UX evaluation results was assessed. However, **Study I** included three separate evaluation studies with different products and perceptions of employees working in different roles in the company, therefore improving the quality of the study. Next, findings related to the feasibility of specific methods and tools for long-term UX evaluation (**RQ2b**: "How can user experience evaluation methods and tools support the long-term user experience evaluation in product development?") are limited in a sense that they are derived from single case studies during **Studies I**, **II** and **V**. Another limitation in answering **RQ2b** is that the implications concerning the feasibility of the used UX evaluation methods are based on experiences of the researchers, and not observed when actual development teams in companies were utilizing these methods. However, the studies were conducted in close collaboration with practitioners in companies, with representative users from their customer base. In addition, more than one researcher participated in each study, therefore diminishing the possibility of a single researcher's dominating view. Finally, regarding **RQ3** ("How can the utilization of usage data logging be supported in product development?"), the results to support usage data logging in product development, e.g., guidelines, although limited to a single case study, were derived from practitioners in different roles over several data collection points (**Study V**). Furthermore, multiple evaluation methods, as proposed by the MILC approach (Shneiderman & Plaisant, 2006), were utilized in **Study V** to improve validity of the data.

Internal validity refers to causality between the collected data and the results, i.e., how well the logical reasoning can defend the research conclusions (Yin, 2003; Dubois & Gibbert, 2010). However, the studies included in this thesis were primarily explorative and did not aim to prove causality, as would be the case in, e.g., experiments (Mayo, 1996). When reporting implications from the research, they were supported with examples from the collected data, e.g., quotes from the participants' responses or descriptions of events during the studies, to clarify the reasoning. Apart from the employee surveys in **Study I**, two or more researchers always participated in data analysis, therefore decreasing threats to validity. Furthermore, the Experience Goal Elicitation Process model (**Study III**), as a main theoretical contribution of the thesis, was iteratively developed from three initial versions by collecting feedback from study participants who had conducted experience design cases in practice. However, the evaluation and further development of the resulting model are concerns for future research.

External validity is concerned with the generalizability of the findings from the study context into other settings. As case studies do not allow statistical generalization, analytical generalization is recommended instead, where generalizations are made from empirical observations to theory (Yin, 2003). In **Study III**, implications from the survey responses regarding nine different design cases were utilized in design of the Experience Goal Elicitation Process model. Using web survey as a data collection method may have posed some limitations to the richness of the data, and interviews with each participant might have provided more elaborate comments regarding the utilization of UX goals in each case study. However, all the participants had an opportunity to comment on the initial process models and the instructions for defining and evaluating UX goals, allowing the researchers to consider different viewpoints when defining the final versions.

Study I included perceptions of product development practitioners in one company developing digital sports products, posing a limitation to the generalizability of the results. However, the external validity was increased by conducting the employee surveys in three different long-term UX evaluation studies and including practitioners with different roles in the company. Although the perceived usefulness of the long-term UX evaluation results reflected views of practitioners only in the studied domain, it is argued that most of the identified aspects regarding the usefulness of UX evaluation results were general enough to, for example, motivate the long-term UX evaluations of other digital, commercial products. However, the actual utilization of long-term UX evaluation results in industry contexts is something that should be studied further, also in other contexts, as the results in this topic were limited due the small number of respondents in the follow-up survey (see Section 4.2.1).

Considering the external validity of **Study V**, contributions related to supporting the utilization of usage data logging (**RQ3**) are derived from a case study with a flexible manufacturing systems supplier company and are therefore most applicable in similar contexts. However, some of the guidelines (see Section 4.3.2) include similar findings from related research in other domains, therefore encouraging the generalization of specific guidelines into the broader setting. In practice, development teams utilizing any design guidelines should be critical in what is applicable in their current product development context.

Ecological validity refers to the relevance of the research findings in the real world. The data during this research work was collected in a real-life context, including real companies and design cases, products on the market, employees, and actual or potential customers or end-users. The case studies were conducted in a real-world setting in close collaboration with practitioners from industry, and UX evaluations included users utilizing the products in natural use contexts over time. Therefore, the ecological validity of the research is considered to be high.

5.4 Future Research

This doctoral research has inspired several avenues to continue the presented research work, related to the limitations and insights gained from the current research. Next, the possible topics for future research are discussed.

First, only a few earlier studies have explored the experience-driven design in practice and especially how UX goals are utilized over product development life cycle. The candidate shares the views with Roto et al. (2017) regarding the future research topics for UX goals utilization in experience design, including a need for more empirical research studying in which conditions UX goals "*can be translated to design implications, and how complex the relations can be*" (Roto et al. 2017). Furthermore, keeping the product development team's focus on experience over the product development life cycle is difficult (Roto et al. 2017). More empirical research studying the utilization of UX goals in different product development contexts is required to assess how UX goals can support product development activities. Furthermore, Roto et al. (2017) propose that when introducing the idea of UX goals to industry, the road should go through strategic operations, for example, by utilizing company-wide experience goals (Roto et al. 2015). Empirical research regarding such an approach is needed in the future.

The Experience Goal Elicitation Process and the instructions for defining UX goals could be iterated further in future studies. In the results, examples of stakeholders who had participated in UX goal definition and different means for communicating UX goals were reported. Future studies could aim at justifying with more rigor which stakeholders related to product development should participate in the UX goal definition, how this participation should be realized, what the potential contribution of each participant in this process is, and how the UX goals should be prioritized and chosen. According to **P1**, approaches to communicating UX goals between stakeholders include at least written and verbal communication, but also artifacts, such as personas and sketches. The survey presented in **P1** did not include the respondents' perceptions regarding how successful the used approaches were in communicating the UX goals. In future research, it would be interesting to study how well these different approaches support the communication of the intended experiences during different phases of the development process. This would also require research on the perceptions of different stakeholders participating in the product development regarding the used UX goals and approaches for their communication.

Furthermore, empirical findings from experience design projects conducted by practitioners from industry, instead of researchers from academia, could provide additional viewpoints to how UX goals can support product development activities.

Empirical research regarding retrospective UX evaluation methods and tools, such as DrawUX, is required in different contexts to inform other researchers and practitioners interested in utilizing these methods. For instance, the DrawUX tool should be evaluated in a systematic validation study, to evaluate its feasibility in supporting UX design work in product development. In addition, more studies inspecting the actual utilization of long-term UX evaluation results in industrial contexts could provide further evidence from the benefits of long-term evaluations in practice. One obvious step for future research would be to collect a database of experiences from the utilization of specific methods, for example, for long-term UX evaluation, in different contexts to guide the choices of suitable UX evaluation methods to support product development practitioners. Websites already presenting various UX evaluation methods, such as the All About UX website[1], could act as a central hub for finding data related to specific methods and experiences reported from their utilization in different product development cases.

Finally, while **Study 5** in the manufacturing automation context suggested that usage data logging has several potential benefits for practitioners working in different roles in product development, the realization of these benefits remains a topic for future research. In particular, it

[1] https://www.allaboutux.org/ (access date 2018 September 12)

would be interesting to study what kinds of insights and benefits logged usage data can provide for stakeholders in marketing, sales and user training, and study the customers' viewpoints regarding the value gained from usage data logging in the manufacturing automation context. Interesting questions relate especially to data ownership and the value proposition for customers in sharing logged usage data with the supplier company. More research on these topics could provide useful advice for other supplier companies when interested in utilizing logged usage data collected from their customers' systems. Finally, more example studies of how logged usage data is utilized to improve the UX of the product would be valuable. For instance in **Study 5**, logged usage data could help pinpointing challenges in the UI, but would still require more qualitative approach from product designers to understand why the product is used in a specific way. Such studies would make it more evident how beneficial usage data logging can be in UX evaluation. In future, one possible approach could be the combination of logged usage data with sensor data measuring users' emotional responses during product use, such as measurements of the galvanic skin response (GSR) for the intensity of emotional arousal. However, providing an easy or automated way for designers to collect such sensor data from users in similar fashion as logging usage data today, is a challenge for future studies.

6 Conclusions

During this doctoral research, three approaches related to supporting product development activities were explored: utilization of UX goals, long-term UX evaluations, and usage data logging. The findings and contributions were derived from explorative case studies with companies and their employees, real products on the market and experiences from actual or potential customers utilizing these products in real use contexts.

The research contributions of this work are both theoretical and practical, aiming at supporting product development teams, and especially UX designers and UX researchers, working in industry and academia. First, the contributions to support the utilization of UX goals include a notation for describing UX goals, a description of characteristics for a good UX goal, and instructions for defining and evaluating UX goals. A model to illustrate the Experience Goal Elicitation Process was created to structure the fuzzy front-end of experience design. Second, in the results, it was reported how product development practitioners perceived the usefulness of long-term UX evaluation results in three evaluation studies. Furthermore, the identified benefits and challenges with long-term UX evaluation methods utilized in the included studies were presented, especially regarding methods that support retrospective reporting of experiences with products. These findings aim to support product development teams in choosing suitable methods and tools for their long-term evaluation needs. Third, the results of this research work include a summary on the perceived benefits of usage data logging for product development purposes in the manufacturing automation context and a set of questions to inspire discussions regarding the feasibility of usage data logging. These findings can inform product development teams that are considering using usage data logging to support their work, also in domains outside the studied one. The collaborative development of visual data analytics tools for logged usage data is proposed as a potential approach to supporting the utilization of usage data logging in product development, and guidelines to support such collaborative development projects between academia and industry are provided.

This research work has summarized several insights from empirical case studies in close collaboration with industry. The findings have potential in supporting UX researchers and UX practitioners in their product development activities. In the future, the proposed methods and approaches can be further iterated as more case studies are conducted in different product development contexts, with a focus on UX goals, long-term UX evaluations, or utilization of logged usage data.

References

Adams, A., Lunt, P., & Cairns, P. (2008). A qualitative approach to HCI research. In Cairns, P., & Cox, A. (Eds.), Research Methods for Human-Computer Interaction (pp. 138-157). Cambridge University Press, Cambridge, UK.

Bargas-Avila, J.A., & Hornbaek, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In Proc. CHI 2011 (pp. 2689-2698). ACM Press.

Bhavnani, Y., Rodden, K., Guarnotta, L. C., Lynn, M. T., Chizari, S., & Granka, L. (2017). Understanding mobile phone activities via retrospective review of visualizations of usage data. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI ’17 (pp. 1-10). http://doi.org/10.1145/3098279.3119841

Bloch, P.H. (1995). Seeking the Ideal Product Form: Product Design and Consumer Response. Journal of Marketing, V95, July, 17.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: capturing life as it is lived. Annual review of psychology, 54(1), 579-616.

Cagan, J., & Vogel, C.M. (2001). Creating Breakthrough Products. Upper Saddle River, NJ: Prentice Hall.

Carpendale, S. (2008). Evaluating Information Visualizations. Information Visualization: Human-Centered Issues and Perspectives, 4950/2008, 19-45.

Chamberlain, A., Crabtree, A., Rodden, T., Jones, M., & Rogers, Y. (2012) Research in the Wild: Understanding ‘In the Wild’ Approaches to Design and Development. Designing Interactive Systems 2012 - Workshop.

Cooper, A., Reimann, R., & Cronin, D. (2007). About Face 3: The Essentials of Interaction Design. Indianapolis, Indiana: Wiley. p. 610. ISBN 978-0-470-08411-3.

Crisan, A., Gardy, J.L., & Munzner, T. (2016). On Regulatory and Organizational Constraints in Visualization Design and Evaluation. In Sedlmair, M., Isenberg, P., Isenberg, T., Mahyar, N., Lam, H. (Eds.) Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16) (pp. 1-9). New York, USA: ACM Press.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. Journal of Nervous and Mental Disease, 175, 526-537

Denzin, N. K., & Lincoln, Y. S. (1994). Handbook of qualitative research. Thousand Oaks: Sage.

Desmet, P. M. A., & Schifferstein H. (2011). From Floating Wheelchairs to Mobile Car Parks: a Collection of 35 Experience-Driven Design Projects. The Hague: Eleven Publishers.

Desmet, P. M. A., & Hekkert, P. (2007). Framework of product experience. International Journal of Design, 1, 57-66.

Drachen A., Seif El-Nasr M., & Canossa A. (2013) Game Analytics – The Basics. In Seif El-Nasr M., Drachen A., Canossa A. (Eds.), Game Analytics. London: Springer.

Dubois, A., & Gibbert, M. (2010). From complexity to transparency: managing the interplay between theory, method and empirical phenomena in IMM case studies. Industrial Marketing Management, 39(1), 129-136.

Eisenhardt, K. M. (1989). Building theories from case study research. Academy of Management Review, 14(4), 532-550.

Fenko, A., Schifferstein, H.N.J., & Hekkert, P. (2010). Shifts in sensory dominance between various stages of user-product interactions. Applied ergonomics 41(1), 34-40.

Ferreira, D., Gonçalves, J., Kostakos, V., Barkhuus, L., & Dey, A. K. (2014). Contextual experience sampling of mobile application micro-usage. In Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'14) (pp. 91-100). http://doi.org/10.1145/2628363.2628367

Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. In Proceedings of the 2004 conference on Designing interactive systems processes, practices, methods, and techniques - DIS '04 (pp. 261-268). http://doi.org/10.1145/1013115.1013152

Forsell, C., & Johansson, J. (2010). An heuristic set for evaluation in information visualization. In Santucci, G. (Ed.) Proceedings of the International Conference on Advanced Visual Interfaces (AVI '10) (pp. 199-206). New York, USA: ACM Press.

Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., & Landay, J. A. (2007). MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In Proceedings of the 5th international conference on Mobile systems, applications and services (pp. 57-70). New York, USA: ACM Press. http://doi.org/10.1145/1247660.1247670

Gotz, D., & Stavropoulos, H. (2014). Decisionflow: Visual analytics for high-dimensional temporal event sequence data. IEEE transactions on visualization and computer graphics, 20(12), 1783-1792.

Grimes, C., Tang, D., & Russell, D. M. (2007). Query Logs Alone are not Enough. In Proc. of WWW 07 Workshop on Query Log Analysis: http://querylogs2007.webir.org

Grossauer, C., Holzmann, C., Steiner, D., & Guetz, A. (2015). Interaction visualization and analysis in automation industry. In Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15) (pp. 407-411). New York, USA: ACM Press.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., Tour-angeau, R. (2009). Survey Methodology. John Wiley & Sons Inc.

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred system design. Behaviour and Information Technology, 22(6), 397-410.

Guzdial, M., Walton, C., Konemann, M., & Soloway, E. (1993). Characterizing Process Change Using Log File Data. Tech. Rep. GIT-GVU-93– 44.

Harbich, S., & Hassenzahl, M. (2016). User Experience in the Work Domain: A Longitudinal Field Study. Interacting with Computers, 29(3), 306-324. http://doi.org/10.1093/iwc/iww022

Hartson, R., & Pyla, P. S. (2012). The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Elsevier.

Hassenzahl, M. (2013). User experience and experience design. In Soegaard, M., & Dam, R. F. (Eds.), The encyclopedia of human–computer interaction, 2nd Ed. Aarhus, Denmark: The Interaction Design Foundation. https://www.interaction-design.org/encyclopedia/user_experience_and_experience_design.html.

Hassenzahl, M. (2010). Experience design: Technology for all the right reasons. San Rafael, California: Morgan & Claypool.

Hassenzahl, M. (2008). User experience (UX): Towards an experiential perspective on product quality. In Proceedings of the 20th International Conference of the Association Francophone d"Interaction Homme-Machine (IHM '08.) (pp. 11-15). http://doi.org/10.1145/1512714.1512717

Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. Hum-Comput. Interact. 19(4), 319-349.

Hassenzahl, M. (2003). The thing and I: Understanding the relationship between user and product. In Blythe, M., Overbeeke, C., Monk, A. F., & Wright, P. C. (Eds.), Funology: From usability to enjoyment. Human-Computer Interaction Series, vol 3 (pp. 31-42). Dordrecht: Springer.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Szwillus G., & Ziegler J. (Eds), Mensch & Computer 2003. Berichte des German Chapter of the ACM, vol 57 (pp. 187-196). Stuttgart, Leipzig: Vieweg+Teubner Verlag.

Hassenzahl, M., & Monk, A. F. (2010). The Inference of Perceived Usability from Beauty. Human-Computer Interaction 25, 235-260. Taylor & Francis. doi:10.1080/073700242010500139.

Hassenzahl, M., & Tractinsky, N. (2006). User experience – A research agenda. Behaviour and Information Technology, 25(2), 91-97. http://dx.doi.org/10.1080/ 01449290500330331.

Heer, J., & Kandel, S. (2012). Interactive analysis of big data. XRDS, 19(1), 50-54.

Hekkert, P., Mostert, M., & Stompff, G. (2003). Dancing with a machine: A case of experience-driven design. In Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces (pp. 114-119). http://doi.org/10.1145/782896.782925

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., & Verplank W. (1992). ACM SIGCHI Curricula for Human-Computer Interaction. Technical Report. New York, USA: ACM Press SIGCHI. Retrieved 15 July 2014.

Hilbert, D. M., & Redmiles, D. F. (2000). Extracting usability information from user interface events. ACM Computing Surveys, 32(4), 384-421. http://doi.org/10.1145/371578.371593

Hullett, K., Nagappan, N., Schuh, E., & Hopson, J. (2012). Empirical Analysis of User Data in Game Software Development. In Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Lund (pp. 89-98). ACM Press. doi: 10.1145/2372251.2372265

Holzmann, C., Lettner, F., Grossauer, C., Wetzlinger, W., Latzelsperger, P., & Augdopler, C. (2014). Logging and Visualization of Touch Interactions on Teach Pendants. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14) (pp. 619-624). New York, USA: ACM Press.

International Organization for Standardization. (1999). ISO 13407: 1999, Human-centered design processes for interactive systems. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardisation. (2010). ISO FDIS 9241-210: Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems. Geneva, Switzerland: International Organization for Standardization.

Isomursu, M. Tähti, M. Väinämö, S., & Kuutti, K. (2007). Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. International Journal of Human-Computer Studies, 65(4), 404-418.

Jain, J., Rosenbaum, S., & Courage, C. (2010). Best practices in longitudinal research. In CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10) (pp. 3167-3170). New York, USA: ACM Press. doi: https://doi.org/10.1145/1753846.1753946

Jordan, P.W. (1998). Human factors for pleasure in product use. Applied Ergonomics 29(1), 25-33.

Kaasinen, E., Roto, V., Hakulinen, J., Heimonen, T., Jokinen, J., P., P., Karvonen, H., Keskinen, T., Koskinen, H., Lu, Y., Saariluoma, P., Tokkonen, H., & Turunen, M. (2015). Defining User Experience Goals to Guide the Design of Industrial Systems. Behaviour & Information Technology 34(10), 976-991. Taylor & Francis. doi: 10.1080/0144929X.2015.1035335

Kahneman, D., Fredrickson, B.L., Schreiber, C.A., & Redelmeier, D.A. (1993). When More Pain Is Preferred to Less: Adding a Better End. Psychological Science 4(6), 401-405.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. Science 306(5702), 1776-1780.

Karapanos, E., Jain, J., & Hassenzahl, M. (2012a). Theories, methods and case studies of longitudinal HCI research. In CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12) (pp. 2727-2730). New York, USA: ACM Press. doi: https://doi.org/10.1145/2212776.2212706

Karapanos, E., Martens J., & Hassenzahl M. (2012b). Reconstructing Experiences with iScale. Int. J. Hum-Comput. St. 70(11), 849-865.

Karapanos, E., Martens, J., & Hassenzahl, M. (2010). On the Retrospective Assessment of Users' Experiences Over Time: Memory or Actuality? In CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10) (pp. 4075-4080). New York, USA: ACM Press. doi: https://doi.org/10.1145/1753846.1754105

Karapanos, E. Zimmerman, J. Forlizzi, J., & Martens, J. B. (2010). Measuring the dynamics of remembered experience over time. Interacting with Computers 22(5), 328-335.

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J. B. (2009). User experience over time: an initial framework. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09) (pp. 729-738). New York, USA: ACM. doi: https://doi.org/10.1145/1518701.1518814

Karvonen, H., Koskinen, H., & Haggrén, J. (2012a). Defining User Experience Goals for Future Concepts. A Case Study. In Väätäjä, H., Olsson, T., Roto, V. and Savioja, P. (Eds.), In NordiCHI2012 UX Goals 2012 Workshop Proceedings, Tampere (pp. 14-19). TUT Publication series.

Karvonen, H., Koskinen H., & Haggrén, J. (2012b). Enhancing the User Experience of the Crane Operator: Comparing Work Demands in Two Operational Settings. In Proceedings of the 30th European Conference on Cognitive Ergonomics, Edinburgh (pp. 37-44). UK: Edinburgh Napier University.

Karvonen, H., Koskinen, H., Tokkonen H., & Hakulinen, J. (2014). Evaluation of User Experience Goal Fulfillment: Case Remote Operator Station. In Shumaker R., & Lackey S. (Eds), Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality (VAMR 2014). Lecture Notes in Computer Science, vol 8526 (pp. 366-377). Springer, Cham.

Ketola, P., & Roto, V. (2009). On User Experience Measurement Needs - Case Nokia. International Journal on Technology and Human Interaction (IJTHI), 5(3), 78-89.

Kim, J. H., Gunn, D. V, Schuh, E., Phillips, B. C., Pagulayan, R. J., & Wixon, D. (2008). Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08) (pp. 443-452). New York, USA: ACM Press. doi: https://doi.org/10.1145/1357054.1357126

Kohavi, R., Henne, R. M., & Sommerfield, D. (2007). Practical Guide to Controlled Experiments on the Web. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '07) (pp. 959-967). New York, USA: ACM Press. doi: https://doi.org/10.1145/1281192.1281295

Koskinen, H., Karvonen H., & Tokkonen, H. (2013). User Experience Targets as Design Drivers: A Case Study on the Development of a Remote Crane Operation Station. In Proceedings of the 31st European Conference on Cognitive Ergonomics (ECCE '13). Article 25, 9 pages. New York, USA: ACM Press. doi: http://dx.doi.org/10.1145/2501907.2501956

Kremer, S., Michailidou, I., von Saucken, C., & Lindemann, U. (2014). User Experience Milestones Structuring the Development of Experience Products. In Marcus A. (Eds), Proceedings of the

International Conference of Design, User Experience, and Usability (DUXU 2014). Lecture Notes in Computer Science, vol 8520 (pp. 308-318). Springer, Cham.

Kujala, S. (2008). Effective user involvement in product development by improving the analysis of user needs. Behaviour & Information Technology, 27(6), 457-473.

Kujala, S., & Miron-Shatz, T. (2013). Emotions, experiences and usability in real-life mobile phone use. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13) (pp. 1061-1070). New York, USA: ACM Press. doi: https://doi.org/10.1145/2470654.2466135

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

Kuusinen, K., & Väänänen-Vainio-Mattila, K. (2012). How to make agile UX work more efficient: Management and sales perspectives. In Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12) (pp. 139-148). New York, USA: ACM Press. doi: https://doi.org/10.1145/2399016.2399037

Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. Computers in Human Behavior, 43, 35-48. doi: http://doi.org/http://dx.doi.org/10.1016/j.chb.2014.10.048

Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2012). Empirical Studies in Information Visualization: Seven Scenarios. IEEE Transactions on Visualization and Computer Graphics, Institute of Electrical and Electronics Engineers, 18(9), 1520-1536.

Law, E. L., Hassenzahl, M., Karapanos, E., Obrist, M., & Roto, V. (2015). Tracing Links between UX Frameworks and Design Practices: Dual Carriageway. In Proceedings of HCI KOREA 2015 (pp. 188-195). Hanbit Media, Inc.

Law, E. L., Roto, V., Hassenzahl, M., Vermeeren, A., & Kort, J. (2009). Understanding, Scoping and Defining User Experience: A Survey Approach. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09) (pp. 719-728). New York, USA: ACM Press. doi: https://doi.org/10.1145/1518701.1518813

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. New Directions for Methodology of Social and Behavioral Science 15, 41-56.

Libai, B., Muller, E., & Peres, R. (2013). Decomposing the value of word-of-mouth seeding programs: acceleration versus expansion. J. Mark. Res. 50, 161-176.

Liu, N., Liu, Y., & Wang, X. (2010). Data logging plus e-diary: towards an online evaluation approach of mobile service field trial. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI '10) (pp. 287-290). New York, USA: ACM Press. doi: https://doi.org/10.1145/1851600.1851650

Lopes, P., & Roy, B. (2015). Dynamic Recommendation System Using Web Usage Mining for E-commerce Users. Procedia Computer Science 45, 60-69.

Lu, Y., & Roto, V. (2014). Towards meaning change: experience goals driving design space expansion. In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14) (pp. 717-726). New York, USA: ACM Press. doi: https://doi.org/10.1145/2639189.2639241

Mahlke, S. (2005). Understanding users' experience of interaction. In Proceedings of the 2005 annual conference on European association of cognitive ergonomics (EACE '05) (pp. 251-254). University of Athens.

Mahlke, S. (2008). User Experience of interaction with technical systems. Theories, methods, empirical results, and their application to the design of interactive systems. Saarbrücken, Germany: VDM Verlag.

Mahlke, S., & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07) (pp. 915-918). New York, USA: ACM Press. doi: https://doi.org/10.1145/1240624.1240762

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Martín, H., Bernardos, A. M., Iglesias, J., & Casar, J. R. (2013). Activity logging using lightweight classification techniques in mobile devices. Personal and Ubiquitous Computing, 17(4), 675-695. http://doi.org/10.1007/s00779-012-0515-4

Mayo, D. (1996). Error and the Growth of Experimental Knowledge. University of Chicago Press.

Mcdonagh, D., & Storer, I. (2015). Mood Boards as a Design Catalyst and Resource: Researching an Under-Researched Area. The Design Journal, 7(3), 16-31. Taylor & Francis Online.

Medler, B., John, M., & Lane, J. (2011). Data cracker: developing a visual game analytic tool for analyzing online gameplay. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11) (pp. 2365-2374). New York, USA: ACM Press. doi: https://doi.org/10.1145/1978942.1979288

Mendoza, V., & Novick, D. (2005). Usability over time. In Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information (SIGDOC '05) (pp. 151-158). New York, USA: ACM Press. doi: http://dx.doi.org/10.1145/1085313.1085348

Mitchell, T.R., & James, L.R. (2001). Building better theory: time and the specification of when things happen. Acad. Manag. Rev. 26, 530-547.

Munzner, T. (2009). A Nested Model for Visualization Design and Validation. IEEE Transactions on Visualization and Computer Graphics, 15(6), 921-928.

Norman, D. (2004). Emotional Design – Why We Love (or Hate) Everyday Things. New York: Basic Books.

Norman, D. A. (2009). THE WAY I SEE IT - Memory is more important than actuality. Interactions 16(2), 24-26.

Novick, D., & Santaella, B. (2012). Short-term methodology for long-term usability. In Proceedings of the 30th ACM international conference on Design of communication (SIGDOC '12) (pp. 205-212). New York, USA: ACM Press. doi: http://dx.doi.org/10.1145/2379057.2379097

Olson, G. M., Herbsleb, J. D., & Rueter H. H. (1994). Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. Hum.-Comput. Interact. 9(4), 427-472. doi: http://dx.doi.org/10.1207/s15327051hci0903&4_6

Ouden, E.d., Yuan, L., Sonnemans, P.J.M., & Brombacher, A.C. (2006). Quality and Reliability Problems from a Consumer's Perspective: an Increasing Problem Overlooked by Businesses? Quality and Reliability Engineering International 22(7), 821-838.

Patton, M.Q. (2001). Qualitative Research and Evaluation Methods (3rd ed.) London: Sage Publications.

Perer, A., & Shneiderman, B. (2008). Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08) (pp. 265-274). New York, USA: ACM Press.

Ployhart, R. E., & Ward, A.-K. (2011). The “Quick Start Guide” for Conducting and Publishing Longitudinal Research. J. Bus. Psychol. 26(4), 413-422. doi:10.1007/s10869-011-9209-6

Rajeshkumar, S., Omar, R., & Mahmud, M. (2013). Taxonomies of User Experience (UX) evaluation methods. In 3rd International Conference on Research and Innovation in Information Systems (ICRIIS’13), Kuala Lumpur (pp. 533-538). doi: 10.1109/ICRIIS.2013.6716765

Reichheld, F.F. (2003). The one number you need to grow. Harvard Business Review 81(12), 46-54.

Rogers, Y., Sharp, H., & Preece, J. (2011). Interaction design – beyond human-computer interaction. (3rd ed.) (p. 585). Chichester: Wiley.

Rohn, J. A. (2007). How to organizationally embed UX in your company. Interactions 14(3), 25-28.

Roto, V., Kaasinen, E., Heimonen, T., Karvonen, H., Jokinen, P. P. J., Mannonen, P., Nousu, H., Hakulinen, J., Lu, Y., Saariluoma, O. P., Kymäläinen, T., Keskinen, T., Turunen, M., & Koskinen, H. (2017). Utilizing Experience Goals in Design of Industrial Systems. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 6993-7004). New York, USA: ACM Press.

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (Eds.). (2011). The UX White Paper. Outcome of the Dagstuhl Seminar on Demarcating User Experience, Germany. http://www.allaboutux.org/uxwhitepaper.

Roto, V., Lu, Y., Nieminen, H., & Tutal, E. (2015). Designing for User and Brand Experience via Company-wide Experience Goals. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (pp. 2277-2282). New York, USA: ACM Press.

Roto, V., Nuutinen, M., & Smedlund, A. (Eds.) (2014). UX HOW? A Glimpse of User Experience for B2B Industry – Issue 2. User Experience and Usability in Complex Systems –program. Finnish Metals and Engineering Competence Cluster (FIMECC). Retrieved 11 November 2017. https://issuu.com/vttfinland/docs/ux_booklet_english?e=5313536/10315158

Sanders, E. B.-N., & Dandavate, U. (1999). Design for Experiencing: New Tools. In Overbeeke, C.J., & Hekkert, P. (Eds.) Proc. of the First International Conference on Design & Emotion, Holland (pp. 87-92).

Saraiya, P., North, C., Lam, V., & Duca, K. (2006). An Insight-Based Longitudinal Study of Visual Analytics. IEEE Trans. Visualization and Computer Graphics, 12(6), 1511-1522.

Schacter, D.L. (1999). The seven sins of memory. Insights from psychology and cognitive neuroscience. Am. Psychol. 54, 182-203.

Sedlmair, M., Isemberg, P., Baur, D., & Butz, A. (2011). Information Visualization Evaluation in Large Companies : Challenges, Experiences and Recommendations. Information Visualization Journal, 10(3), 248-266. http://doi.org/10.1177/1473871611413099

Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design study methodology: Reflections from the trenches and the stacks. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2431-2440. http://doi.org/10.1109/TVCG.2012.213

Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What Is Satisfying About Satisfying Events ? Testing 10 Candidate Psychological Needs. Journal of Personality and Social Psychology 80(2), 325-339. DOI=10.1037//O022-3514.80.2.325

Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B. and 13 others. (2006). Creativity Support Tools: Report from a U.S. National Science Foundation Sponsored Workshop, International Journal of Human–Computer Interaction 20(2), 61-77.

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization (BELIV '06) (pp. 1-7). New York, USA: ACM Press.

Sim, G., Nouwen, M., Vissers, J., Horton, M., & Slegers, K. (2016). Using the MemoLine to capture changes in user experience over time with children. International Journal of Child-Computer Interaction 8, 1-14. http://doi.org/10.1016/j.ijcci.2016.07.001

Stake, R. E. (1995). The art of case study research. Thousand Oaks, CA: Sage.

Stolper, C. D., Perer, A., & Gotz, D. (2014). Progressive visual analytics: User-driven visual exploration of in-progress analytics. IEEE transactions on visualization and computer graphics 20(12), 1653-1662.

Swallow, D., Blythe, M., & Wright, P. (2005). Grounding experience: relating theory and method to evaluate the user experience of smartphones. In Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics (pp. 91-98).

Taylor, S. J., Bogdan, R., & DeVault, M. (2016). Introduction to Qualitative Research Methods: A Guidebook and Resource (4th ed.). Wiley.

TENK. (2012). Responsible conduct of research and procedures for handling allegations of misconduct in Finland. Guidelines of the Finnish Advisory Board on Research Integrity 2012. http://www.tenk.fi/en/tenk-guidelines

Tory, M., & Möller, T. (2005). Evaluating visualizations: Do expert reviews work? IEEE Computer Graphics and Applications, 25(5), 8-11. http://doi.org/10.1109/MCG.2005.102

Yin, R. K. (2003). Case study research: Design and methods (3rd ed.). Thousand Oaks, CA: Sage.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07) (pp. 493-502). ACM, New York, NY, USA. DOI: https://doi.org/10.1145/1240624.1240704

Varsaluoma, J., & Kentta, V. (2012). DrawUX: Web-Based Research Tool for Long-Term User Experience Evaluation. In Proceedings of the NordiCHI 2012, the 7th Nordic Conference on Human-Computer Interaction, Copenhagen, Denmark (pp. 769-770). New York, USA: ACM Press.

Vaughan, M., & Courage, C. (2007). SIG: capturing longitudinal usability: what really affects user performance over time? In CHI '07 Extended Abstracts on Human Factors in Computing Systems (CHI EA '07) (pp. 2149-2152). New York, USA: ACM Press. doi: https://doi.org/10.1145/1240866.1240970

Vaughan, M., Courage, C., & Rosenbaum, S. (2008). Longitudinal usability data collection: art versus science? In CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08) (pp. 2261-2264). New York, USA: ACM Press. doi: https://doi.org/10.1145/1358628.1358664

Vermeeren, A.P.O.S., Law, E.L.C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10) (pp. 521-530). New York, USA: ACM Press. doi: http://doi.acm.org/10.1145/1868914.1868973

Väätäjä, H., Heimonen, T., Tiitinen, K., Hakulinen, J., & Turunen, M. (2015). Suppliers' expectations on usage data analytics of complex industrial systems. In Proceedings of ISPIM Conferences, (December) (pp. 1-9). International Society for Professional Innovation Management ISPIM.

Väätäjä, H., Olsson, T., Roto, V., & Savioja, P. (Eds.). (2012). How to Utilize User Experience Goals in Design? In UX Goals 2012 Workshop Proceedings. Tampere: TUT Publication series.

Väätäjä, H., Savioja, P., Roto, V., Olsson, T., & Varsaluoma, J. (2015). User experience goals as a guiding light in design and development – Early findings. In INTERACT 2015 Adjunct proceedings (pp. 521-527). University of Bamberg Press.

Wahlström, M., Karvonen, H., Kaasinen, E., & Mannonen, P. (2014). Designing for Future Professional Activity – Examples from Ship Bridge Concept Design. In Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics (pp. 3965-3975), Krakow, Poland.

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

Wongsuphasawat, K., Guerra Gómez , J.A., Plaisant, C., Wang, T.D., Taieb-Maimon, M., & Shneiderman, B. (2011). LifeFlow: visualizing an overview of event sequences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11) (pp. 1747-1756). New York, USA: ACM Press.

Wright, P., & McCarthy, J. (2010). Experience-Centered Design: Designers, Users, and Communities in Dialogue. Synthesis Lectures on Human-Centered Informatics 3(1), 1-123. doi: https://doi.org/10.2200/S00229ED1V01Y201003HCI009

Wright, P., & McCarthy, J. (2004). Technology as Experience. Cambridge, MA, USA: MIT Press.

Wright, P., McCarthy, J., & Meekison, L. (2003). Making sense of experience. In Blythe, M.A., Overbeeke, K., Monk, A. F., & Wright, P. C. (Eds.) Funology: From Usability to Enjoyment. The Netherlands: Kluwer Academic Publishers.

ORIGINAL PAPERS

P1

THE FUZZY FRONT END OF EXPERIENCE DESIGN: ELICITING AND COMMUNICATING EXPERIENCE GOALS

by

Varsaluoma, J., Väätäjä, H. Kaasinen, E., Karvonen, H. and Lu, Y, 2015.

In Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OzCHI '15). ACM New York. pp. 324-332.

The Fuzzy Front End of Experience Design: Eliciting and Communicating Experience Goals

Jari Varsaluoma, Heli Väätäjä
Tampere Univ. of Technology
Finland
firstname.lastname@tut.fi

Eija Kaasinen, Hannu Karvonen
VTT Technical Research Centre of Finland Ltd
firstname.lastname@vtt.fi

Yichen Lu
Aalto University
Finland
yichen.lu@aalto.fi

ABSTRACT
When starting an experience design process, designers should first determine the experience to aim for. In the fuzzy front end of the experience design process, there are often several alternative sources for gaining insight and inspiration in defining this experience. In this paper, we describe our findings from two surveys about experience goal setting and approaches to communicate about these goals with stakeholders. The results from researchers working on 9 different experience design cases suggest that "empathic understanding of the users' world" is the most used source of insight and inspiration in defining experience goals. As an end result, we propose the model for Experience Goal Elicitation Process to clarify the fuzzy front end of experience design and instructions to support designers in defining and evaluating experience goals.

Author Keywords
Experience goal, UX goal, experience design, fuzzy front end, survey study, Experience Goal Elicitation Process

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
In experience design, the emotional and experiential elements are the main starting point of design activities. The intended user experience (UX) is taken as the primary objective of the design process (Hekkert et al., 2003). ISO 9241-210 (2010) defines UX as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service". Hassenzahl (2003) divides UX into pragmatic (e.g., usability and utility) and hedonic (e.g., stimulation and identification) aspects of product use. Similarly, Mahlke (2007) presents a model of UX components with instrumental (e.g., ease of use) and non-instrumental system qualities (e.g., visual attractiveness). In addition, he also links the perceptions of these qualities to emotional reactions which are presented as the third UX component, during user's interaction with the system (Mahlke, 2007)

OzCHI '15 , December 07 - 10 2015, Melbourne, VIC, Australia

DOI: http://dx.doi.org/10.1145/2838739.2838761

Because experiences with interactive products and services are subjective, dynamic and context-dependent (Hassenzahl & Tractinsky, 2006), it is suggested that designers can only aim to facilitate specific experiences among the users, i.e., design *for* an experience (Wright et al., 2003; Sanders & Dandavate, 1999). According to Desmet and Schifferstein (2011), two important challenges in experience design are: 1) to determine what experience to aim for and 2) to design something that is expected to evoke that experience. In this paper, we focus on the first challenge.

The "fuzzy front end" refers to actions at the beginning of the development process, when the targeted product or service is not yet decided and making changes to the target result is still inexpensive (Khurana & Rosenthal, 1998). The first step in designing for an experience is to define the *experience goals* that concretize what the users are intended to experience before, during or after interacting with the product or service. Clearly defined experience goals, to which the project team commits to already during the fuzzy front end phase, can help the team by "keeping user experience in focus through the multidisciplinary product development and marketing process" (Kaasinen et al., 2015).

In this paper, we present results from web survey studies with nine (9) respondents working on nine different experience design cases in the field of human-computer interaction (HCI). The surveys were conducted in connection with an academic conference workshop in NordiCHI2014. The workshop focused on the first phases of experience design in real life design cases where the design was driven by the intended experiences, described as "experience goals". The research questions directing our research were:

- Where can insight and inspiration be gained from to define experience goals?
- What are the characteristics of a good experience goal?
- How should experience goals be communicated to stakeholders?

As a result of our findings and prior research, we propose a model illustrating the elicitation process of experience goals and different approaches for communicating them among stakeholders. The model aims to clarify the fuzzy front end of experience design for HCI academics and practitioners. We also summarise our learnings in a more practical set of instructions to support designers when defining and evaluating experience goals. Our findings contribute to the body of knowledge of experience design research in the field of HCI.

EXPERIENCE/UX GOALS

The first academic workshop to collect cases of UX goal (in this paper, we use term "experience goal") utilisation was held in a NordiCHI2012 conference (Väätäjä et al., 2012; Väätäjä et al., 2015). The domains of the case studies varied from workplace to consumer applications and education. The workshop participants defined a good experience goal as something that 1) helps aiming the design as a guiding light, 2) is measurable, 3) describes positive emotions, and 4) is a way to communicate the desired experience with other people. Furthermore, experience goals were considered useful in keeping the focus on important issues and providing inspiration.

In their recent study, Kaasinen et al. (2015) identified five approaches to gain insight and inspiration for experience goal-setting in industrial environments. The approaches were derived from four industrial design case studies with companies and were supplemented with literature study. Table 1 presents definitions for the five identified approaches (Kaasinen et al., 2015).

Brand	UX Goals Derived from Company and Brand Image
Theory	Deriving UX Goals from Scientific Understanding of Human Beings
Empathy	Inspiration from Designer's Empathic Understanding of Users' World
Technology	UX Goals Identified Based on Possibilities and Challenges of a New Technology
Vision	Inspiration from Investigating the Deep Reasons for Product Existence and Envisioning Renewal

Table 1. Five approaches to gain insight and inspiration for UX-goal setting. (Kaasinen et al., 2015)

Karvonen et al. (2012) suggest that the made design solutions should be traceable back to the originally defined experience goals during later design phases. In this way, it is possible to measure and evaluate the fulfilment of the experience goals in different phases of the design work, such as when evaluating the designed product with users (Karvonen et al., 2014). In their experience design case of a remote operator station (ROS) for container gantry crane operation in port yards, a combination of methods was used to evaluate the fulfilment of the original experience goals. User interviews, testing sessions with a simulation version of the ROS, and UX questionnaires (for measuring UX and usability of the system) provided evidence to evaluate if the original claims for target experience goals had been achieved. They concluded that a modified version of the Usability Case method (see Liinasuo & Norros, 2007) could be a suitable approach for evaluating the fulfilment of experience goals in the future.

STUDY DESIGN

The goal of the study was to understand how experience goals are created and communicated among stakeholders. The study consisted of four phases: 1) a survey for experience design cases, 2) the analysis of the results and creation of model prototypes and instructions, 3) a follow-up survey, and 4) the iteration of the most promising model and the instructions.

Participants

The respondents were researchers working on experience design related cases, and had submitted their case papers to a workshop about experience design in the field of HCI. Prior to the workshop, we received 11 responses from 16 possible authors and co-authors (i.e., a 69%

response rate). Three identical responses were received from the same experience design case, so two of them were removed from the data set. Therefore, in total, nine responses were used in further analysis. All nine respondents came from academic institutions, either universities or research organizations, situated in Sweden, Finland, Germany or United Kingdom.

Ten responses were received to the follow-up survey. Despite two reminders, only five of the original nine respondents answered the follow-up survey. Table 2 summarises how these five described their expertise in the HCI and experience design fields to be at the time of the first survey. Three of them were researchers responsible for defining, designing and/or evaluating the solution, one was a researcher conducting a user study and one acted as project manager and design lead. All five had more than four years of experience from experience design related activities and had worked in several experience design projects (see Table 2). Four out of five respondents had more working experience from academia than from industry. In addition to the five original respondents, three HCI researchers from the Tampere University of Technology and two of the workshop organisers gave feedback about the proposed models and the instructions.

Based on the follow-up survey responses, case C4 (see Table 4 for the case descriptions) was an academic research project, while cases C3, C5, C6 and C8 were collaborative projects between academic and industrial partners.

Web Surveys

Table 3 presents the questions used in the first web survey. The participants were first asked to choose a specific case of experience design and answer the questions 2 to 5 (see Table 3) based on their experiences during the case. In the second question about the sources of insight and inspiration for the used experience goals, we utilized the five approaches presented by Kaasinen et al. (2015), but also left an open-ended option for alternative sources. The last question was related to the definition of experience goals in general. Here, we used a projective technique called sentence completion (e.g., Soley & Smith, 2008).

The follow-up web survey was conducted to iterate the

Based on the case described in your paper, or thinking about some other case, please answer the following questions.

1. Describe the topic of the case briefly

2. Where did you get insight and inspiration to define what experience to aim for? (Choose all that apply)

   a) Company or brand image (Brand)

   b) Scientific understanding of human beings (Theory)

   c) Empathic understanding of the users' world (Empathy)

   d) Possibilities and challenges of a new technology (Technology)

   e) Reasons for product existence and envisioning renewal (Vision)

   f) Something else, what?

3. Who participated in defining the targeted experiences (experience goals)?

4. What were the targeted experiences (experience goals) in your case?

5. How (in what form) did you communicate the targeted experiences?

6. Based on you own experience, please complete the following sentence: In my opinion a good experience goal is…

Table 3. Experience design case survey questions.

Topic	n	Range	Mean (SD)
HCI related working experience from academia? (years/months)	4 (1 missing)	3 y 3 m – 14 y	12 y 9 m (3 y 11 m)
HCI related working experience from industry? (years/months)	4 (1 missing)	1 y – 6 y	3 y 5 m (2 y 2 m)
Working or research experience from Experience Design (e.g., experience goals, design for experience, experience evaluation) related activities? (years/months)	5	4 y 2 m – 15 y	7 y 8 m (4 y)
How would you evaluate your own expertise and knowledge from theories and research related to experience design field? (Very low 1 – 5 Very high)	5	3 – 5	4 (0.63)
How would you evaluate your own expertise and knowledge from conducting experience design work in practice? (Very low 1 – 5 Very high)	5	3 – 5	4 (0.63)
How many other experience design projects had you worked on before the one reported here? (i.e. projects following experience design process, with defined experience goals)	5	6 – 11 projects	9.4 projects (1.85)

Table 2. The follow-up survey results regarding the respondents' expertise at the moment of the first survey [n=5].

proposed models and instructions. This survey also included more detailed background questions regarding 1) the respondent's role in the experience design case, and 2) working experience in HCI field and experience design related activities at the time of the first survey. In addition, we asked the respondents to choose their favourite of the three Experience Goal Elicitation Process model prototypes, justify their choice and suggest improvements to the model. Finally, we asked how understandable our instructions for "defining and evaluating experience goals" were, did they agree or disagree with the instructions, and how they could be improved?

Process

A link to the first web survey was sent by e-mail to the participants of the workshop and to all the authors of each accepted position paper of the workshop. Two researchers analysed the results of each open question and categorised similar responses to their own groups. Based on the survey results and previous literature, the researchers outlined several prototype models for an Experience Goal Elicitation Process and a set of instructions for defining and evaluating experience goals. The models and the instructions were iterated by a group of HCI researchers at the Tampere University of Technology and separately by each author of this paper, resulting in three alternative versions of the model and an updated set of instructions.

Approximately 8 months after the first survey, a link to a follow-up survey was sent to all workshop participants and organizers. An abridged version of the survey was also sent to HCI researchers working at the Tampere University of Technology. Based on the feedback received from the workshop participants, the workshop organizers, and the fellow researchers at the university, two researchers further iterated the most promising version of the model and updated the instructions.

RESULTS

This section describes the main results and discusses the findings. At the end of this section, we present the model of an Experience Goal Elicitation Process and instructions to support designers in defining and evaluating experience goals.

Sources for Insight and Inspiration When Defining Experience Goals

Table 4 presents short descriptions of the cases and what *sources for insight and inspiration* (see Kaasinen et al., 2015) the participants reported in each case *when defining experience goals*.

In the reported case studies, "empathy" (7/9 responses) and "visioning" (5/9) were the most often used sources. "Brand" and "vision" were mentioned in all three industrial cases, which may indicate that these cases had clear business purposes, which were different from more research-driven cases. In seven out of nine cases, three or more different sources were used for insight and inspiration when defining experience goals. This suggests that the respondents prefer to combine multiple sources of information when defining experience goals.

The reported five "Other" sources for insight and inspiration were 1) the environment (case C4), 2) co-design cued by site visits (C5), 3) the values of an author to whom the museum is based on (C5), 4) previous published pilot work related to the user group (C7), and 5) rules and functionalities created by the paper generation (C9). Kaasinen et al. (2015) include co-design in the empathic approach, according to the original idea of co-design by Sanders and Dandavate (1999). However, they also question whether co-design should be an approach of its own.

Our results seem to confirm the findings from previous studies by Väätäjä et al. (2012; 2015) and Kaasinen et al. (2015) where "Empathy" (e.g., user studies) was the most often used source for inspiration when defining experience goals. However, it is not evident if our respondents interpreted the given options for the sources of insight and inspiration in a similar way, since the provided descriptions in the questionnaire were brief. For example, it may not have been evident to the respondent from C5 that co-design was included in the "Empathy". Therefore, more elaborate descriptions for the sources, possibly with some examples, would be required in any future studies to avoid misinterpretations.

Case description	Empathy	Vision	Theory	Technology	Brand	Other	Total
C1. Industry: UI design for a tool to manage customer information in product development	X	X			X		3
C2. Industry: Developing paper machine quality control system	X	X			X		3
C3. Industry: Concept design approach Innoleap for industrial work activity. Example of a ship command bridge design case.		X	X	X	X		4
C4. Entertainment: Considering quality of experience in a location-based augmented reality horror adventure	X		X			X	3
C5. Entertainment/Education: Use of mobile augmented reality and outdoor education principles to create something for families visiting a museum.	X	X	X	X		XX	6
C6. Education/Well-being: Designing technology to combat (cyber)bullying in classrooms	X			X			2
C7. Health care/Well-being: Enhancing patient agency in spinal cord injury (SCI) rehabilitation	X		X			X	3
C8. Marketing: Studying experiences from packaging design using Online Research Community method.	X						1
C9. Informatics: Enhancing archival UIs with UX techniques		X		X		X	3
Total	7	5	4	4	3	5	

Table 4. Case descriptions and the sources of insight and inspiration when defining experience goals [n=9].

Case	Researcher	Designer	Topic expert/ Specialist	Developer	Management/ Employee/Client	User/ Target user	Students	Total
C1. Industry		X		X			X	3
C2. Industry	X			X		X		3
C3. Industry		X			X	X		3
C4. Entertainment	X	X						2
C5. Entertainment/Education	X		X		X			3
C6. Education/Well-being			X			X		2
C7. Health care/Well-being	X	X	X		X			4
C8. Marketing	X							1
C9. Informatics				X				1
Total	5	4	3	3	3	3	1	

Table 5. Participants in the definition process of the experience goals in each case [n=9].

Who Participated in Defining the Experience Goals?

When considering "Who participated in defining the experience goals?" the results in Table 5 shows that researchers (5/9 cases) and designers (4/9, e.g., UX designers, graphic designers and game designers) participated most often in the definition process of the experience goals. Topic experts or specialists (3/9) included an expert panel of educators, outdoor educator specialists, and consultants. Management / employees / clients (3/9) category included museum management and employees (C5) and the spinal injury unit (SIU) director (C7).

We noticed that although "Empathy" was the most often mentioned source for inspiration (7/9, see Table 4), users participated in the definition of the experience goals only in three cases. This seems to suggest that although understanding users is necessary for designers to become familiar with the design context, the experience goals are not always derived from the users. This may indicate that experience design seeks possibilities rather than aims at solving existing problems or evident needs. However, it is also possible that user participation during the experience goal definition was not feasible for some reason. Overall, when not involving users in the definition process of experience goals, there is a risk of basing the design on stereotypical views or assumptions. Still, designers with much previous experience from designing for a specific user group could arguably manage without actual user participation, but our data does not tell how experienced the respondents were with similar target user groups. However, some of the respondents did have a substantial experience from different experience design cases.

Chosen Experience Goals and Approaches to Communicate Them

Table 6 shows that the reported *targeted experiences* are rather unique in each case. However, industrial cases C2 and C3 shared a common experience goal: the sense/feel of control. Also, entertainment related cases C4 and C5, both utilizing augmented reality, had a similar goal: to provide an overall experience of curiosity.

Some of the reported goals seem to contain aspects related to good usability, such as "learnability" (C2) and "ease of use" (C7). Furthermore, in some cases, the usability quality was described from an experiential aspect, such as "feeling of efficiency" (C3). Some of the goals do not seem to be related to experiential aspects, such as "support outdoors education" or "dialogue". However, the complete definitions of these goals may have included also the experiential aspects. Still, it seems that in some cases, the chosen goals were a blend of experiential and more pragmatic goals.

Table 6 also shows *how the targeted experiences were communicated* among stakeholders. Brain-storming sessions, workshops and meetings (3/9 responses) were the most common ways to verbally discuss the experiences to aim for. Written reports and documentation (2/9) and sketch-level scenarios (2/9) were also used. For example, in the case C5 "around forty design concepts were sketched down and presented to stakeholders". In the case C7, "a generalized timeline of a patient's journey through the SIU" was created based on user observations during an ethnographic study.

According to the survey results, experience goals are communicated between stakeholders in writing, verbally and by using artefacts, such as personas and sketches. An interesting theme for future research could be how well these different ways manage to communicate the intended experience, in what phases during the development process they are used and how.

Aspects of a Good Experience Goal

Based on the responses to the sentence completion task, *a*

Case	Chosen experience goals	Approaches to communicate experience goals
C1. Industry	-	Documents, Verbal presentations
C2. Industry	1) Learnability, 2) Awareness, 3) Feel of control, 4) Success	-
C3. Industry	1) Being one with the ship and the sea, 2) Feeling of community, 3) Feeling of efficiency, 4) Feeling of trust towards peers, 5) Sense of control	Scenarios, Sketching
C4. Entertainment	1) Overall experience of curiosity, tension and 'black-humour' horror, 2) Feeling of presence, 3) Speculative play, 4) Support trajectories as journeys through hybrid spaces	Audiovisual material, Free play
C5. Entertainment/ Education	1) Arouse curiosity, 2) Focus on natural and cultural landscape, 3) Communicate author's life and authorship, 4) Support outdoors education, 5) Sustainable experience over time	Bodystorming, Brainstorming, Moodboards, Personas, Scenarios, Sketching
C6. Education/ Well-being	1) No-blame strategy: not blaming bullies, 2) Positivity, 3) Kind authority, not strict or punishing, 4) Dialogue	Reports, Academic publications
C7. Health care/ Well-being	1) Patient-centredness, 2) Ease of Use, 3) Ownership, 4) Network Navigation, 5) Projection	Workshops, Verbally in meetings, Ad-hoc interaction, Patient journey timeline
C9. Informatics	1) Bring user experience of archives closer to modern day web	-

Table 6. Chosen experience goals and approaches to communicate them. No answers from C8 [n=8].

good experience goal is 1) expressed clearly (so that all stakeholders understand it in the same way) (4/9 responses) and 2) precise enough to guide the design (4/9). Furthermore, the goal 3) should be achievable (3/9) and 4) involve emotion or the feeling users have while interacting with the product/service (3/9). In individual responses, a good experience goal 5) is grounded in research, 6) comes directly from the end user, 7) is related to the context of use, 8) can be evaluated, and 9) is a principle that drives creativity.

In the follow-up survey, one of the respondents commented that an experience goal does not necessarily have to be realistically achievable, as it can still act as a target, even if it cannot be fully reached.

Our results are mostly in line with the earlier findings by Väätäjä et al. (2012, 2015). However, the requirements of clear and precise descriptions of the goals were more distinctively emphasized in our findings.

The Model for Experience Goal Elicitation Process and Instructions for Defining Experience Goals

When evaluating the proposed model prototypes in the follow-up survey, respondents highlighted that experience goal elicitation is an iterative process. One of the respondents justified the choice for his/her favourite model: "Because it shows an iterative process. Cloud is a symbol for possible approaches (indicating there might be others)". Another respondent considered the boundaries between the "sources for insight and inspiration" and "approaches for processing information" to be blurry: "Participatory design can be one way of gaining such emphatic understanding, so the boundaries between the different boxes are not always clear to me."

Figure 1 illustrates the resulting model for Experience Goals Elicitation Process. On the left are potential sources for insight and inspiration, as described by Kaasinen et al. (2015). In the middle, information from the sources is iteratively processed by stakeholders using different approaches, such as brainstorming or co-design (examples come from the studied cases). The sources are overlapping with the approaches that "are also approaches to building sources for insight and inspiration", as one respondent commented. The iterative process produces a list of usually verbally described tentative experience goals, which are then prioritized. After making the selection of the target experience goals, they can be communicated to all stakeholders through different means, such as sketches, personas, and use scenarios (examples come from the studied cases). The means change depending on the stage of the project. The whole process is iterative, and during the communication with the stakeholders the experience goals can be further refined and more data can be gathered from the sources.

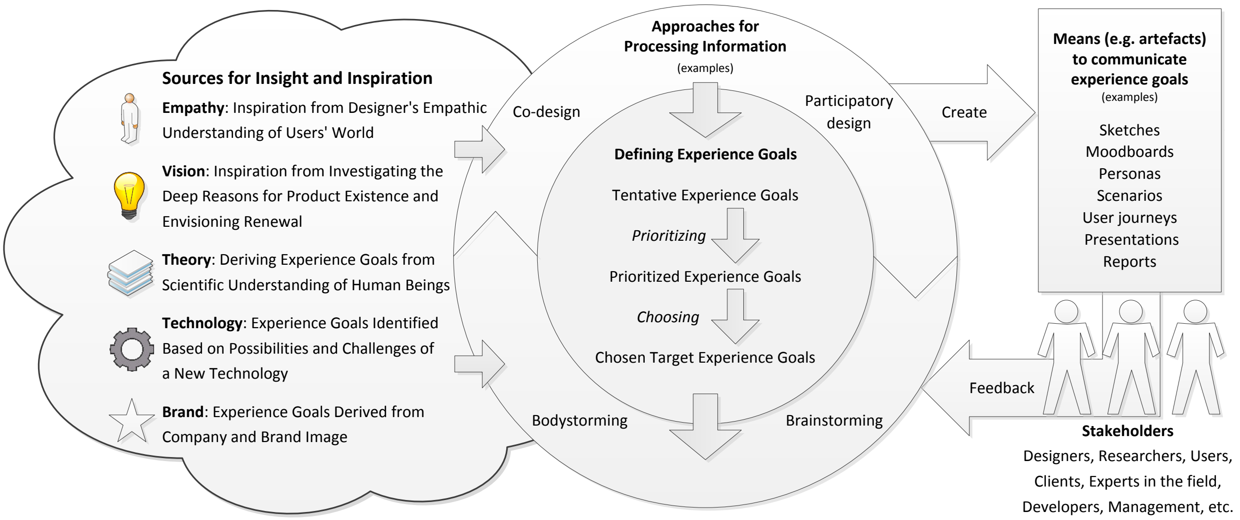

Figure 1. The model for Experience Goals Elicitation Process.

Instructions for defining and evaluating experience goals	
Describe, prioritize & choose	1. Use/choose methods and means to describe experience goals so that all stakeholders can create a shared and similar understanding. 2. Consider possible user requirements connected with the experience goals. You can also describe emotions or feelings the user is aimed to experience. 3. Describe goals precisely enough to make them actionable for designers in the design process. Describe also the reasoning behind the goals (why) as designers need to select the proper means of conveying (how) the experience (what). 4. Prioritize the experience goals to aim for and choose goals that can realistically be achieved (or at least targeted).
Communicate & iterate	5. Plan what means (e.g. artefacts) to use to communicate the experience goals for stakeholders. 6. Iterate the goals as you learn more throughout the design process. Revise what deliverables to use if you find better ways of communicating.
Measure & evaluate	7. If experience is measured, operationalize the experience goals and select appropriate (qualitative) metrics for evaluation. 8. Plan how to trace the later design solutions back to experience goals so that it is possible to evaluate the fulfilment of the goals in different phases of the design work.

Table 7. Instructions to support designers when defining and evaluating experience goals.

Although experience goals defined at the very beginning of an experience design project should guide the whole design process, in practice it is possible that the original goals are iterated later on, as designers learn more about the users and the context where the product or service will be used.

Finally, in Table 7, we provide a set of instructions to support the definition and evaluation of experience goals, derived from our results and previous literature (Väätäjä et al., 2012; Karvonen et al., 2012; Karvonen et al., 2014). Although not an exhaustive list to cover all possible steps in the experience design process, our aim is that these instructions can act as a check-list to support the beginning of the experience goal definition process and that different aspects of good experience goals are discussed with stakeholders.

CONCLUSIONS

In this paper, we have focused on one of the first challenges in experience design process: determining what experience to aim for. Our aim was to understand how experience goals are defined and communicated with stakeholders in the fuzzy front end of experience design. We have reported our findings from two survey studies of 9 experience design cases. As a result, we proposed a model of Experience Goal Elicitation Process and instructions to support the definition and evaluation of experience goals.

There are several limitations of our current study that should be taken into account in studies to follow: 1) the small sample size of rather dissimilar experience design cases, 2) lack of experience design projects ran by practitioners from industry, and 3) our data was based on participants' memories of the events, not necessarily what actually happened during the design process. From research perspective, it can be challenging to obtain a large sample of, especially industry-driven, design projects that actually follow the experience design process. Therefore, a more in-depth approach utilizing interviews and observations with practitioners could provide more valuable insights for experience design research and practice, even from a smaller number of cases.

We aim to iterate the model for Experience Goal Elicitation Process with practitioners from industry, since the participants of our current study were mainly from academia although several of them had work experience also from industry. Also, the model could be further iterated to better illustrate the experience design process in more specific domains, such as consumer electronics, educational games, or assistive technology, where different sets of stakeholders might be important.

Interesting research questions that have formed during this study include: 1) how experience goals can be prioritized and chosen, 2) how well different artefacts support the communication of experience goals for different stakeholders, and 3) how experience goals can be transformed to measurable design targets (e.g., Karvonen et al., 2014). With more research on these topics, we can hope to clear up the "fuzziness" of the fuzzy front end of experience design.

REFERENCES

Desmet, P. M. A. and Schifferstein H. From Floating Wheelchairs to Mobile Car Parks: a Collection of 35 Experience-Driven Design Projects. The Hague: Eleven Publishers, (2011).

Hassenzahl, M. The Thing and I: Understanding the Relationship Between User and Product. In Blythe, M.A., Overbeeke, K., Monk, A. F., Wright, P. C. (eds.) Funology: From Usability to Enjoyment. Kluwer Academic Publishers, The Netherlands, (2003).

Hassenzahl, M. and Tractinsky, N. User Experience - a research agenda [Editorial]. Behavior & Information Technology, 25(2), (2006), 91-97.

Hekkert, P., Mostert, M. and Stompff, G. Dancing with a machine: a case of experience-driven design. In Proc. DPPI 2003, ACM New York (2003), 114-119.

ISO 9241-210 (2010): Ergonomics of human-computer interaction – Part 210: Human centered design process for interactive systems. Geneva:International Standardiziation Organization (ISO).

Karapanos,Kaasinen, E., Roto, V., Hakulinen, J., Heimonen, T., Jokinen, J., P., P., Karvonen, H., Keskinen, T., Koskinen, H., Lu, Y., Saariluoma, P., Tokkonen, H. and Turunen, M. Defining User Experience Goals to Guide the Design of Industrial Systems. Behaviour & Information Technology journal, Taylor & Francis, (2015). DOI: 10.1080/0144929X.2015.1035335

Karvonen, H., Koskinen, H. and Haggrén, J. Defining User Experience Goals for Future Concepts. A Case Study. In Väätäjä, H., Olsson, T., Roto, V. and Savioja, P. (eds.) NordiCHI2012 UX Goals 2012 Workshop Proceedings, Tampere: TUT Publication series (2012), 14-19.

Karvonen, H., Koskinen, H., Tokkonen, H. and Hakulinen, J. Evaluation of User Experience Goal Fulfillment: Case Remote Operator Station. In Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality, Springer International Publishing (2014), 366-377.

Khurana, A. and Rosenthal, S.R. Towards holistic ‘front-end’ in new product development. J. Prod. Innovat. Manag. 15, 1 (1998), 57-74.

Liinasuo, M. and Norros, L. Usability Case - Integrating Usability Evaluations in Design. In COST294-MAUSE Workshop, (2007), 11-13.

Mahlke, S., and Thüring, M. Studying antecedents of emotional experiences in interactive contexts. In Proc. CHI 2007, ACM (2007), 915-918.

Sanders, E. B.-N. and Dandavate, U. Design for Experiencing: New Tools. In Overbeeke, C.J., and Hekkert, P. (eds.): Proc. of the First International Conference on Design & Emotion, (1999), 87-92.

Soley, L. C. and Smith, A. L. Projective Methods for Social Science and Business Research. The Southshore Press, Milwaukee (2008).

Väätäjä, H., Olsson, T., Roto, V. and Savioja, P. (eds.) How to Utilize User Experience Goals in Design? In UX Goals 2012 Workshop Proceedings. Tampere: TUT Publication series (2012).

Väätäjä, H., Savioja, P., Roto, V., Olsson T. and Varsaluoma, J. User experience goals as a guiding light in design and development – Early findings. In INTERACT 2015 Adjunct Proceedings. Univ. of Bamberg Press (2015), 521-527.

Wright, P., McCarthy, J. and Meekison, L. Making sense of experience. In: Blythe, M.A., Overbeeke, K., Monk, A. F., Wright, P. C. (eds.) Funology: From Usability to Enjoyment. Kluwer Academic Publishers, The Netherlands, (2003).

P2

EXPLORING MOTIVATIONAL ASPECTS AND USER EXPERIENCE OF MOBILE MATHEMATICS LEARNING SERVICE IN SOUTH AFRICA

by

Varsaluoma, J., Väätäjä, H. and Walsh, T, 2016.

In Proceedings of the 20th International Academic MindTrek Conference (AcademicMindTrek'16). ACM, New York. pp. 159-168.

Exploring Motivational Aspects and User Experience of Mobile Mathematics Learning Service in South Africa

Jari Varsaluoma
Tampere University of Technology
Korkeakoulunkatu 10,
33720 Tampere
+358 40 849 0856
jari.varsaluoma@tut.fi

Heli Väätäjä
Tampere University of Technology
Korkeakoulunkatu 10,
33720 Tampere
+358 40 849 0720
heli.vaataja@tut.fi

Tanja Walsh
University of York
Deramore Lane, Heslington East,
York
YO10 5GH, United Kingdom
tanja.walsh@york.ac.uk

ABSTRACT

The rapid rise in the number of mobile phone users in developing countries has created an opportunity to research and develop new mobile learning services. However, designing mobile learning services that are both motivating and provide positive user experiences is crucial for being successful in mobile learning markets. In this paper, we study the motivational factors for studying mathematics and using mobile mathematics learning service Microsoft Math in South Africa. The relationship of learners' motivational factors, UX, math skills and behavioral intentions are studied. No significant relationships were found between UX and motivational factors. As a result, we provide 1) the summaries of the identified motivational factors for studying mathematics and for using mobile mathematics learning services, 2) UX goals for designing mobile learning services for developing countries, and 3) implications for conducting remote mobile surveys in developing countries.

CCS Concepts

• Human-centered computing~Empirical studies in HCI • Applied computing~E-learning.

Keywords

Motivation; User experience; Mobile learning; Informal learning; UX Goals; Developing countries.

1. INTRODUCTION

Access to mobile devices, such as mobile phones, has increased tremendously in the developing countries during the last decade. Today, developing world is more mobile than the developed world as most phones are owned by people living in low-income regions [25]. From mobile learning perspective, there is huge potential for new mobile learning services as more learners in developing countries now have access to mobile internet. Design and development of mobile learning (m-learning) services has strongly focused on learners in western countries and more research is needed regarding the design of mobile learning services for learners in developing countries [21].

In this paper, we study the motivational and user experience (UX) factors and their relationships regarding the use of a mobile mathematics learning service by high-school students in South Africa. Other factors, such as math skills and behavioral intentions were also studied. Developing mobile learning services that are motivating and provide positive UX for their users is crucial in being successful in mobile learning markets. By understanding factors that motivate learners in using mobile learning services and that provide positive experiences, in different cultures and learning contexts, designers will have better changes in creating mobile learning services that are not only taken into use but also used over longer periods of time.

Subjective data was collected with a remote mobile survey from the learners who had been using the evaluated mobile mathematics learning service. In addition, objective data on use of the service was gained by analyzing the logged usage data from the same respondents. This approach gave us a more comprehensive image on how learners evaluate and actually use the service.

Based on the findings, we propose three UX goals for supporting the design of mobile mathematics learning services for learners at high-school level to be part of the math curricula and didactics for developing countries. We also share our experiences in conducting remote mobile surveys in South Africa to support researchers working in similar context in the future. Finally, topics for future research are suggested.

2. RELATED WORK

2.1 Motivation in mobile learning

The self-determination theory states that motivation can be classified as *intrinsic motivation*, *extrinsic motivation* and *amotivation* [1]. Intrinsic motivation is related to performing the activity because it is perceived as fun and pleasurable. External rewards, such as money and public recognition, drive a person to perform in case of extrinsic motivation. Different types of extrinsic motivation have been proposed, two of them being *external regulation* and *identified regulation* [1]. External regulation occurs when individual acts only in a pursue of a reward or to avoid sanction. Identified regulation occurs when a person values the activity and chooses to act by oneself, however still with a specific goal in mind, such as learning a new skill. Amotivated person has no motivation, intrinsic or extrinsic, when completing a task. Both intrinsic and extrinsic motivations can be present at the same time.

Sharples et al. (2007) define mobile learning as "the processes of coming to know through conversations across multiple contexts amongst people and personal interactive technologies". A more practical definition was used in MoLeNET programme [18]: "The exploitation of ubiquitous handheld technologies, together with wireless and mobile phone networks, to facilitate, support,

AcademicMindtrek'16, October 17-18, 2016, Tampere, Finland
© 2016 ACM. ISBN 978-1-4503-4367-1/16/10…$15.00
DOI: http://dx.doi.org/10.1145/2994310.2994369

enhance and extend the reach of teaching and learning." Several more definitions for mobile learning (or m-learning) are available in m-learning literature.

Jones et al. (2006) define *informal learning* as "learning that is outside institutional contexts". They suggest six factors explaining why using mobile devices for informal learning could be motivating: 1) control (over goals), 2) ownership, 3) fun, 4) communication, 5) learning-in-context, and 6) continuity between contexts. These factors were further iterated in a workshop "Affective factors in learning with mobile devices" [19], where motivational factors and barriers for mobile learning were discussed. The conclusions from the workshop were that the six proposed motivational 'features' require more iterations as there is overlapping with other concepts (e.g. cool and fun, control and ownership). Affect and especially motivation were seen as key areas for further research. Understanding learners' previous and current use of technologies were also considered as important factors in understanding motivation to use mobile technologies.

Venkatesh (2000) studied in a longitudinal study how intrinsic motivation is related to the perceived ease of use by using the concept of computer playfulness operationalized as feeling spontaneous and creative while using the system. Users use intrinsic motivation (computer playfulness) as one anchor when forming perceived ease of use about a new system [23]. With increasing experience in using the system, perceived enjoyment becomes more important in forming the perception of ease of use [23].

2.2 Mobile learning in South Africa

Vainio et al. (2015) studied the effect of South-African cultural context on user experience of mobile mathematics service in a longitudinal research with over 30 South African schools during three years. Their aim was to identify culturally sensitive areas in the local context of mobile learning services in order to give insight for localization of these services.

To understand the learning culture in South Africa, or in any learning culture, one needs to know the educational system. In 2009, there were 24693 public schools in South Africa (12 million pupils) and 1174 private schools (386098 pupils) [6]. The student-to-teacher ratio in South Africa is high, 16 to 1 in private schools (primary, secondary, middle and combined) [6]. According to the 2003 Trends in Mathematics and Science Study [20], there is a challenge in mathematics teaching in South Africa because the international average maths score in 2003 was 467 whereas in South Africa it was 264. In addition, there is discrepancy in mathematics achievement across provincial, gender, economic and racial divides [20]. To solve the challenges, several initiatives has been established, focusing on training teachers to meet current and future requirements and ensuring that adequate measures for pupils to move from secondary education into higher education (HE) institutions or the labour market [8]. However, the role of technology is not specified at all, therefore leaving the field open for research into the possibilities offered by technology.

Vainio et al. (2015) found several issues to consider in South-African learning context: The first issue is the level of mobile network coverage, which is a fundamental requirement for mobile learning services. South Africa has the most advanced telecommunications network in Africa. This is a good starting point for developing mobile mathematics learning service in South Africa. The second issue is the level of mobile penetration: in 2009, South Africa had 46.4 million mobile telephones in use, as well as 4.4 million Internet users [6]. The third issue to consider is the language used at school. South Africa has about 50 million people of diverse origins, cultures, languages, and religions and eleven official languages are recognized in the constitution. Two of these languages are of European origin: English and Afrikaans. Although English is commonly used in public and commercial life, it is only the fifth most spoken home language. The education sector does not totally reflect the multilingual nature of South Africa. English is often used as the medium of instruction at the expense of Afrikaans and African languages [6]. Fourthly, in the study of Vainio et al. (2015) the content of the mobile mathematics material was aligned with the South African curriculum and level of maths. In a context where math results tend to decline substantially from Grade 9 to Grade 10 in South African Public Schools, the pupils who used the Nokia Mobile Mathematics service regularly (completing more than 15 practice exercises and tests) achieved results for Grade 10 were 7% better on average than their peers.

The sixth issue is how schools and education systems allow the use of mobile technology during school hours. In the study of Vainio et al. (2015), 81% of the schools participating in the study had an ICT policy or school code of conduct that restricts the use of mobile telephones during school time. The seventh issue is to understand how much pupils and teachers use the mobile learning service. In 2009, 85% of the pupils had mobile phones with SMS capacities and 64% of them were able to use the browser based learning system with their mobile phones. The average posts per week for those pupils who used browser-based service was 3.99 posts per week and for SMS based service users 1.69 per week. In 2010, there were altogether 2875 registered users and 1528 of them were active users. Of these pupils, 75% reported that they had their own mobile phones and 67% reported that their mobile phone could download browser-based service. In addition, 17% reported that they could access a shared mobile telephone, which could download browser-based service, at home. However, 13% of case study pupils were unable to use either their own or a shared mobile phone.

Vainio et al. (2015) found that the pupils' attitude towards learning mathematics improved when they had been using mobile learning service and learning mathematics was seen enjoyable or fun. Vainio et al. (2015) found also that in South African environment it is important to encourage the informal use of the service as the churn rate of teachers at school is high and the information disappears when a teacher leaves the school. This finding encouraged the development of the service towards a more informal way of using it: independently and outside school hours. Because of the low teacher-to-pupil ratio in public schools, mobile learning service could help teachers to keep up with the pupils' learning progress as well as give teachers more ways to communicate with the pupils than in a traditional class set up.

2.3 User Experience Goals

Various available definitions for UX underline the richness and complexity of this concept. However, in the field of HCI, it is widely agreed that UX is dynamic, context-dependent and subjective [12]. ISO 9241-210 (2010) standard defines UX as "*a person's perceptions and responses that result from the use or anticipated use of a product, system or service*". The components of UX include pragmatic or utilitarian (related to usability and utility) and hedonic or non-utilitarian (stimulation, identification visual aesthetics etc.) aspects of product use [4, 14]. Furthermore, when interacting with the system, user's perceptions of these qualities are linked to emotional reactions, which are presented as the third UX component [14].

Designing for experiences is increasingly in the interest of both Human-Computer Interaction (HCI) practitioners and researchers. User experience (UX) goals are one means that can be used to describe the experiential requirements for an interactive system, and guide their design and evaluation [24]. They concretize the experiences that are intended to be experienced by the users of the interactive systems [24]. Commonly agreed experience goals help the entire project team by "keeping user experience in focus through the multidisciplinary product development and marketing process" [10].

Väätäjä et al. (2015) report the characteristics of good experience goals based on the questionnaire responses of researchers working on nine case studies varying from workplace to consumer applications and education [24]. A good experience goal 1) helps the design as a guiding light, 2) is measurable, 3) describes positive emotions, and 4) is a way to communicate the desired experience with other people. Furthermore, experience goals were considered useful in keeping the focus on important issues and providing inspiration.

Five sources to gain insight and inspiration for setting experience goals in industrial environments have been suggested [10]. The sources were derived from four industrial design case studies with companies, and were supplemented with a literature study (ibid.). Sources include brand, theory, empathy for users, possibilities and challenges of technology, and vision. Varsaluoma et al. (2015) report based in nine case studies that empathy – empathic understanding of the user's world – is the most often used source for inspiration to set the experience goals [22].

For example, Lu and Roto (2015) describe using the experience goal approach that drives the design ideation process in educational context [13]. An e-learning tool was developed for beginner forklift truck drivers to reduce the workload of senior drivers who traditionally teach the new drivers. Based on interviews with warehouse workers as well as the designers' own experience of learning to drive a truck, three experience goals for the design were identified: 1) Security by feeling guided even without a human teacher present; 2) Competence by balancing the feeling of incompetence and over-confidence; 3) Stimulation by the enjoyment of learning. Solution included a virtual eye that was implemented that followed the driver and provided feedback in natural language, the self-assessment of the success of the exercise to support reflection, and by gamified elements for feedback. However, user experience goals have rarely been reported and mentioned as driving and guiding the design of educational systems. Our work contributes to this body of knowledge, by deriving user experience goals from the findings to support designing mobile mathematic learning services for developing countries.

3. METHOD

A remote mobile survey was designed by a research team in Finland and iterated together with context experts from South Africa. Data was collected using two methods: 1) a questionnaire accessible with mobile devices and 2) log-data describing the actual usage patterns of the service.

3.1 Microsoft Math Service

The studied service was Microsoft Math (earlier Nokia Mobile Mathematics) that is an international mobile mathematics learning service, intended for high-school students to practice mathematics especially informally, e.g. outside school hours. The service is accessible through web browsers for all data-enabled handsets. Using the service does not cause any mobile data transfer costs for learners in South Africa.

The service was launched 2008 in South Africa, its content following the local curriculum. Starting with 300 learners in 2009, in 2012 the service had reached 50 000 learners. In 2014, the service was available nationwide to all learners and was also launched in Tanzania with localized mathematics content.

Microsoft Math offers math examples, theory and exercises on different difficulty levels (see Figure 1 for screenshots). Users can collect points by completing quizzes, compete with each other, create study groups (or classrooms) and send messages within the groups. The service is freely available at https://math.microsoft.com/.

3.2 Instruments

A mobile survey was chosen as a data collection method as we expected it to be fast, inexpensive and reliable way of reaching many learners around South Africa. To make certain that the participants would have an easy access to the questionnaire without any mobile data transfer costs, we decided to integrate the questionnaire into the Microsoft Math platform. Although this set some restrictions for the format of the questions that could be used, we expected to get a better response-rate when there were no additional data-transfer costs for the participants. Furthermore, the questionnaire could be advertised in the front page of the service itself and answering the questionnaire would be easy in the familiar environment for the users of the service. Also, by using the existing platform, we could be certain that the questionnaire would function correctly with every respondent's mobile device.

The questionnaire was divided into two parts to decrease the

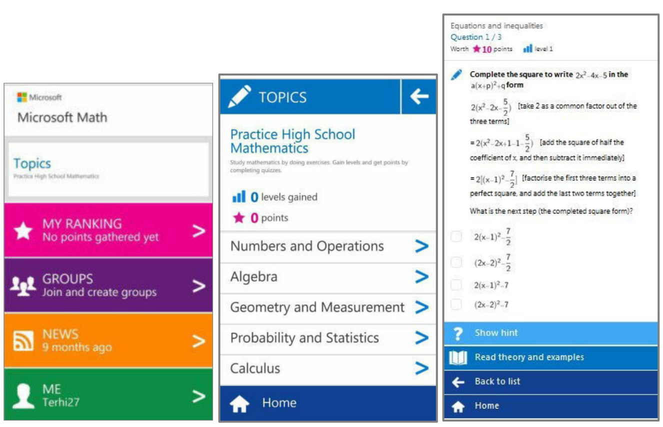

Figure 1. Microsoft Math screenshots.

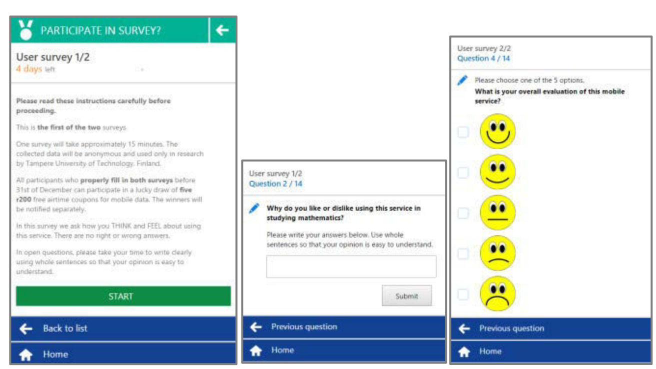

Figure 2. Examples of the mobile questionnaire pages in Microsoft Math.

participant fatigue, allowing a break for the respondents between answering both parts while using mobile devices. Both parts had 14 questions, each one presented on a separate page (see Figure 2 for examples). Due the restrictions of the platform, only single-choice Likert scale questions (with five options) and open-ended questions were used. The questions were divided into five main topics: 1) user experience, 2) motivation, 3) behavioral intentions, 4) context, and 5) background and math skills. In addition to the questionnaire, data logging was used to gather actual usage data from the participants who answered the questionnaire. This data had been gathered since the participants had registered to the service. Table 1 presents the questions and logged usage data that were used for measuring the factors of these topics. To keep the mobile questionnaire short, only a few questions were used for each topic.

User experience. Two questions measured the overall UX (Q1 and Q2, see Table 1) and one the fulfillment of hedonic goals (Q3). For overall UX, emoticons were used to offer a non-verbal way to evaluate how learners experienced the system (see Figure 2). Ease of use (Q2), related to pragmatic usability, was derived from the UMUX scale [2] and Q3 from the Attrakdiff2 questionnaire to measure the hedonic quality Beauty [5]. To keep the questionnaire short, no other aspects of UX were measured.

Motivation. We wanted to understand how motivated (i.e. "interested") learners were and what motivated them to study mathematics in general (Q4, Q5). The motivation to use the service was measured in a similar fashion (Q6, Q7), but also an abridged version of the Situational Motivation Scale (SIMS) [3] was used (Q8). SIMS measures four motivational constructs: intrinsic motivation, identified regulation (extrinsic motivation), external regulation (extrinsic motivation) and amotivation. The whole amotivation subscale and one item on each remaining scale were removed to reduce the length of the questionnaire. Also, to make the SIMS questions easier to comprehend and to better fit the scale on mobile screens, the scale was changed from "Does not correspond at all 1–7 Corresponds exactly" to "Strongly disagree 1–5 Strongly agree".

From the log data, our aim was first to measure the number and length of log-in sessions, but these data were not comparable because some users had used the "remember me" option for log-in settings. Instead, we decided to use the number of completed separate quizzes for the measurement of motivation to use the service (L1). In Microsoft Math, each mathematical topic had quizzes that consisted of three problems to be solved. Normal quizzes were between difficulty levels 1 to 10, after which a bonus level 11 could be repeatedly completed to improve the user's total score. Upon discussions with the developers, it was revealed that some learners may have earlier repeatedly completed bonus quizzes in order to win competitions that were based on collected points. For this reason, we did not use the total score or count the repetitions of the same quizzes for measuring the motivation by actual usage.

Behavioral intentions. Participants were asked how likely they were a) going to continue to use the service (Q9) and b) recommend the service to their friends (Q 10).

Context. We were interested to see if the service is most often used outside school hours, suggesting that it supports informal learning (Q11). Also, do learners prefer to use the service or more traditional methods for studying mathematics (Q12)?

Background and math skills. Measured background variables were the respondents' age, gender, occupation and math skills. Perceptions of the respondents' own math skills were measured with two task difficulty scales asking 1) how hard math is in general (Q14) and 2) how hard the math questions in the service are (Q15). Since the math tasks in the service follow the curriculum in South Africa, Q15 seemed to be a suitable measurement for perceived math skills.

3.3 Participants and Procedure

The questionnaire was advertised in the news-section of the Microsoft Math service and was open for one month in December 2014. Two additional news-messages were sent near the completion of the study. Respondents who completed both parts of the questionnaire could participate in a lucky draw of five free airtime coupons for mobile data worth 200 ZAR (~ USD 16.50) each.

Table 1. Main topics and used questions. Q#=questionnaire and L#=log data. Note that the order of the questions has been altered from the questionnaire to improve readability.

User experience	**Q1.** What is your overall evaluation of this mobile service? (Emoticons: Sad 1–5 Happy) **Q2.** This service is… (Difficult to use 1–5 Easy to use) **Q3.** This service is… (Ugly 1–5 Beautiful)
Motivation	**Q4.** I am interested in studying mathematics in general. (Strongly Disagree 1–5 Strongly Agree) **Q5.** Why are you interested or not interested in studying mathematics in general? **Q6.** I like to use this mobile service in studying mathematics. (Strongly Disagree 1–5 Strongly Agree) **Q7.** Why do you like or dislike to use this service in studying mathematics? **Q8.** *Abridged version of SIMS with 9 questions. Example question for measuring intrinsic motivation:* Why do you use this service to study mathematics? Because I think that this activity is interesting (Strongly Disagree 1-5 Strongly Agree) **L1.** Number of completed separate quizzes.
Behavioral intentions	**Q9.** How likely are you going to continue using this service in future? (Not at all likely 1–5 Very likely) **Q10.** If your friends had a mobile phone access, how likely would you recommend this service to your friends? (Not at all likely 1–5 Very likely)
Context	**Q11**. When do you most often use this service? (On my free time / During school or work hours) **Q12.** Which one would you prefer to study mathematics? (This service / Textbook with pen and paper)
Background & Math skills	**Q13.** Year of birth, gender and current occupation. **Q14.** In general, how hard is math for you? (Very hard 1–5 Very easy) **Q15.** In general, I think the math questions in this service are… (Very hard 1–5 Very easy)

65 responses were received to the first part (opened 91 times) of the questionnaire and 53 responses to the second part (opened 65 times). In total, 53 respondents (22 female, 42%) answered both parts of the questionnaire and their responses were used in the analysis. Only three of these respondents had not answered both parts of the questionnaire during the same day, suggesting that the length of the survey was not an issue for a majority of the respondents.

Respondents' age varied from 14 to 42 (M = 18.3, SD = 3.87). One respondent was less than 16 years (14) and three were more than 19 years old (21, 30 and 42). 45 (85%) of the respondents were in a primary school or college, 5 (9%) were higher education students, one (2%) in working life and 2 (4%) in other life situation ("none of the previous").

Most participants assessed (scale: "Very hard 1 – 5 Very easy") math in general (Q14: M = 3.74, SD = 1.02) and the math questions in the service as easy (Q15: M = 3.66, SD = 0.85). 5/53 (9%) respondents considered math in general to be very hard or hard (responded 1 or 2) and only three (6%) considered the questions in the service to be very hard or hard. 36 (68%) rated math in general to be easy or very easy and 31 (59%) the questions in the service to be easy or very easy. Spearman's correlation showed a significant, moderate relationship between the responses to Q14 and Q15 ($r_s = .47$, $p < .001$). The responses suggest that our sample was somewhat biased in a sense that majority of the respondents considered themselves to be skilled in math.

3.4 Analysis of the results

Qualitative data (Q5 and Q7, see Table 1) were initially coded based on similar responses to the questions. The initial coding was then viewed by another researcher, after which another round of analysis was carried out by the first researcher, resulting in the final coding and categorization of the open comments. We also coded the resulting categories for Q5 with the intrinsic and extrinsic motivational types [1]. Quantitative data were analyzed with SPSS. Shapiro-Wilk tests for normality showed that the data deviated from a normal distribution, and therefore non-parametric tests were used in further analysis. Log data (L1 in Table 1) from each respondent were used in SPSS for evaluating the actual usage of the service and for making subgroup analyses.

Cronbach's alphas for the abridged SIMS was acceptable for the external regulation subscale ($\alpha = .82$), but not for the other two subscales: intrinsic motivation ($\alpha = .40$) and identified regulation ($\alpha = .29$). Since the removal of any items did not significantly increase the alpha, each item of these two subscales was analyzed individually.

Initial UX goals were derived from the analysis results by the researcher who was the most familiar with the data. The initial goals were then discussed and iterated within the research team, resulting in the final set of proposed UX goals.

4. RESULTS

4.1 User Experience

The overall experience that respondents had with Microsoft Math was very positive (Q1: M = 4.74, SD = 0.52). Majority also found the service to be easy to use (Q2: M = 4.74, SD = 0.71) and beautiful (Q3: M = 4.85, SD = 0.36). No significant correlations were found between the three UX factors.

4.2 Motivation

Based on the responses, participants were very interested in studying mathematics in general (Q4: M = 4.75, SD = 0.62). They also liked to use the mobile service to study mathematics (Q6: M = 4.74, SD = 0.63). Spearman's correlation showed a statistically significant, but weak relationship between the two factors (r_s = .34, $p < .05$).

Motivation to study mathematics. Table 2 shows the categorization of reasons for why studying mathematics was found interesting or not. No negative reasons were reported. Based on the results, the main types of motivation for studying mathematics seemed to be both extrinsic (identified regulation), and intrinsic.

The most often mentioned motivation (25/53 responses, 47%) for studying mathematics related to learners' future profession or career, where they would require math skills. For example:

Mathematics is fun and interesting but above all mathematics opens many opportunities to becoming successful because in many fields of work maths is required as it shows how mentally sharp you are and how good you are at solving problems.

Mathematics was also seen important for solving problems in real life:

I am interested in learning maths generally because I believe that I will need it in future as I will be taking a career that needs maths and it also helps me to solve real life problems with patience.

15 (28%) respondents seemed to be intrinsically motivated as they simply liked to study mathematics because it is interesting or fun:

I am interested in studying mathematics in general because it is the only subject I find interesting! it is fun, cool and not time

Table 2. Reasons why studying mathematics is interesting or not interesting (Q5, no negative responses)

What motivates learners in studying mathematics?	Responses
For my future (identified regulation): Math is studied in order to secure future studies, profession or career. It is something that everyone needs, it can make life easier and help solving real life problems on personal and global scale.	25/53
I like to study mathematics (intrinsic motivation): Studying mathematics is interesting, amazing or fun. I enjoy the challenges, complexity and working with numbers. Practicing math and solving problems gives me satisfaction, sense of accomplishment, optimism, confidence and pride. Math gives me motivation to think, new ideas and can free my mind. It has become a hobby for me.	19/53
Improve my math skills (identified regulation): I want to increase my understanding, math skills or thinking capacity.	14/53
For school (external regulation): Math is needed at school, it can help with other topics and improve my marks.	4/53
Math is easy (intrinsic motivation): I am good in math and it is easy for me.	2/53
Help fellow learners (identified regulation): I want to help fellow learners who are struggling with math.	1/53
It supports my hobby (identified regulation): I combine math with my hobby, chess.	1/53

wasting! I love doing calculations than theory.

…or because they enjoyed the challenges that mathematics offers:

I love mathematics because math makes me to think out of the box. It brings everyday challenges that need to be solved. What I like the most about math is that you have to be a critical thinker.

Maths is a very interesting subject. It can be challenging at times but once you overcome a challenge you get filled with joy, confidence and pride.

14 (27%) respondents were motivated by the improvement of their math skills and thinking capacity. As one respondents commented:

I believe everything in life is about maths. learning maths in general increases my intelligence, my speed of thinking accurately and solving problems I face in my life. Having a great knowledge of maths in general makes life very easy and nice to live.

Other motivations for studying mathematics in general were that 1) math is needed at school and improving marks (4 responses), 2) math was found as easy subject (2), 3) learner wanted to help others struggling with mathematics (1), and 4) mathematics supported another hobby (1).

Motivation to use Microsoft Math. Table 3 shows the categorization of reasons why respondents liked or disliked using Microsoft Math to study mathematics.

More than half of the respondents' comments (27/53, 51%) related to the content offered by the Microsoft Math as a motivational factor to use the service. The wide variety of questions and clear examples on different topics were liked:

I like the variety of questions asked about a topic, and the manner in which the theory is explained after I get it wrong or right.

I like this service because it is an interactive way of studying mathematics and it shows you the answers and the correct way of solving the problems for those who struggle to get to it.

I like to use this service to study mathematics. The reason is that it has all contents of mathematics and easy to access. I like it again because it helps me a lot, I was struggling with Math but now I'm much better! Thank you for the service.

One respondent also appreciated the change to learn more on topics that were rarely discussed at school:

This software/application is a great studying tool as it challenges us to do better and it gives us totally new fields to explore in our mathematics subject, fields that we rarely do at school.

15 (28%) respondents stated that they either wanted to or already had improved their math skills and marks with the help of the service. As some of the respondents commented:

I like it because it makes me have a better look and understanding on answering and solving the questions. This service has also, as I believe, improved my paper 1 maths, so it's so awesome and simplified.

I like to use this service because it really improves my knowledge in maths, using this service had also improved my marks in school.

I like to use it because it helps me. I have a problem with maths and since I started using it my maths skills have improved and I now understand a lot of things and ways to do them differently.

8 (15%) respondents appreciated the good accessibility of the mobile service, e.g. when compared with traditional textbooks:

I like to use this service because through my phone I can easily get access to math's whenever and wherever I want unlike when

Table 3. Reasons why learners liked or disliked using Microsoft Math (Q7).

What motivates or unmotivates learners to use Microsoft Math to study mathematics?	Responses
High-quality content supports informal learning: The service helps to understand math questions and different topics by offering theory and clear explanations of correct solutions. New topics are learnt by reading examples and tips. High quality contents with wide variety in questions also in topics rarely explored at school.	27/53
Improves my math skills: Has improved or will improve my math skills and math marks	15/53
Accessibility: Easy to access anywhere and anytime (unlike textbooks).	8/53
Fun way of learning mathematics: Fun, enjoyable, exciting or awesome way of learning mathematics.	7/53
Supplements teaching at school: Similar content helps with questions that will be asked at school or topics that learners did not understand in class. Can be more understandable than textbooks.	7/53
Competition: Competing with fellow learners.	6/53
Ease of use: Simplified and easy to use	5/53
Makes me use more time on studying: I am more active and get to practice mathematics a lot. It gives me courage to practice math. Less time wasted on luxurious services.	4/53
Efficiency: Fast, efficient and saves time (compared to paging textbooks)	3/53
Prizes: Competitions and possibility to win prizes	2/53
Inexpensive: Uses only little airtime	2/53
Measure my progress: Points tell which topics to focus on more and help measuring my knowledge.	2/53
Novel learning experience: Interactive way of studying mathematics. Modern learning experience.	2/53
Unmotivating (1 response for each): Not enough questions on each level. Examples are not clear. Slow loading times. Feedback not replied fast enough.	4/53

using books which I can't carry everywhere I want to go.

I like to use this service because it saves time, improves my psychological power, so there is no need to spend most time paging like in textbooks It is like my teacher in the pocket because it provides with both questions and answers and I can use it anywhere at any time.

I'm someone who is always online, so I like to use this service because I'm also able to practice and enjoy maths without getting to my books. I get to be more educated and skilled in maths through this service.

7 (13%) reported that the using the mobile learning service for studying mathematics was fun, exciting or enjoyable:

It is a fun way for practicing maths and one gets to compete with other students across the country.

It's a much better and a fun way of practicing maths.

In 7 (13%) responses it was emphasized that Microsoft Math can supplement teaching at school. One respondent found it easier to comprehend than textbooks:

I like to use this service because it makes mathematics simpler and more understandable than textbooks…

Others liked that the content is similar to what is discussed at school and it can therefore help understanding topics discussed in class. Also, using the service can help when preparing for exams:

I like it because it helps me with future questions that will be asked at school. I also like it because most of the questions the service provides to me are common to the questions we are given at school so it gives me more confidence to answer the questions.

This service helps me to understand things I did not even understand in class.

I like to use this service because l noticed that the questions l get to answer here are actually the questions that appear during exams so with this service l stand a better chance of scoring higher marks.

6 (11%) learners were especially motivated by competing with each other. For some this was related to why using the service was considered fun:

…Also I like to be competed since competition is a performance stimulant that's why I like this service because through it I can compete with many people.

It is a fun way for practicing maths and one gets to compete with other students across the country.

I like to use this service because I get to measure my knowledge up to so far and I get to see how good am I compared to other students

5 (9%) respondents mentioned ease to use as a motivational factor to use the service. 4 (8%) had noticed positive change in their behavior, such as using more of their free time for studying or getting more courage to practice mathematics. 3 mentioned efficient and fast usage of the service. Only 2 respondents mentioned the following as motivational factors: 1) possibility to win prizes from competitions, 2) using the service is inexpensive, 3) measuring one's progress by following the points gained in different topics, and 4) novel experience for studying mathematics.

Only four negative aspects were mentioned by single respondents: 1) not enough questions on each level (respondent preferred to have 5-6 questions instead of 3), 2) examples are not clear, 3) slow loading times, and 4) feedback not replied fast enough.

In conclusion, the open comments also indicate that majority of the respondents liked to use the Microsoft Math service and that there were several factors that motivated them. Especially 1) the content provided by the service and 2) the respondents' interpretation that the service already has or will improve their math skills were considered important aspects when using the service.

Motivation to use Microsoft Math based on SIMS. Table 4 presents the results of the abridged version of SIMS (Q8). Cronbach's alpha was acceptable only for the subscale External regulation. The results from SIMS seem to be in line with the qualitative data regarding motivation to study mathematics in general. Ratings were high for all intrinsic motivation items and very high for all identified regulation items. This suggests that the use of the service is motivated by the activity itself but even more so because respondents see the value for themselves in studying mathematics with the service. External regulation was in the middle of the scale with higher SD, suggesting that perhaps external rewards such as prizes from competitions or competing with fellow students have motivated some respondents to use the service, but clearly not everyone.

Table 4. Results from the abridged version of SIMS (Q8).

Why do you use this service to study mathematics? (Strongly Disagree 1-5 Strongly Agree) (n=53)	Mean (SD)
INT1 Intrinsic motivation: Activity is interesting	4.81 (0.59)
INT2 Intrinsic motivation: Activity is pleasant	4.43 (0.87)
INT3 Intrinsic motivation: Activity is fun	4.57 (1.05)
IDE1 Identified regulation: I am doing it for my own good	4.85 (0.36)
IDE2 Identified regulation: Activity is good for me	4.87 (0.39)
IDE3 Identified regulation: Activity is important for me	4.91 (0.35)
EXT External regulation subscale	3.03 (1.35)

Spearman's correlation showed statistically significant, but weak correlations between the following motivational scales:

- IDE2 & INT1 ($r_s = .43, p < .01$)
- IDE2 & INT3 ($r_s = .39, p < .01$)
- EXT & INT2 ($r_s = .33, p < .05$)
- EXT & IDE1 ($r_s = .39, p < .01$)

A weak relationship was found between motivations to use the service because it was interesting or fun (intrinsic motivation) and good for oneself (identified regulation). Also, motivation to use the service based on external regulations was weakly related to pleasantness (intrinsic motivation) and doing the activity for one's own good (identified regulation).

Motivation to use Microsoft Math based on logged usage data. In addition to the subjective measurements, we calculated the number of separate completed quizzes in order to estimate how motivated the respondents were to use the service. On average, respondents had completed 70,96 separate quizzes, with SD 76,87. 10/53 (19%) respondents had completed less than 10 separate quizzes, while 14 (32%) had completed more than one hundred, the highest number being 329. Interestingly, 9/10 respondents who had completed less than 10 separate quizzes stated that they were very interested (Q6) in using the service for studying mathematics. However, log data suggests that these learners did not actually use the service that much. This stresses the importance of studying actual usage data in order to learn what users really do with mobile learning services.

4.3 Behavioral intentions

Behavioral intentions were measured with two questions (Q9 and Q10). Based on the results, the respondents were very likely to use the service in the future (Q9: $M = 4.77$, SD = 0.70) and also recommend the service to their friends who had access to mobile phone (Q10: $M = 4.87$, SD = 0.39).

4.4 Context

49/53 (93%) used the service more often during their free time than during school or working hours. Also, 45 (85%) preferred to study mathematics with the studied service when compared with a traditional textbook, pen and paper. Although schools may restrict the use of mobile phones [21], these results suggest that the service has been successful in motivating the learners towards informal learning. Also, the available service seems to be compelling from the learners' perspective when compared with traditional methods.

4.5 Correlations between measurements and comparisons of subgroups

Next, we will report 1) significant correlations between measurements of different factors and 2) significant differences between subgroups that were created based on background variables.

Correlations. Spearman's **c**orrelation showed the following statistically significant correlations with $r_s > .50$.

- Interest in studying mathematics in general (Q4) & IDE3 (Q8) ($r_s = .51$, $p < .001$).
- Hedonic quality beauty (Q3) & willingness to recommend the service (Q10) ($r_s = .52$, $p < .001$)

Only weak relationships were found between the following UX and motivational factors:

- Overall UX (Q1) & I like to use this service (Q6) ($r_s = .33$, $p < .05$)
- Hedonic quality beauty (Q3) & (IDE1 (Q8) ($r_s = .41$, $p < .005$) / IDE2 (Q8) ($r_s = .34$, $p < .05$))

Subgroups. For further analysis, we divided the respondents into the following subgroups: 1) female (22) and male (31), 2) novice (39) and expert users (14), and 3) math-talents (25) and non-talents (28). "Expert users" had completed 100 or more separate quizzes. "Math-talents" had rated Q14 and Q15 as 4 or 5. For each subgroup pair, Mann-Whitney U Test was used to detect any significant differences in the measurements.

No statistically significant differences between genders were found among any of the questionnaire measurements.

We found a statistically significant difference between novice and expert users in external regulation (Q8, $U = 175$, $p < .05$) suggesting that external rewards may have motivated learners who had completed more separate quizzes.

Statistically significant difference was found between math-talented and non-talented groups in SIMS item INT1: "This activity is interesting" (Q8, $U = 275$, $p < .05$), suggesting that those who evaluated themselves to be talented in math were also more interested in using the Microsoft Math service. Another notable finding was that math-talented were more likely to keep using the service in the future (Q9, $U = 275$, $p = .058$), however this difference was not statistically significant ($p > .05$).

5. DISCUSSION

The results suggest that learners were motivated in using Microsoft Math service and that their user experience was positive in terms of overall UX, ease of use and beauty. There was also a weak relationship between those who liked to study mathematics in general and use Microsoft Math. Majority also preferred the service to using textbooks.

Relationships between UX and motivational factors were not evident. Weak relationships were found between overall UX and "like to use the service" (motivation). The weak relationship between hedonic quality beauty and two identified regulation items seems accidental in this context.

Majority of the learners liked studying mathematics and this intrinsic motivation was also evident in open comments. However, the most often mentioned reason for studying mathematics related to future plans and ambitions (identified regulation). Previous studies have also shown that African youth generally appreciates education as a means to gainful employment or status in society [15]. The motivations for using Microsoft Math based on SIMS seem to be in line with the motivation to study mathematics in general, as the means for all intrinsic motivation and identified regulation items were high.

Extensive content that supported informal learning seemed to be that main motivation for using Microsoft Math. Other detected factors such as accessibility, supplementing teaching at school, ease of use, efficiency and inexpensive usage were also likely to support informal learning. Personal experiences of improved math skills or marks were considered highly motivating. We assume that shared stories of such experiences would also be an effective way of motivating fellow learners to try out the service.

Surprisingly, inexpensive use of the service (no airtime costs) was mentioned only by two respondents. Considering that majority of the users are young, one could expect this to be crucial aspect for using such service outside school hours. It is possible that this aspect was not considered as the service has been free to use since it was launched.

Playful elements such as competing with other learners by collecting points or measuring your progress motivated some respondents. Few mentioned the possibility of winning actual prizes. There was also statistically significant difference between novice and expert users in motivational factor external regulation. This could relate to the fact that some learners had completed significantly more quizzes than others, perhaps in order to win competitions or to see their name on the high-score board. Therefore, it appears that including such playful elements can fuel extrinsic motivation for some users in mobile mathematics learning context.

Regarding the negative comments on using Microsoft Math, one learner found the examples difficult to comprehend. However, more details would be required on learning what aspects were not explicit. Another practical challenge mentioned was slow loading times. This is likely to hinder positive user experience, as even if the service would be optimized for older mobile device models, data transfers over a mobile network could be slow at times or be prone to blackouts. Finally, answering user feedback too slowly is likely to give a bad impression to users.

Interestingly, a moderate relationship was found between the hedonic quality beauty and willingness to recommend Microsoft Math to friends. This seems unusual for mobile learning context and would be interesting to see if a similar relationship can be found with other mobile learning services.

Finally, it seemed that math-talented were more likely to keep using the service in the future. This raises a question of how to motivate those with weaker skills in math to keep using mobile learning services also in the future?

UX Goals. Based on the identified motivational factors to use the studied mobile math service, we propose the following UX goals to support designing mobile learning services for developing countries. The proposed goals *autonomy* and *competence* reflect the fundamental psychological needs discussed by Sheldon et al. (2001).

- *Autonomy* – to feel freedom by choosing when and how to study using the mobile learning service.
- *Competence* – to feel successful by achieving new goals in the learning process.
- *Efficiency* – to feel that no time is wasted in the use of the mobile learning service.

The *autonomy* goal could be supported with good accessibility and by enabling inexpensive usage. Content should support informal learning for example with clear examples, step-by-step instructions and theory reading.

The *competence* goal could be achieved for example by offering sufficient challenge in tasks and emphasizing the personal progress with playful elements such as collecting points, competing with fellow learners and rewarding from achievements.

The *efficiency* goal is supported by instant feedback from learning exercises, easy to use user interface and short loading times.

Implications for conducting remote mobile surveys. Based on our experiences in conducting this study, we present the following implications to support researchers conducting mobile surveys in developing countries:

- Open questions can work with for mobile surveys. We were positively surprised by the long responses to open questions. It seems that learners in South Africa were used to writing longer responses using mobile devices.
- Longer questionnaires can be divided into two parts to decrease respondent fatigue. However, we experienced that the learners could have answered the whole questionnaire in one session.
- Logged (objective) usage data can support subjective statements.
- Free airtime coupons for mobile data as rewards. Coupon codes can be easily sent to respondents (e.g. lottery winners) via SMS.

Limitations. Our study was limited at least in the following aspects: 1) sample size was small in comparison with the potential number of users for Microsoft Math at the time of conducting the survey, 2) our recruitment process may have attracted respondents who were already interested in maths and therefore the results should not be generalized in other contexts without consideration, 3) a majority of the respondents considered themselves to be skilled in math, suggesting that our results do not reflect the views of those learners who struggle with mathematics, 4) better measurement for math skills would have been test scores from national math tests. However, we did not have access to this data as the survey was kept anonymous. Alternatively, we could have asked the respondent for their most recent math score.

6. CONCLUSIONS

In this paper, we have studied the motivational and UX factors of Microsoft Math service used by learners in South Africa. Results from a mobile survey include the summaries of motivational factors for studying mathematics in general and with Microsoft Math service. However, no meaningful statistically significant relationships were identified between UX and motivational factors. Finally, we summarize our findings by reporting 1) UX goals to support design work of mobile learning services for developing countries and 2) implications for conducting remote mobile surveys for developing countries.

Future studies could focus on understanding motivational aspects in mobile learning with a more in-depth approach, such as interviews. Replicating this study with larger sample size including learners who are less skilled in mathematics, in other countries or cultures, or with other measured UX factors could be interesting. Studying the views of other actors, such as teachers or parents, could provide a more comprehensive image of the possibilities and challenges of mobile learning services in the future. For example, in an exploratory study taking place in the rural India, parents perceived that learning to use technology should be emphasized in education and that mothers could also benefit from educational applications that support lifelong learning [11]. However, cultural issues such as gender attitudes are still a major challenge that can hinder girls' opportunities for mobile learning (ibid.). Finally, longitudinal studies of motivational factors and UX with mobile learning services could help understanding how to keep learners motivated and how to support positive UX over time.

7. ACKNOWLEDGMENTS

We are grateful for Microsoft for their help in data collection. This study was part of the CCD MobiLe project, supported by Tekes – the Finnish Funding Agency for Innovation.

8. REFERENCES

[1] Deci, E. L. and Ryan, R. M. 1985. Intrinsic motivation and self-determination in human behavior. New York: Plenum.

[2] Finstad, K. 2010. The Usability Metric for User Experience. *Interacting with Computers*, 22, 5, 323–327. DOI=10.1016/j.intcom.2010.04.004

[3] Guay, F., Vallerand, R., and Blanchard, C. 2000. On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24, 3. Retrieved from http://link.springer.com/article/10.1023/A:1005614228250

[4] Hassenzahl, M. 2003. The Thing and I: Understanding the Relationship Between User and Product. In *Funology: From Usability to Enjoyment*, M. Blythe, C. Overbeeke, A.F. Monk and P.C. Wright, Ed. Hum. Com. Int. vol. 3, 31-42.

[5] Hassenzahl, M., Burmester, M., and Koller, F. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer* 2003: Interaktion in Bewegung, J. Ziegler and G. Szwillus, Ed. B.G. Teubner, Stuttgart, Leipzig, 187-196.

[6] ICT Development Associated Ltd. 2011. Transformation-Ready: The strategic application of information and communication technologies in Africa. Available at http://siteresources.worldbank.org/EXTINFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/282822-1346223280837/Education_Fullreport.pdf

[7] ISO 9241-210: 2010. "Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems," *International Standardization Organization (ISO). Switzerland*.

[8] JIPSA Annual Report. 2008. Joint Initiative on Priority Skills Acquisition. Available at http://www.thepresidency.gov.za/docs/reports/jipsa/annualreport08.pdf

[9] Jones, A., Issroff, K., Scanlon, E., Clough, G., McAndrew, P., and Blake, C. 2006. Using mobile devices for learning in informal settings: is it motivating? In *IADIS International Conference on Mobile Learning*, 14-16 July 2006, Dublin, IADIS Press, 251-255.

[10] Kaasinen, E., Roto, V., Hakulinen, J., Heimonen, T., Jokinen, J., P., P., Karvonen, H., Keskinen, T., Koskinen, H., Lu, Y., Saariluoma, P., Tokkonen, H., and Turunen, M. 2015. Defining User Experience Goals to Guide the Design of Industrial Systems. *Behaviour & Information Technology journal*, Taylor & Francis. DOI=10.1080/0144929X.2015.1035335

[11] Kumar, A., Tewari, A., Shroff, G., Chittamuru, D., Kam, M., and Canny, J. 2010. An exploratory study of unsupervised mobile learning in rural India. *Proceedings of the 28th International Conference on Human Factors in Computing Systems CHI 10*, 743-752. DOI=10.1145/1753326.1753435

[12] Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S., and Kort, J. 2009. Understanding, scoping and defining user experience: a survey approach. In *Proc. CHI 09*, ACM, NY, 719-728.

[13] Lu, Y., and Roto, V. 2015. Evoking meaningful experiences at work – a positive design framework for work tools. *Journal of Engineering Design*. DOI=10.1080/09544828.2015.1041461

[14] Mahlke, S., and Thüring, M. 2007. Studying antecedents of emotional experiences in interactive contexts. In *Proc. CHI 07*, ACM, NY, 915-918.

[15] Masita-Mwangi, M., Mwakaba, N. M., Ronoh-Boreh, F., and Impio, J. (2012). Building a Case For M-learning in Africa: African Youth Perspectives on Education. In *Proc CHI 12 Case Study*, ACM, NY, 521-536.

[16] Sharples, M. 2006. Big Issues in Mobile Learning. *Report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative*. 2006. <hal-00190254>

[17] Sheldon, K. M., Elliot, A. J., Kim, Y., and Kasser, T. 2001. What Is Satisfying About Satisfying Events ? Testing 10 Candidate Psychological Needs. *Journal of Personality and Social Psychology* 80(2), 325-339. DOI=10.1037//O022-3514.80.2.325

[18] MoLeNET programme: http://web.archive.org/web/20100830073550/http://www.molenet.org.uk/

[19] Sharples, M., Taylor, J., Vavoula, G. 2006. A Theory of Learning for the Mobile Age. Andrews, R. and Haythornthwaite, C., Ed. *The Sage Handbook of Elearning Research*, Sage publications, 221-247. <hal-00190276>

[20] TIMMS. 2003. *International Mathematics Report*. Available at http://timss.bc.edu/timss2003i/mathD.html

[21] Vainio, T., Walsh, T. and Varsaluoma, J. 2015. CROSS-CULTURAL DESIGN OF MOBILE MATHEMATICS LEARNING SERVICE FOR SOUTH AFRICAN SCHOOLS. *IADIS International Journal on WWW/Internet*. 01/2015; 12(1):81-93. ISSN: 1645-7641, http://www.iadisportal.org/ijwi/papers/2014121106.pdf

[22] Varsaluoma, J., Väätäjä, H. Kaasinen, E., Karvonen, H., and Lu, Y. 2015. The Fuzzy Front End of Experience Design: Eliciting and Communicating Experience Goals. In *Proceeding of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction OzCHI '15*. ACM, New York, NY, 324-332. DOI= http://dx.doi.org/10.1145/2838739.2838761

[23] Venkatesh, V. 2000. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research* 11, 4 (Dec. 2000), 342-365.

[24] Väätäjä, H., Savioja, P., Roto, V., Olsson T., and Varsaluoma, J. 2015. User experience goals as a guiding light in design and development – Early findings. In *INTERACT 2015 Adjunct Proceedings*. Univ. of Bamberg Press, 521-527.

[25] World Bank. 2012. Information and Communications for Development 2012: Maximizing Mobile. Washington, DC: World Bank. DOI: 10.1596/978-0-8213-8991-1; website: http://www.worldbank.org/ict/IC4D2012. License: Creative Commons Attribution CC BY 3

P3

USEFULNESS OF LONG-TERM USER EXPERIENCE EVALUATION TO PRODUCT DEVELOPMENT: PRACTITIONERS' VIEWS FROM THREE CASE STUDIES

by

Varsaluoma, J. and Sahar, F. 2014.

In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI'14). ACM, New York. pp. 79-88.

Usefulness of Long-Term User Experience Evaluation to Product Development: Practitioners' Views from Three Case Studies

Jari Varsaluoma
Tampere University of Technology
Korkeakoulunkatu 6, P.O. Box 589, 33101
Tampere, Finland
jari.varsaluoma@tut.fi
+358 40 8490856

Farrukh Sahar
Tampere University of Technology
Korkeakoulunkatu 6, P.O. Box 589, 33101
Tampere, Finland
farrukh.sahar@tut.fi

ABSTRACT
Understanding the temporal aspects of user experience (UX) has received increasing attention in the HCI community. However, little empirical evidence is available on how practitioners in product development companies evaluate the usefulness or actually use long-term UX evaluation data in their work. In this study, we explore how practitioners (e.g., managers, designers and UX specialists) evaluate the usefulness of long-term UX evaluation results to their own work. Three case studies were conducted with longitudinal and retrospective methods in a company developing interactive sports products. Our findings suggest that long-term UX evaluation provides results that are perceived as interesting, relevant and useful by practitioners. Potential uses for the results were e.g., verifying practitioners' expectations, planning future work, understanding changes in UX, the development of future products, and updating current software products. Future research should focus on how to provide long-term UX evaluation results in more efficient manner to benefit product development.

Author Keywords
Usefulness; evaluation; long-term; longitudinal; user experience; usability; product development; case study.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Today, several companies developing interactive products have adopted user studies as a regular part of their product development processes. Traditional user research methods often focus on the first experiences and learnability problems that novice users have with interactive products. However, previous research suggests that conventional usability testing methods may not reveal the problems that can cause frustration for more experienced users over time [11]. Indeed, there has been an increasing interest in HCI field towards the temporal aspects of usability and user experience (UX) [4, 7, 15, 16].

There are no exact definitions for terms *long-term UX* and *longitudinal research* in HCI literature. However, several of the proposed UX models consider the temporal aspects of UX [e.g. 10]. Also, an emerging definition states that longitudinal research looks beyond the initial UX (or learning experience) [4]. Longitudinal research "is ideal for studying how and when users transition from novice to expert, as well as addressing issues such as abandonment or adoption rates, learnability, comfort with technology, productivity, and evolution of user perceptions" [4]. In this paper, *long-term UX evaluation* refers to longitudinal and retrospective studies that focus on understanding the change in product UX over time.

Motivation and benefits to conduct long-term studies have been addressed by the HCI research community [4, 7, 15, 16]. However, there is lack of empirical research on how practitioners in companies utilize results from long-term UX evaluations in their work and how *useful* this information is from practitioners' perspective. We argue that providing useful UX evaluation results that can support product development is a key factor in motivating stakeholders to invest in conducting long-term studies in future.

In this paper, we explore the *usefulness of long-term UX evaluation results for practical work* over three case studies in one company. The questions that motivated this research were:

- What kind of long-term UX evaluation results are the most useful for practitioners (e.g., managers, designers, UX specialists), who participate in the development of interactive products?

NordiCHI '14, October 26 - 30 2014, Helsinki, Finland
Copyright 2014 ACM 978-1-4503-2542-4/14/10…$15.00.
http://dx.doi.org/10.1145/2639189.2641218

- For what purposes are the reported long-term UX evaluation results seen as useful?
- How do practitioners actually use the long-term UX evaluation results in their own work?

During the case studies presented in this paper, information was collected on users' experiences with products and how UX relates to other aspects, such as customer loyalty. Long-term studies can result in a vast amount of information that can be beneficial to practitioners in different positions, e.g. management, marketing and design. Therefore, we were interested to explore how managers, in addition to UX specialists and designers, would use the reported long-term UX evaluation results. Due the exploratory nature of this study, we let the practitioners freely describe, what use (if any) they had for the reported results.

By presenting new empirical research results, this study can help building the body of knowledge for long-term research in HCI. The issues highlighted in this paper can contribute to the ongoing discussions and motivate the future research of long-term UX evaluation practices for both industry and academia.

First, we present an overview of the current long-term UX research practice in HCI, followed by discussion on measuring the usefulness of long-term UX evaluation results. The research process chapter presents the three case studies and the personnel surveys for measuring the usefulness of the long-term UX evaluation results. The results chapter describes the findings from the personnel surveys. In discussion, the main findings are reviewed and their meaning discussed. Finally, the research is summarized in the conclusions, with the limitations of the study and implications of the findings for future research.

BACKGROUND

Long-Term UX Research Practice in HCI

In [17], three perspectives for HCI studies were presented based on the time period the study covers. Typical usability tests are a *micro* perspective studies (one to two hours), while longer-term studies are divided into a *meso* perspective (e.g., 5 weeks) and *macro* perspective studies (from years to the whole product lifecycle). While macro perspective studies are rare in HCI, the number of published meso perspective studies has been increasing since 2006, judging by the number of workshops and other events around the topic [7].

Long-term studies are not dependent on any specific method, and using a combination of quantitative and qualitative methods is encouraged [4]. Retrospective methods such as CORPUS [17], iScale [6] and UX Curve [9] can be cost-effective alternatives to repeated measurement methods, such as the Experience-Sampling Method (ESM) [2] and the Day Reconstruction Method (DRM) [5]. Retrospective studies rely on users' memories of experiences and are prone to biases. However, memories can guide customers' future behavior and what experiences will be reported to others [12]. Therefore memories of product use can be relevant information for product development purposes. Lastly, data logging methods (e.g., usage logs) provide an interesting viewpoint for observing changes in product use over time [4, 7].

Longitudinal studies are useful for studying change over time, as they include two or more observations or measurements with the same users [17]. However, in quantitative longitudinal studies a minimum of three measurements is advised to differentiate true change from measurement error [14]. The length of previous longitudinal studies in HCI varies from few weeks to three years [4]. In order to track the change over time, some of the dimensions (e.g., tasks, users, measures, or products) have to stay constant over the study period. As the same participants use the studied product over time, they will get more experienced with the product. Therefore, no longitudinal survey samples the exactly same users twice [13]. This should be considered especially when the learning process itself is of interest, e.g. how long it takes and why for new users to learn to use a product efficiently?

How to decide the timing and frequency for measuring long-term UX? Considering longitudinal studies in general, if no theoretical guidance is available for deciding the measurement times, Ployhart and Ward [14] propose to 1) consider "natural" measurement occasions for the studied phenomenon, 2) conduct interviews or observations with subject matter experts, and 3) review literature that studied similar phenomena. Few studies in HCI literature discuss the most beneficial measurement times with interactive products. In a longitudinal study by Kujala and Miron-Shatz [8], DRM [5] and questionnaires were used to study 22 users' experiences with new mobile phone models. After the first week with DRM, more retrospective measurements were conducted on the 6th day, after 2.5 months and after 5 months of product usage. Surprisingly, some basic usability problems were reported still after 2.5 months. Another study [13] used a cross-sectional approach to study differences between novice and expert users regarding frustration episodes. Although the sample size was small, results suggest that studies where applications are used beyond a year may not be beneficial, as the most observable differences occur within three to six months from the beginning of use.

In practice, the number and times of UX measurements can depend on several factors, including the product itself (e.g. use frequency, the estimated length of the product learning period and product life cycle), users' characteristics (e.g. previous experience with similar products), available research resources, measured factors, and stakeholders' demand for receiving actionable results. Finding a balance between the length of a single survey and the number of measurements is important, since each measurement requires effort from the participants and participant drop-out is common for longitudinal studies [4, 14]. Overall, it seems that more empirical research is required as the work towards building a rich body of knowledge for long-term UX research in HCI continues.

Usefulness of Long-Term UX Evaluation Results to Product Development

As user research, be it long-term or short-term, is conducted in a product development company, the probable goal is to provide useful information to be used in specific phases of the product development process. To our knowledge, measuring the *usefulness of long-term UX evaluation results to work practice* has received little attention in HCI literature. Usefulness has been measured before regarding the use of different HCI methods. In [1], the perceived usefulness of different HCI methods by HCI practitioners was measured using a pen-and-paper questionnaire and a web survey. Participants rated the usefulness of the provided HCI methods for different phases of the development process (start, mid, and end phase) using a rating scale of 1 (Not at all useful) to 5 (Very useful). In addition, the participants were asked what methods they had actually used in the different product development phases.

In the current study, the evaluation of the *usefulness* of the long-term evaluation results was supported with additional measurements that we considered meaningful. These related factors included: 1) what is considered *interesting* in the results, 2) what is *relevant* (similar to importance) in the results for each practitioner's work, 3) the *novelty value* of the results, 4) *likeability to utilize* the results, and 5) the *actual use* of the results in practice. Our hypothesis was that information rated as interesting, novel or relevant to the practitioners' own work would also have more potential of being useful. However, it is possible that information that is considered e.g., relevant to one's work, can be considered uninteresting, or vice versa. Furthermore, although the reported likelihood to utilize the results in future might relate to the usefulness of the reported information, an observation or measurement of actual use of the results is required to properly evaluate their usefulness.

METHOD

Between the years 2011 and 2013, three case studies evaluating long-term UX of products were conducted with one Scandinavian company developing interactive digital sports equipment. The studies were a part of a joint research project between a university and the company.

The focus of this paper is in how the practitioners in the company evaluated the usefulness of the long-term UX evaluation results. Detailed results of the case studies are not in the scope of this paper and therefore only the type of the reported results is presented. Next, we describe the case studies (briefly), the personnel surveys, and how they were conducted.

Case studies

Table 1 summarizes the case studies and their research methods. Web surveys were mainly used, since all the studies were international. Both qualitative and quantitative questions were used to collect data on users' experiences with the products. All the participants were contacted via the company's customer database and were chosen based on a screening survey. One of the authors participated closely in the design of the studies. In total there were five sessions (DC and SWb were reported in two parts) during the three case studies where the results were presented to the company personnel.

Case Study 1: Diving Computer (DC)

The first case study DC evaluated the UX of a diving computer and its associated software after the first months of usage. Another objective was to study how a new software update would affect the UX. Furthermore, the study acted as a pilot for using the iScale tool [6] in a remote study.

Two retrospective measurements were carried out using a web survey with the Attrakdiff questionnaire and iScale. Attrakdiff provides quantitative data describing user's perceptions towards the evaluated product [3]. 33 users, all male, answered the first survey (part 1). The time of product usage varied between the participants from one month to five months. 21 of the participants continued to the second survey after a software update was released. The second survey was sent to the participants after each of them had used the product for six months.

Both results presentations (DC part 1 & 2) of the study included: 1) customer background information e.g., previous experiences from the brand and similar products, 2) expectations from the product before purchase (in retrospect) and how the expectations were met, 3) the

Case study	Studied product	Number of respondents	Measured product usage period	UX evaluation methods
1. DC	Diving computer	Part 1: 33 Part 2: 21	Part 1: Varied from 1 to 5 months Part 2: 6 months	Retrospective: Web survey with Attrakdiff [3] + iScale [6]
2. SWa	Sports watch A	25	From 3 to 6 months	Longitudinal/Retrospective: Weekly web survey
3. SWb	Sports watch B	Part 1: 111 Part 2: 104	Part 1: From 1-2 to 4-5 months Part 2: From 1-2 to 7-8 months	Longitudinal/Retrospective: Monthly web survey with Attrakdiff [3]

Table 1. Case study summary. Results from the case studies DC and SWb were reported to the company in two parts.

attractiveness of the diving computer and its associated software, 4) satisfaction with the product and why, 5) the importance of the product for oneself and why, 6) willingness to recommend the product to friends and why, 7) an abridged version of Attrakdiff, and 8) summary of the iScale curve shapes with positive and negative experiences.

Case Study 2: Sports Watch A (SWa)

The goal of the second study SWa was to understand how the UX of different product components associated with a sports watch can affect the evaluation of the product's overall UX. Furthermore, changes in the UX after the initial learning period were studied. The studied product consisted of the sports watch as the main unit, two sensor units and a web service.

The study was carried out as a weekly web survey, following a repeated measurement design. 25 participants (4 female) were chosen, each with over two months of use experience with the product. The data collection phase lasted for two months and included eight weekly surveys per participant.

The results presentation included: 1) customer background information, 2) the number of different sports activities where the product had been used, 3) overall positive and negative feelings with the product over time 4) willingness to recommend the product to a friend over time, 5) the number of reported positive and negative experiences over time for each product component, 6) the summaries of positive and negative experiences for each product component, 7) experience quotes from users, and 8) component-specific design ideas based on the experiences.

Case Study 3: Sports Watch B (SWb)

The studied product in the third case study SWb consisted of a sports watch, a sensor unit, installable software, and a web service. The research goal was twofold: first, to understand the customer journey since the beginning of use, including users' expectations and their fulfilment. Second, to learn how the UX changes over time and what factors affect these changes (e.g. software updates).

The study consisted of six monthly web surveys. The final report included results from 104 participants (7 female). The study covered the experiences with the product from the first and second month until the eighth and ninth month of usage, depending on the date of the product purchase.

A preliminary report (SWb part 1) was created to present the main results from the first three surveys (with still 111 participants), including: 1) customers' background information, e.g., previous experiences with similar products, relationship with the brand, 2) expectations before the product purchase (in retrospect), 3) how easy it was to take the product into use and need for support, 4) challenges when starting product usage, 5) product use frequency over the first three months, 6) satisfaction with each product component over time, 7) willingness to recommend the brand to a friend over time and why, and 8) summary of positive and negative experiences that the users reported with the product.

The final report (SWb part 2) included: 1) customers' background information, 2) how easy it was to take the product into use and need for support, 3) product use frequency over six months, 4) satisfaction with each product component over time, 5) willingness to recommend the brand to a friend over time and why, 6) Attrakdiff measurements over time, 7) the most important product qualities over time (based on Attrakdiff), 8) expectations before product purchase (in retrospect), 9) how expectations were fulfilled, 10) quantitative analysis results, e.g., how emotions relate to the willingness to recommend, satisfaction with the product, and Attrakdiff measurements, 11) summary of positive and negative experiences that the users reported with the product over time, 12) new design ideas based on the users' suggestions, 13) summary of the experiences that users reported on the 6th (last) survey in more detail, and 14) the conclusions of the study, including the first steps in the customer journey, how expectations were fulfilled, and changes in UX over time.

Personnel Surveys

Procedure

Before each of the five results presentation sessions, a contact person from the company informed stakeholders about the upcoming presentation. Project team members presented the results in a live meeting with a Microsoft PowerPoint. Paper surveys were used to gather feedback from the session participants, who were free to answer the survey during or after the presentation. The presentation slides and case study data files were delivered to the company contact person after each presentation.

In order to verify if practitioners had actually used the research results from the case study SWb in their work, a follow-up survey was administered in 2014, nine months after the last SWb results reporting session. During this time, the company had launched a new model of the sports watch that had been studied in the case study SWb. An email invitation to answer a web survey was sent to the participants who had attended either of the case study SWb results sessions. Reports from the case study SWb had been available in the company's intranet for the last nine months. Participants were asked to skim through the reports before answering the survey.

Survey questions

Table 2 presents the questions for each of the personnel surveys. New questions were added in SWa and SWb part 1. SWb part 2 contained more results than any of the previous presentations and we decided to modify the survey to follow the structure of the presentation. For each results section in the presentation there was a separate page in the survey, with three new questions: 1) "How useful these results are for you? (1=Not at all useful, 5=Very useful)", 2) "What is the novelty value of these results for you? (1=No novelty value, 5=High novelty value)", and 3) "Any comments, feedback or thoughts from these specific results?" Each page included pictures of the presentation slides as a memory aid. Since the usefulness was asked separately for each results section, the question about relevant information was excluded. However, the question about "what was interesting in the results" was kept as a summary question (see Table 2).

In the follow-up survey, the participants were asked if they had read or used either of the case study SWb reports in their work. If the report had not been used, the participant was asked why not. Otherwise, the participant was asked that what kind of information he or she had been looking for in the report and for what purpose. Furthermore, the respondent was asked how useful the information in the reports had been in the participant's own work, on a scale 1 to 7, where 1 = not at all useful and 7 = very useful. The same scale was used at the end of the survey to ask how useful the long-term user studies are in general from the participant's point of view, continued with an open-ended question: "Why? Please clarify your answer".

Participants

30 individuals from the company answered at least in one of the five personnel surveys (52 responses in total). 20 (67%) answered only in a single survey. Two participants answered in all the five surveys. Figure 1 presents a summary of the personnel survey participants. Each response was categorized into one of the four categories based on participant's title and work tasks. The categorization was done in cooperation with two employees from the company. "Manager, high-level" category included titles such as Business Unit Director, Design Manager and Program Manager. "Manager" category included e.g., Product Manager, Product Concept Manager and Team Manager. "Designer/UX Specialist" category included e.g., UI Designer, Interaction Designer and UX specialist.

Unfortunately, despite two reminders over one month, only five responses were received to the follow-up survey: two "high-level managers", two "managers" and one "other".

Analysis

The responses to the open-ended questions were content analyzed by one of the authors. If a single answer entailed several different aspects, each one was coded as an individual item. Similar items between responses were linked into appropriate categories that were named to describe the items in them. Data for each question from each survey were first analyzed separately. After this, similar categories from all five surveys were combined and the categorization descriptions updated as necessary. If the same participant had similar responses to the same question in different surveys, items from each response were added up as separate items for their categories. Due the small number of participants, no statistical tests were conducted to compare the follow-up survey results with the previous surveys.

Question	DC part 1 & 2	SWa	SWb part 1	SWb part 2
"What kind of user information would be the most beneficial for you?"			x	x
"Was there something especially interesting in these results? What?"	x	x	x	x
"Was there something relevant to your own work in these results? In what way?"	x	x	x	
"Where would you use the relevant results? (e.g., specific phases of product development)"	x	x	x	
"How likely will you utilize the presented results in your own work? (Not at all likely 1-7 Very likely)"	x	x	x	x
"What was missing in the results or would have been more useful for you?"		x	x	x

Table 2. The questions (five open and one Likert-scale) used in personnel surveys during the case studies.

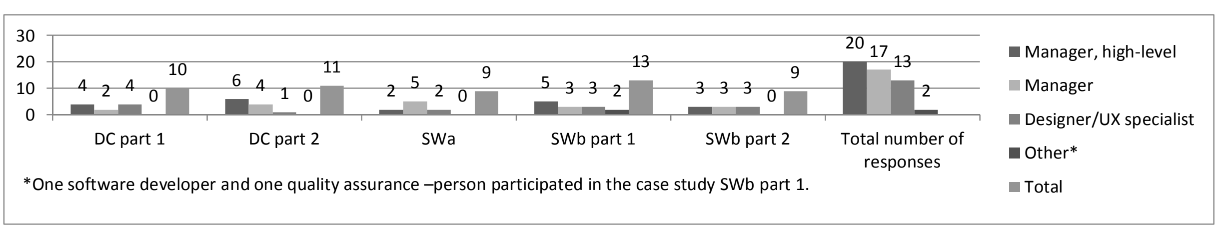

Figure 1. Personnel categorization and the number of responses to the personnel surveys in the case studies. In total, 52 responses were collected from 30 separate participants.

RESULTS

1) What kind of information was considered interesting in the long-term UX evaluation results? Figure 2 presents a summary of what the personnel found interesting in the long-term study results. After summarizing all the five surveys, there were in total 14 different categories with 70 items related to what was considered interesting. Nine items were unique. From 52 possible responses, on seven occasions (14%) a participant left this question unanswered.

The three largest categories with 8 responses (15%) were related to: 1) comparing the results with the participant's own expectations, 2) the positive and negative experiences that users reported, and 3) how UX changed (or did not change) over the measurement period. User satisfaction with the product, UX of specific components, and factors that affected UX were also found as interesting topics (5 responses each).

It seems that practitioners had initial expectations for the results based on their subjective knowledge: "Expected results. Good that initial 'hunch' of UX predicted results were aligned with actual results" (ID8, Designer/UX), or previous research: "Very consistent with previous research findings and our subjective understanding about the topic" (ID15, Manager, high-level). Two high-level managers commented, on different case studies, that most of the findings were already known to them: "Interesting, yes, but not much new info, most of this is already known by us." (ID25).

2) What kind of information in the long-term UX evaluation results was considered relevant to the practitioners' own work? Figure 3 summarizes what the participants found the most relevant for their own work in the reported results. The analysis resulted in 12 categories with 56 items (11 unique). This question was not included in the fifth personnel survey (SWb part 2). From the 43 possible responses, two were blank (5%). The same participants had left the previous question about interesting results blank.

The most repeated responses (8 in both categories, 19%) were: 1) all the results were relevant and 2) the results show where to focus next, e.g., improving a specific product component/feature or promoting specific aspects of the product in future. Furthermore, the positive and negative experiences with the product and user feedback in general were found relevant (7, 16%). It was noted that a single participant gave similar comments in three different surveys for categories "Taking the product into use" (ID8, Designer/UX) and "Long-term UX" (ID2, Manager).

Two participants did not seem to make any distinction between what was interesting or relevant content in case the SWa presentation, as they answered the later question: "things mentioned above" (ID16, Manager, high-level) and "see previous" (ID17, Manager).

3) Where the information that was considered relevant could be used? Figure 4 illustrates where the participants could have used the research results from the first four presentations. The qualitative analysis resulted in 12 categories with 52 items, from which eight were unique. Two participants did not respond to this question.

Majority of the responses (12, 28%) related to proposing, concepting and defining new products: "Good points for defining product specifications and requirements before actual development project" (ID20, Manager). Two of the second largest categories with seven responses (14%) indicated that the findings could be used e.g., in updating the current product, and as an input to future product development to avoid some of the reported problems. For example, one designer commented: "We can use these results because software development is still going on daily" (DC part 1, ID10). For case SWb part 1, there were comments related to current software and future products: "Considering the content of the next software releases…" (ID26, Manager, high-level), and "Found problems will be very likely to be fixed in future products" (ID23, Manager, high-level).

4) What kind of information was considered novel and/or useful in the results of case study SWb part 2? Participants' evaluation of the novelty and usefulness of the result section are provided in Table 3. The most useful results seem to have been the satisfaction scores, the detailed experiences with the product and the conclusions of the study. All the results, apart from the quantitative analysis, were rated above average in usefulness, with mean 3.9 or higher (scale being from 1 to 5). Although the least useful, the quantitative results were seen as the most novel results. It seems that the presentation time and content were not enough to communicate the quantitative analysis results to the audience properly: "This part was a bit difficult to comprehend and would have needed a little bit more practical explanation. Still, the content was interesting", (ID9, Designer/UX).

Attrakdiff results were seen as the least novel, but the differences with other result sections were small. In overall, the novelty of the results was rated slightly above average. However, the standard deviations of the novelty values were slightly higher than those in the usefulness scores. This indicates that there were more differences among the participants in what was considered as novel information when compared with what was seen as useful information.

5) How likely the participants were going to use the results in their own work? Table 4 presents the mean values and standard deviations for each survey. There were seven missing responses, six of them for the case study DC part 2. All the mean scores were above average, suggesting that majority of the participants were planning to use some of the results in their work. Case SWa had the lowest score and the highest standard deviation, indicating that some of the participants did not see clear usage for the research results. One participant who gave the lowest rating (3) also commented that the product "is not in the core of my responsibilities" (ID17, Manager). Other low score (4) came from a manager (ID20), who stated that they had received similar feedback from other sources and that "most of the detailed findings being taken into account already".

6) Was there some information missing in the results? After the analysis of 25 responses it seems that the practitioners had been looking for more quotes from users and raw data in the case study SWa presentation (7 responses). Also, the most important usability problems and more insight of problem severity were missed by four participants for case SWb part 1. Furthermore, three participants asked in case SWa, that how many separate users had reported similar comments, since this had not been evident in the presentation. The rest of the comments/questions were unique and related to e.g., comparing different user groups, recommendations on how to improve/maintain UX over time, how to increase the recommendation rate, top 5 positive feedback of the UI, and more detailed data about the product components that received negative feedback. Interestingly, one manager (ID25) noted that while there

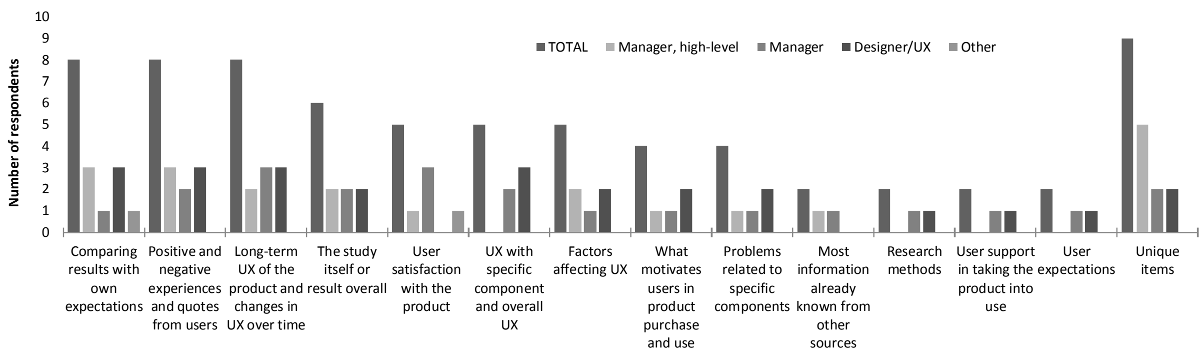

Figure 2. Responses to the question "Was there something especially interesting in these results? What?" Asked in all the five personnel surveys (n=52).

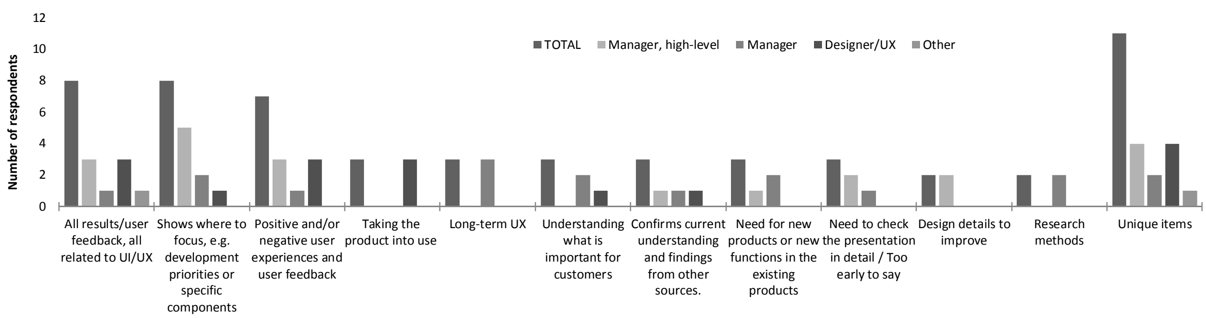

Figure 3. Responses to the question "Was there something relevant to your own work in these results? In what way?" Asked in the first four personnel surveys (n=43).

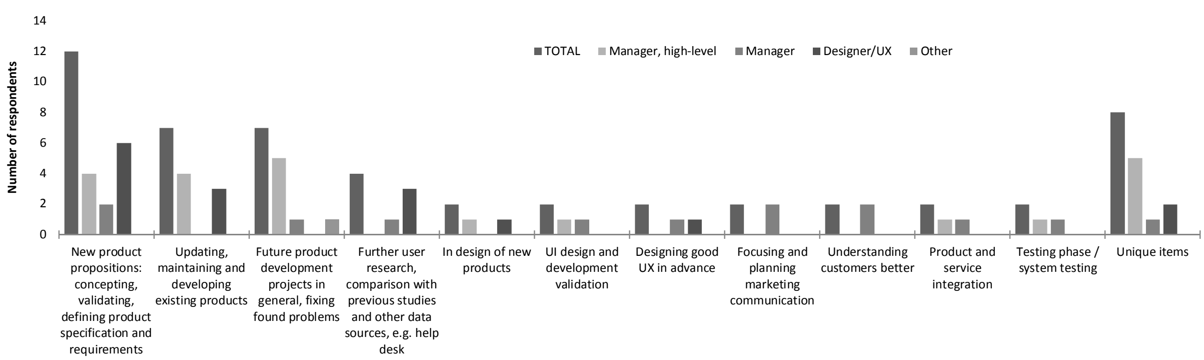

Figure 4. Responses to the question "Where would you use the relevant results?" Asked in the first four personnel surveys (n=43).

Result presentation sections for case study SWb part 2	Usefulness Mean (SD)	Novelty Mean (SD)
1. Attrakdiff measurements over time and the most important product qualities	3.9 (0.6)	**3.1** (0.8)
2. Satisfaction with each product component over time	**4.1** (0.6)	3.2 (1.2)
3. Willingness to recommend the brand to a friend over time and why	4.0 (0.7)	3.2 (1.1)
4. Expectations before the product purchase (in retrospect) and how they were fulfilled	4.0 (0.5)	3.4 (0.7)*
5. Quantitative analysis results, e.g., how emotions relate to the willingness to recommend, satisfaction with the product, and Attrakdiff measurements.	**2.9** (1.1)	**4.0** (1.0)
6. Summary of positive and negative experiences that users reported over time and design ideas based on the users' suggestions	3.9 (0.9)	3.3 (1.0)
7. A detailed summary of the experiences that users reported in the 6th (last) survey	**4.1** (0.6)	3.4 (1.1)
8. Conclusions of the study, including the first steps in the customer journey, fulfillment of expectations, and changes in UX over time	**4.1** (0.8)	3.3 (1.0)

* One answer missing, therefore n=8 for this question.

Table 3. Responses to the questions "How useful these results are for you? (1=Not at all useful, 5=Very useful)" and "What is the novelty value of these results for you? (1=No novelty value, 5=High novelty value)" Asked in case study SWb part 2 (n=9).

	DC part 1	DC part 2	SWa	SWb part 1	SWb part 2	Total
Mean (Std Dev)	6.60 (0.52)	5.83 (0.75)	**4.89** (1.45)	6.08 (1.04)	**6.63** (0.52)	6.02 (1.12)

Table 4. Responses to the question "How likely will you utilize the presented results in your own work? (Not at all likely 1-7 Very likely)". Asked in all the five personnel surveys (45 responses, after seven missing values).

was little new information in the results of case study SWb part 1, the results may also be a bit "outdated" due the rapid development cycle nowadays.

7) What kind of information from the case study SWb results did the practitioners use in their work during the next nine months? Two high-level managers, two managers and one quality assurance person answered the follow-up survey. One manager and the quality assurance person had not used either of the case study SWb reports, while the other respondents had read and/or utilized both of the reports from one to three times. The manager's (ID20) reasons for not reading the reports or using the information were that 1) there has not been a project where to use the information, and 2) lack of time. The quality assurance person (ID21), who had not participated in the part 2 presentation, replied that "I was not aware / I forgot that this document even existed" (SWb part 2) and that the presentation already gave the needed information (SWb part 1).

Regarding both reports of the case study SWb, three of the participants had been looking for information about: 1) "consumer long-term usage" and "using the findings for future work" (ID2, Manager), 2) "who buys the product and why", "who are our users and what they are experiencing", and "what kind of products we should make" (ID3, Manager, high-level), and 3) "enhancement ideas", "feature priorities", "usability pros and cons", and "as motivational feedback to development team to help them understand how important different UX aspects are" (ID6, Manager, high-level).

8) How useful the results of the case study SWb had been (after nine months) and how useful long-term user studies are seen in general? On a scale 1 (not at all useful) to 7 (very useful), the mean rating for the usefulness of part 1 results was 5 (SD 1.2) and for part 2 results 4.8 (SD 1.3). One manager (ID2) gave the rating 7 while other four participants rated the usefulness to 4 and 5. When looking at the previous surveys, all the participants, apart from the quality assurance person, had been very likely to utilize the results in their own work, as they gave the rating 6 or 7. Only one participant (ID20) who gave high ratings in the likeliness to utilize the results (6 and 7 in both SWb surveys) had not used them in his or her own work. The same participant also stated that while the information was important and useful, they receive "quite a lot of feedback continuously, that are often around the same topics as the study."

The mean rating was 5 (SD 1.4) for the question: "In general, how useful do you see the information from long-term user studies for your own work?" (scale 1-7). The answers were nearly identical to the ratings of the usefulness of the case SWb research results. Some of the reasons why long-term user studies were seen useful, were: 1) "they give us insights on a longer term usage of our products which would be difficult for us to do internally at this level" (ID2), 2) "to see effect on software update" (ID6), and 3) "to learn how experience changes with learning and after 'honey moon'" (ID6).

DISCUSSION

Our results indicate that a majority of practitioners, both managers and designers/UX specialists, found the long-term UX evaluation results interesting and relevant for their work. Also, the mean ratings for usefulness and likeliness to utilize the results were high. However, the number of separate categories and unique items in what was considered interesting and relevant suggests that in order to serve the needs of different practitioners in the company, the long-term studies should be versatile in what they measure. Since long-term product evaluations may require a substantial amount of time and resources, the early involvement of stakeholders and careful scoping of the long-term study is recommended [4].

It is interesting to note that while nearly all the results in case study SWb part 2 were considered highly useful, their mean novelty values were lower. A possible explanation for this could be stated in the open comments: similar findings had been received from other sources, and the results were mostly in line with the practitioners' own expectations or current understanding about the topic. This notion underlines one of the challenges with long-term studies: receiving the research results can take too long for them be as beneficial as they could be. As one manager (ID25) commented, the results can be too "outdated" for today's rapid development cycles. Also the fact that more practitioners came to listen to the case SWb preliminary report (part 1) than the final report suggests that the earlier results have more value for practice. One proposed solution is that the ongoing results are published periodically during the study [4]. Preliminary reports could be provided only of the measurements that are fast to analyze and sought after by stakeholders, therefore having better changes to still influence the design and development of the next product version. Alternatively, more systematic utilization of the other data sources that may provide similar information, such as customer care, could be developed.

All the products evaluated in the case studies were already available in the market. Therefore, it is no surprise that the results were mainly planned to be used in proposing, concepting and developing future products. However, for software products, the long-term results were still relevant as they could be used in the upcoming software updates. This raises an interesting question regarding the long-term UX of software: as software updates may alter the product (e.g., user interface), how does this affect the UX over time and how comparable the measurements are as one more dimension (the product, in addition to the learning user) changes? Although software updates add complexity to the evaluation as users may update their products at different times, the feedback regarding updates could be of major importance for stakeholders. Therefore, when planning long-term UX evaluation of software products, the estimated dates for update releases can be beneficial measuring points.

Due the small number of responses to the follow-up survey, it is difficult to make conclusions of the actual use of long-term results. However, it seems that lack of time and opportunities to utilize the information, being content with learning about the results in the first place, or simply not being aware of the available information can be reasons for not utilizing the results. Furthermore, the perceived usefulness of the results seems to decrease over time as three out of five participants rated the usefulness lower in the follow-up survey. Still, these results are not generalizable as the sample size was small. Also, no designers/UX specialists participated in the follow-up survey. This highlights the challenge of high drop-out rates in longitudinal studies, especially in industrial setting, where employees change or even the company can change its owner in the middle of the study [14].

The reasons why long-term studies were seen useful in general seem to echo some of the previously discussed findings from this study. Long-term studies can help understanding how the UX changes over time through learning and how software updates affect the UX. Also, the insight of longer-term usage that the six-month SWb study offered was something that would be challenging to achieve with internal resources. This hints that studies of this extent are not common in the company involved. However, if similar information is available through other feedback channels, even less systematically collected (e.g. via customer care), the perceived usefulness of long-term UX evaluation results seems to diminish. Still, in our case the results of carefully planned long-term studies seemed to have value for practitioners by confirming their own expectations and subjective understanding of the topic.

Apart from the products studied in this paper, the development of other product types might benefit even more from understanding how, when and why UX changes over time. Possible examples could be: 1) practitioners in an online gaming company are interested to know how and why the motivation to play their games changes over time, 2) designers (and customers) want to measure how fast a new employee will learn to use a complex factory monitoring system efficiently, and 3) marketing team of an educational software company needs proof that using their software has positive effect on students' test results over time. Since evaluation takes time and product development needs user feedback as soon as possible, long-term studies may be most beneficial for companies that develop updatable software (e.g. web services, mobile applications) or interactive products based on previous product versions (e.g. mobile phones, cars, domestic appliances).

CONCLUSIONS

This study was set out to provide more empirical evidence in how practitioners in companies evaluate the usefulness of long-term UX evaluation. The question was studied through three long-term case studies in a company developing interactive sports products. The results of this study suggest that managers and developers perceive long-term UX evaluation generally interesting, relevant and useful. Practitioners found the results relevant for 1) comparing the results with previous knowledge, 2) understanding the change in UX over time, 3) focusing future work, 4) concepting and development of future products, and 5) updating current software products. However, challenges remain related to the time and resources needed for conducting long-term studies: 1) research results may arrive too late to benefit ongoing product development and 2) other sources, such as customer care, may provide similar information, which decreases the usefulness of the results of long-term studies.

The main limitations of this study were that only one company was involved and the sample size of product development practitioners was small, especially when measuring the actual use of the results from case study SWb. Also, no responses were received to the follow-up survey from designers or UX specialists, who should be the most obvious people to utilize UX evaluation results. Furthermore, practitioners' feedback regarding the usefulness of the results could have been different if they had spent more time inspecting the evaluation reports before answering. However, in reality, busy managers and designers might not have time to inspect lengthy research reports and therefore live presentations may sometimes be the only channel to deliver research results.

This study highlights some of the benefits and challenges related to long-term UX evaluation in practical product development work. The empirical findings can inform HCI practitioners and contribute to future research on how long-term UX evaluations are conducted in industry. In future, more extensive research with different product development companies and their practitioners is required to determine how long-term UX evaluation results are used in practice, especially by designers and UX specialists. Also, little is known on how to actually design memorable and positive long-term user experiences [10]. Another interesting topic would be the ways of speeding up the process for providing actionable results from long-term UX studies.

REFERENCES

1. Bark, I., Følstad, A., and Gulliksen, J. Use and usefulness of HCI methods: results from an exploratory study among Nordic HCI practitioners. *People and Computers XIX — The Bigger Picture*, (2006), 201-217.
2. Csikszentmihalyi, M., and Larson, R. Validity and Reliability of the Experience-Sampling Method. *Journal of Nervous and Mental Disease*, 175 (1987), 526-537
3. Hassenzahl, M., Burmester, M., and Koller F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität In: J. Ziegler & G. Szwillus (Hrsg.), *Mensch & Computer 2003. Interaktion in Bewegung*, (2003) 187-196.
4. Jain, J., Rosenbaum, S., and Courage, C. Best practices in longitudinal research. *Ext. Abstracts CHI '10*, (2010), 3167-3170.
5. Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., and Stone, A. A. A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science 306*, 5702 (2004), 1776-1780.
6. Karapanos, E., Martens, J.-B., and Hassenzahl, M. Reconstructing experiences with iScale. *Int. J. Human-Computer Studies 70*, 11 (2012), 849-865.
7. Karapanos, E., Jain, J., and Hassenzahl, M. Theories, methods and case studies of longitudinal HCI research. *Ext. Abstracts CHI'12*, (2012), 2727-2730.
8. Kujala, S. and Miron-Shatz, T. Emotions, experiences and usability in real-life mobile phone use. *Proc. CHI '13*, (2013), 1061-1070.
9. Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., and Sinnelä, A. UX Curve: A method for evaluating long-term user experience. *Interacting with Computers 23*, 5 (2011), 473-483.
10. Kujala, S., Vogel, M., Pohlmeyer, A., and Obrist, M. Lost in Time: The Meaning of Temporal Aspects in User Experience. Ext. Abstracts CHI '13, (2013), 559-564.
11. Mendoza, V. and Novick, D. Usability over time. *Proc. SIGDOC '05*, (2005), 151-158.
12. Norman, D.A. Memory is More Important than Actuality. *Interactions,* March + April (2009), 24-26.
13. Novick, D. and Santaella, B. Short-term methodology for long-term usability. *Proc. SIGDOC '12*, (2012), 205-211.
14. Ployhart, R. E., and Ward, A.-K. The "Quick Start Guide" for Conducting and Publishing Longitudinal Research. J. Bus. Psychol. 26, 4 (2011), 413-422. doi:10.1007/s10869-011-9209-6
15. Vaughan, M. and Courage, C. SIG: capturing longitudinal usability: what really affects user performance over time? *Ext. Abstracts CHI '07*, (2007), 2149-2152.
16. Vaughan, M., Courage, C., and Rosenbaum, S. Longitudinal usability data collection: art versus science? *Ext. Abstracts CHI '08*, (2008), 2261-2264.
17. von Wilamowitz-Moellendorff, M., Hassenzahl, M., and Platz, A. (2006). Dynamics of user experience: How the perceived quality of mobile phones changes over time. In *User Experience – Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer Interaction*, 74-78.

P4

MEASURING RETROSPECTIVE USER EXPERIENCE OF NON-POWERED HAND TOOLS: AN EXPLORATORY REMOTE STUDY WITH UX CURVE

by

Varsaluoma, J. and Sahar, F. 2014.

In Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content & Services (AcademicMindTrek'14). ACM, New York. pp. 40-47.

Measuring Retrospective User Experience of Non-Powered Hand Tools: An Exploratory Remote Study with UX Curve

Jari Varsaluoma
Tampere University of Technology
Korkeakoulunkatu 6, P.O. Box 589, 33101
Tampere, Finland
+358 40 8490856
jari.varsaluoma@tut.fi

Farrukh Sahar
Tampere University of Technology
Korkeakoulunkatu 6, P.O. Box 589, 33101
Tampere, Finland
farrukh.sahar@tut.fi

ABSTRACT
User experience (UX) can change over time, when the exciting new product becomes an ordinary part of daily life. With experiences, also our evaluation of a product can change, affecting whether we will recommend the product for others. Studying the long-term UX of products with retrospective methods offers an alternative to the resource demanding real-time studies. In this exploratory remote study, two methods from Human-Computer Interaction field, UX Curve and Attrakdiff, were used for measuring the UX of a non-interactive product. UX Curve was found suitable for evaluating changes in pruning shears' pleasantness after three months of use. Pragmatic product aspects mainly affected the changes in product pleasantness over time. The Pleasantness curve's shape related to the number of positive experiences, overall goodness and willingness to recommend the product. Interestingly, Attrakdiff measures proposed that in addition to pragmatic qualities, also hedonic qualities identification and beauty were related to customer loyalty.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces – *evaluation/methodology, theory and methods, ergonomics*

General Terms
Human Factors, Design, Measurement

Keywords
Long-term; retrospective method; user experience; UX; hand tools; pruning shears; UX Curve; Attrakdiff; hedonic; pragmatic.

1. INTRODUCTION
The Human-Computer Interaction (HCI) community still lacks a universally accepted definition for the concept of *user experience* (UX). However, based on a survey study with 275 respondents from industry and academia, researchers and practitioners tend to agree that UX is dynamic, context-dependent and subjective [18]. UX changes over time, starting from expectations before use and the first use experiences with a product. It has been proposed that through momentary and episodic user experiences, user forms a cumulative UX, which is based on reflections of several use episodes [25]. Industry is potentially interested in long-term UX, as measuring the overall product UX is more important than understanding momentary feelings when people evaluate products [18].

AcademicMindTrek '14, November 04 - 07 2014, Tampere, Finland
Copyright 2014 ACM 978-1-4503-3006-0/14/11…$15.00
http://dx.doi.org/10.1145/2676467.2676485

Measuring UX over time can be resource-demanding and cause participant fatigue when studied in real-time or with repetitive measurement methods. Retrospective methods, such as UX Curve [16], offer an alternative for studying long-term UX of products. Although retrospective methods rely on user's memories, it has been argued that these memories affect individual's future decisions and what will be reported to others [11, 22]. Therefore it is relevant to study UX also from longitudinal perspective, in spite of the possible bias in the recalling process. Based on recent research [15], the overall evaluation of product UX seems to be as strongly dependent on memories as on the actual experience episodes. Users with positive memories from a product will have positive stories to share with others, certainly a desired goal for any product development company.

Hassenzahl [5, 6] describes UX as *pragmatic* and *hedonic* product attributes. Pragmatic attributes relate to user's needs to achieve behavioral goals, while hedonic attributes relate to user's self. Previous studies [1] with interactive products have shown that the importance of hedonic quality increases in long run, improving customer loyalty. However, more research is needed about the product attributes related to the long-term experiences with non-interactive products. For example, what is the meaning of hedonic qualities, such as beauty, in the long-term evaluation of pragmatic, non-interactive products?

In this paper, we present an exploratory study conducted in cooperation with a Scandinavian company developing gardening tools. Our aim was to 1) understand what factors affect the changes in the product pleasantness of a pragmatic tool (pruning shears) when evaluated retrospectively after three months of use, 2) explore the suitability of two UX evaluation methods, UX Curve [16] and Attrakdiff [7], in a remote UX evaluation of a non-interactive, pragmatic product, and 3) study how the remembered experiences relate to customer loyalty and the product's perceived pragmatic and hedonic qualities. We chose to apply UX measures on a non-interactive product in an exploratory manner and to gain knowledge of non-interactive product use pleasantness as design information for potential interactive products. Further, pruning shears are very tangible, held in hand, and have an excellent physical interface for use, thus lending themselves to be studied as a tangible interface.

Remote UX evaluation can be less resource-demanding for practitioners in industry, when compared with field studies that require travelling and face-to-face meetings with users. Therefore, it is important to explore the suitability of remote UX evaluation methods also in the context of practical, non-interactive products. Since to our knowledge UX Curve has not been used in remote studies before, we also saw this as an opportunity to explore how participants would react to this rather novel method when only written instructions are provided. Furthermore, as a pen-and-paper method, UX Curve is potentially suitable for respondents who have no access to or have no experience with computers, such as elderly people who are potential users of gardening tools. In addition, more research is needed in different contexts to validate the previous findings utilizing UX Curve method. For example, the relationship between of the remembered experiences and 1) product evaluation and 2) customer loyalty requires additional research with different product types.

Kujala et al. [16] found that satisfied mobile phone users drew improving Attractiveness curves. We were interested to study the relation of curve shapes (depicting pleasantness) and remembered experiences with product qualities measured with Attrakdiff questionnaire [7]. AttrakDiff provides quantitative data describing user's perceptions towards the evaluated product. Although Attrakdiff was originally developed to be used with interactive products, we believe that most of the Attrakdiff word pairs are suitable for evaluating the UX of a pragmatic, non-interactive tool, such as pruning shears.

To our knowledge, this is the first study where UX Curve method has been used remotely. Also, the study has novelty value by focusing on the meaning of hedonic factors in the long-term UX of working tools such as pruning shears.

Due the vast number of different factors related to UX, product designers face challenges in identifying the most critical UX factors for their designs and choosing cost-efficient methods for measuring these factors. It is worthwhile to explore how experience evaluation methods from different domains, e.g. UX and interactive products, could also benefit the designers of pragmatic, non-interactive tools. This study can inform practitioners working with tangible interface products in planning and conducting retrospective UX evaluations. Furthermore, this study can motivate future studies that experiment with methods traditionally used in other domains in order to find potentially beneficial approaches to evaluate users' experiences with different product types. Moreover, the findings can contribute to the future development and utilization of retrospective UX evaluation methods, similar to UX Curve.

2. RELATED STUDIES

2.1 Retrospective long-term UX evaluation

Long-term UX studies can be divided into two categories: repetitive measurements and retrospective evaluations. Repetitive measurement methods, such as the Day Reconstruction Method (DRM) [10], provide rich understanding of the users' experiences over time (e.g. [13]). However, retrospective measurements can be less taxing to both the participants and researchers. In a study by von Wilamowitz-Moellendorff et al. [29] a retrospective CORPUS technique was used to measure the remembered experiences of mobile phone usage over 12 months. More recently, Kujala and Miron-Shatz [15] examined emotions, experience episodes and their impact on overall product evaluation during real-life mobile phone use over a five-month period. They used DRM [10] during the first five days and retrospective questionnaire after 2.5 and 5 months.

Karapanos et al. [12] developed an electronic curve drawing tool iScale that was used to study users' experiences with mobile phones after 4 to 18 months of usage. At the same time, Kujala et al. [16] developed a paper-and-pen curve drawing method called UX Curve that aims at supporting users to recall essential issues related to their experiences. According to [16], UX Curve is best suited for evaluating products that are already in the market and used on a daily basis or repeatedly over time. Also, UX Curve can support users in remembering more details in their experiences when compared with CORPUS interviews [16]. Therefore, UX Curve is potentially suitable for studying the UX of non-interactive products, such as pruning shears. So far, UX Curve has been used in measuring the remembered experiences with interactive products, namely Facebook [17] and mobile phones [16]. The previous findings suggest that with the interactive products the changes in long-term UX are more related to the hedonic qualities than pragmatic ones. Recently, Moschou and Zaharias [21] proposed an alternative version of UX Curve. The method was tested in an empirical study with nine gamers. Findings revealed that the alternative version can be an effective method for measuring the temporality of gamers' experiences.

2.2 UX of non-powered hand tools

According to Haapalainen et al. [4] working with poorly designed hand tools can lead to a serious increase in occupational disorders such as carpal tunnel syndrome and cumulative trauma disorders of other kind [26, 27]. Every year throughout Europe these problems heavily impact on the human suffering as well as on economic losses because of sick leave, medical care, compensation, rehabilitation, and premature pensions. Thus there is a need to improve the product design practices of hand tools, also by studying different evaluation methods that can inform design.

Several studies have been published related to hand tool design. A comprehensive design guide for different kinds of hand tools has been compiled in [2] and [20]. Kadefors et al. [9] have published a list of factors for evaluating the ergonomics of hand tools. Kilbom et al. [14] have studied different plate shears designs, user characteristics and performance. Rok Chang et al. [24] quantitatively evaluated the effects of handle/material types with three different gardening tools (shovel, rake and hoe) by measuring 1) user satisfaction in terms of work performance, 2) ergonomic effectiveness, and 3) subjective judgments of tactile feel and control.

Some published studies specifically focus on the design of pruning shears. In the Eurohandtool project [19] the researchers developed ergonomically oriented design methodology, Quality Function Deployment. This method can be used in improving the ergonomics of pruning shears by integrating the end-user into the design process. In [30], Wakula and Landau evaluated hand-operated pruning shears from five different manufacturers. They employed objective and subjective methods (EMG, EKG, questionnaires) in the laboratory and field studies.

Overall, there are few long-term studies conducted in real use context considering the use of pruning shears. Furthermore, it seems that majority of the publications concerning the design of gardening tools, hand tools or pruning shears are ergonomic-centric and focus solely on the pragmatic aspects such as work performance and effectiveness. Since the word "user experience" has rarely been used in this context, as it relates to the use of more interactive products, this study has novelty value in considering the hedonic aspects of pruning shears in combination with our focus on the shears as a tangible, pleasant user interface.

3. METHOD

3.1 Product

The product evaluated was a pruning shears designed to be used with one hand. The shears had rotating handle mechanism with adjustable grip span and bypassing blades. Shears came with a plastic carrying case and weighted 272 grams. According to the instructions for use, the shears were recommended for cutting live branches of maximum 20 mm in diameter. The participants received the shears at the beginning of the study by mail. We expected that the rotating handle mechanism would add some novelty value, but would require a longer learning period when compared with "traditional" shears design.

3.2 Participants

Employees and students from six Finnish vocational institutes offering education for horticulturalist studies were recruited for the study. First, the teachers responsible for horticulturalist education were contacted via phone or email. The contact persons forwarded an invitation email for their colleagues and horticulturalist students. Due to the recruitment process, it was not possible to calculate how many people actually received the invitation.

The invitation email contained a web link to the screening questionnaire. From 80 respondents who answered the questionnaire, 30 were chosen for this study. The screening criteria were: a participant 1) had previous experience in using different pruning shears, 2) was planning to use pruning shears in the coming summer and autumn, and 3) had none or very little previous experience with the product under study. One participant was dropped out during the study and therefore responses from 29 users were included in the analysis.

Participants (*N*=29, 22 female) were between 20 to 62 years old (M=39.9 years). 22 participants were students and 11 school personnel or in working life. Four of the participants were both studying and in working life. Five participants had used only one brand of pruning shears before, four had used two brands and 20 had experience of three or more different brands. The studied product was revealed to the respondents at the end of the screening questionnaire. All respondents were familiar with the brand, but only one had previous experience in using similar shears (used 2-5 times).

3.3 Procedure

User experiences were measured retrospectively after three months of product usage from the middle of June to the middle of September in 2012. Shears were mailed to the participants at the beginning of the study period. This approach was chosen because we were interested in capturing the experiences from the very beginning of use. It would have been difficult to find enough participants who had bought similar shears right before the study period. In addition, we could not afford to miss the summer season when the product was mainly used in real work context in Finland. Although the setting would have been more realistic if users had bought the products themselves, it is common in the gardening profession that work tools are provided by the employer or educational establishment, and therefore users have to use the tools available. Furthermore, from the product development perspective, it is common practice that the tested products are given to the users for free for purposes of collecting user feedback. Participants were instructed to use the shears at home and work as if they had bought the shears from a store themselves.

After 3 months, a paper questionnaire was mailed to the participants, containing a UX Curve drawing task and separate questionnaire form. Three months usage time was chosen because it was expected to cover the period when the pruning shears were mainly used in the gardening work and to be long enough time for acquiring the remembered experiences during and after the initial learning period.

Participants sent their answers back to the researchers in a ready paid envelope received with the questionnaire. Finally, as a reward, the participants got to keep the shears used in the study and participated in a lottery of one gardening equipment kit (worth 100 €).

3.4 Materials

3.4.1 Initial questionnaire

The initial questionnaire was conducted using an online-tool Webropol. In addition to questions for collecting basic demographic data, users were asked to describe what perfect pruning shears would be like and what three things they expect when choosing pruning shears for themselves.

3.4.2 UX Curve after three months of use

In the UX Curve drawing task user was asked to draw a curve that represents how the *pleasantness* of the product had changed from the first use experiences until three months of usage. Pleasantness was defined to the participants as "how good the product is or feels" and was therefore expected to represent the overall goodness of the product. The horizontal axis (time) was divided into three parts, representing the three months of usage (see Figure 1). In addition to drawing the curve, user was asked to write comments (experience narratives) to explain the reasons why the pleasantness of the pruning shears had changed over time.

To avoid participant fatigue, each user was asked to draw only one curve. In a study with mobile phone users, Kujala et al. [16] found that *Attractiveness curve* provided the highest number of reasons for the changing UX. However, for a practical product such as pruning shears, we decided that *product pleasantness* would be a more familiar term when translated into Finnish. Pleasantness curve should also capture a wider spectrum of pragmatic and hedonic experiences, e.g. feeling in hand, cutting efficiency, and evaluation of visual appearance.

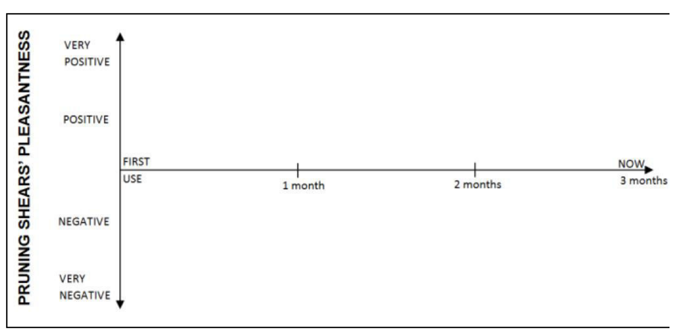

Figure 1. The UX Curve template used in the study. Translated from Finnish.

The UX Curve task was presented on a A3-sized paper form to provide enough space for writing experience narratives that explain the changes in pleasantness. Before drawing the curve, user was prompted to read the instructions with examples on how to add comments to the curve. In order to stimulate the recalling process, user was asked to have the shears available when drawing the curve. After the curve drawing task, user filled in a final questionnaire on paper.

3.4.3 *Questionnaire after three months of use*

In the final questionnaire, the use frequency and use situations of the pruning shears were asked. At the end of the questionnaire, we asked feedback from the UX Curve drawing task and the study in general.

Brand and product recommendation. To predict customer loyalty, that is supposedly related to brand recommendation [23], users were asked if they were willing to recommend the brand to their friends or colleagues on a scale from 0 (not at all likely) to 10 (very likely). In addition, we asked if participants would recommend the evaluated product for others.

AttrakDiff. For comparison to the UX Curve results, we measured the perceived UX of pruning shears with AttrakDiff2 questionnaire [7]. The AttrakDiff questionnaire used in this study consisted of 23 semantic differentials (word pairs) on a 7-point scale. Similar to [5] we used three categories to describe product characters: 1) pragmatic quality (PQ), 2) hedonic quality identification (HQI) and 3) hedonic quality stimulation (HQS). In addition, two semantic differentials were used to evaluate beauty (beautiful-ugly) and overall goodness (bad-good) of the product. The order of the word pairs was randomized. Before the study, a workshop was held where the used word pairs were translated into Finnish by a group of Finnish HCI experts.

3.5 Analysis

Before analysis, the UX Curve forms (depicting pleasantness) were transferred into the digital format. Paper forms were scanned and the curves redrawn on top of the scanned pictures using drawing software. Each curve was placed on a separate drawing layer for an easy creation of collections of different curve types.

Curves were categorized into improving, stable and deteriorating based on the difference between their starting and ending points, similar to Kujala et al. [16]. In addition, each curve was categorized as "positive/neutral/negative start" and "positive/neutral/negative end", depending if the start and end points were above or below the middle line of the vertical product pleasantness scale.

Experience narratives explaining the changes in pleasantness were content analyzed. A single experience narrative could contain several items, both positive and negative, that were coded into appropriate categories. Categories that served as an initial framework were derived from literature and previous studies containing analysis of experience narratives. For experiences related to *pragmatic* product aspects, the following categories were used: Effectiveness, Efficiency, Satisfaction [8] and Utility [3]. For *hedonic* product aspects, the categories were: Stimulation, Identification, Evocation and Beauty [5, 6]. During the analysis, category for "expectations" was added. The categorization was first conducted by one of the authors and iterated together with a second author. Any disagreements were discussed, until consensus was reached.

To assess the internal consistency of the three Attrakdiff categories, Cronbach's alpha coefficient was calculated. Results were as follows: PQ .84, HQI .89 and HQS .75. Three word pairs with low corrected item–total correlation values (r < .29) were removed from the analysis to increase the reliability of the scales: *technical-human* from PQ increasing Cronbach's alpha to .87 and *lame-exciting* together with *easy-challenging* from HQS increasing Cronbach's alpha to .80. This could indicate that these word pairs are not suitable for evaluating a non-interactive, pragmatic product such as pruning shears.

4. RESULTS

4.1 Product use frequency

Product use frequency was measured with two scales: 1) the average number of days per month when shears were used and 2) the longest use time in any single day. Majority of users (20/29) had used the shears between 4-14 days a month. The maximum usage period during any single day for most users (24/29) was from 30 minutes to 5 hours.

4.2 UX Curve results

4.2.1 *Trends of the curves*

From the 29 collected Pleasantness curves, 17 (59%) were improving, 1 (3%) stable and 11 (38%) deteriorating. In addition, as presented in Table 1, majority of the curves started and finished at the positive side of the vertical pleasantness scale. The improving curves are presented in Figure 2. Figure 3 presents the deteriorating curves and the one stable curve (ID42). None of the curves ended at the middle line of the vertical axis ("neutral end").

The number of improving curves and increase in the number of positive end points compared with starting points suggests that the product pleasantness was improving in general. However, as Figures 2 and 3 illustrate, some users experienced major changes in the product pleasantness during the three months. One user (ID10) had stopped using the product after two months and therefore left the curve unfinished.

4.2.2 *Experience narratives*

In total 107 experience narratives explaining the changes in product pleasantness were collected. On average, one participant wrote 3.7 narratives, 41 words and 293 characters. The number of narratives per user varied from 1 to 7. Three narratives were not on the timeline as two users did not mark their narratives' positions on the curve.

159 separate items were identified from the experience narratives. 89 (56%) items were categorized as positive, 11 (7%) as neutral and 59 (37%) as negative. Table 2 shows how the reported items were divided between the three-month timeline (N=27, since 2 positive and 1 neutral items that were not marked on the timeline are not included).

Table 1. Categorization of the Pleasantness curves based on curve starting and ending point.

N=29	**Curve start point**	**Curve end point**
Positive	16 (55%)	21 (72%)
Neutral	4 (14%)	0 (0%)
Negative	9 (31%)	8 (28%)

To compare the number of items between different months, a repeated measures ANOVA test was used. Statistically significant differences were found with the amount of total ($F(2, 56) = 7.166$, $p < 0.005$) and negative items ($F(2, 52) = 7.166$, $p < 0.005$). Post hoc analysis with Bonferroni correction revealed a statistically significant decrease in the number of a) total items between 1^{st} and 2^{nd} ($p < .05$), 1^{st} and 3^{rd} month ($p < .05$), and b) negative items between 1^{st} and 2^{nd} month ($p < .05$). These findings suggest that users report less, especially negative, experiences after the 1^{st} month of use when a retrospective curve drawing method is used. Users' main concerns related to ergonomics, size, weight and the grip span of the shears, aspects that are noticeable during the first use experiences and perhaps therefore reported on the first month. Positive experiences that were reported during the first month related mainly to satisfaction and efficiency, e.g. "light to use", "lightweight", "easy to cut" or "effortless to cut". Users were learning and getting used to the product already during the first months, e.g. "better grip after getting used to" (1^{st} month), "getting used to the rotating handle" (1^{st} month) and "after training, shears have become a bit easier to use" (2^{nd} month). These learning experiences might partly explain the decrease in the number of negative items after the first month.

Categorization of the items is presented in Table 3. 158 (96%) of all items were related to pragmatic aspects, mainly satisfaction, efficiency and utility. Only 6 (4%) items related to hedonic aspects: identification and beauty. However, satisfaction is often based on filling expectations (product worked as it should) or needs for the product (it felt pleasant), and can be considered semi-hedonic. 5 items were related to the carrying case that came with the shears. 9 positive items and 10 negative items were coded into more than one category. For example, item "better grip after getting used to the lever mechanism, several pruning tasks" (ID50) was coded in positive categories efficiency and satisfaction, and "blades get stuck often because of tree bark" (ID22) was coded in negative utility and efficiency. No experiences were related to stimulation or evocation categories.

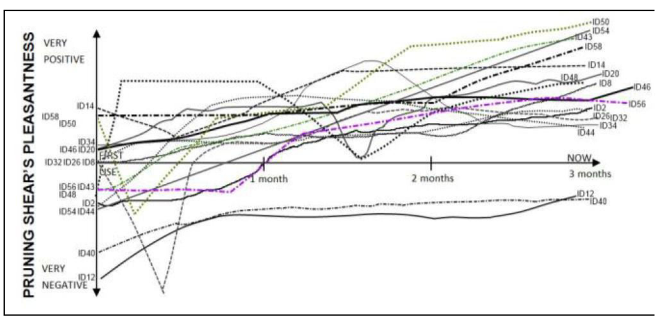

Figure 2. The improving (17) pleasantness curves.

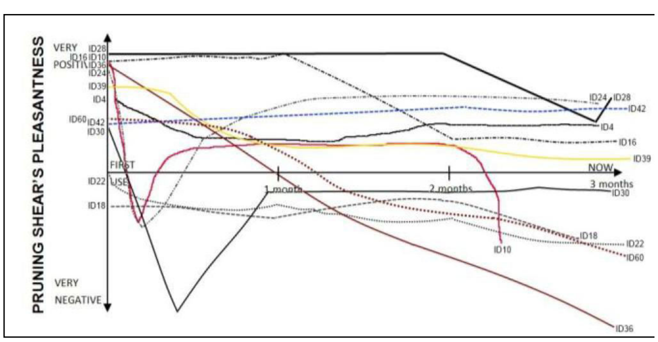

Figure 3. The deteriorating (11) and one stable (ID42) pleasantness curves.

Positive experience items. Satisfaction comments related mainly to ease of use, nice feeling in hand (e.g. ergonomic, sturdy), overall pleasantness and lightness. Efficiency items related to effortless or fast use (for cutting) and learning to use shears over time, e.g. one user (ID56) wrote that "after a while getting used to the shears and could utilize fingertips in the beginning of squeeze [of the handle]", and later continued, "rotating handle gave additional strength for cutting larger branches". Examples from utility items are: "works well despite the dirt" and "works in different situations". Effectiveness related to good cutting power and precise cutting results. Identification items were situations where user's friends had praised the product. Two users commented that shears had pleasant or hi-quality appearance (coded in the beauty category).

Negative experience items. Negative satisfaction items related to the weight (e.g. too heavy to carry in a pocket), size and grip span (too wide for a small hand, especially for females). Only one of the male participants complained about the size, saying that "in the beginning shears felt too big in hand and therefore clumsy". One user did not like the rotating handle. Efficiency items were related to problems with opening the shears after dirt or sand got into the locking or returning mechanism, therefore slowing down the work process. Furthermore, two users thought that the shears were getting dull after one or two months, therefore requiring more effort when cutting. When shears or part of them did not function properly, the item was coded into the utility category. 10 items on the utility category were also coded into other categories, mainly in efficiency. These items related to problems with shears getting stiff from dirt. One user noted that shears were "not suitable for left handed". For effectiveness, one user was not pleased with the quality of the cutting results and another found the blades too big for delicate cutting. Although the shears were not designed for cutting small flowers, this did not stop users from trying. Related to beauty, only one participant commented that shears had "unattractive looks". In addition, one participant reported negative expectations before the first use, but was later satisfied with the product.

4.3 Questionnaire results

4.3.1 Willingness to recommend

The mean score for willingness to recommend the product (on a scale from 0 to 10) was 6.04 (SD: 3.21) and for the brand it was 7.57 (SD: 2.24). Although four respondents were clearly more willing to recommend the brand than the product (difference 4 or higher), there was a strong correlation between the two measurements (Spearman's rank order correlation, $N = 28$, $r = .700$, $p < .001$). We compared both recommendation scores between genders with Mann-Whitney U test, but no statistically significant differences were revealed.

Table 2. Number of items (reasons) that described the changes in the product pleasantness over three months' time.

Items (*N*=27)	1st month	2nd month	3rd month	Total
Positive	38	19	30	87
Neutral	4	5	1	10
Negative	34	11	14	59
Total	76	35	45	156

Table 3. Categorization of the items that affected product pleasantness during the first 3 months of use.

Category (*N*=29)	**Positive**	**Negative**
Satisfaction	47 (48%)	35 (52%)
Efficiency	25 (26%)	18 (27%)
Utility	12 (12%)	11 (16%)
Effectiveness	8 (8%)	2 (3%)
Identification	3 (3%)	-
Beauty	2 (2%)	1 (2%)
Expectations	-	1 (2%)
Total	97 (100%)	68 (100%)

4.3.2 Attrakdiff results

Table 4 illustrates users' evaluation (after three months of product use) of pruning shears' beauty, overall goodness and the three product characters: pragmatic quality (PQ), hedonic quality identification (HQI) and hedonic quality stimulation (HQS).

All mean scores were above the scale average (4). Spearman's Rank Order correlation showed strong relationships between PQ and HQI ($r_s(27) = .782$, $p < .001$), PQ and beauty ($r_s(27) = .702$, $p < .001$), PQ and goodness ($r_s(27) = .767$, $p < .001$), HQI and beauty ($r_s(27) = .710$, $p < .001$), HQI and goodness ($r_s(27) = .886$, $p < .001$), and beauty and goodness ($r_s(27) = .744$, $p < .001$). Stimulation (HQS) did not correlate with any other factor and it also had the lowest mean and standard deviation values. It is possible that stimulation factor is rather consistent for a practical tool such as pruning shears, especially when measured after three months of use, when the product offers less novel or stimulating experiences.

4.3.3 Respondents' feedback from the UX Curve task

19/29 (66%) respondents found the curve drawing task easy to do or clear, but 9/29 (31%) replied that it was challenging to remember all the experiences or their order after three months. One participant thought the task was fun and two other appreciated the free style for reporting experiences. One user found it challenging to evaluate the pleasantness in one scale, when the product had both positive and negative aspects, suggesting that more than one curve types could be used. However, this would have been more taxing for the participants, as two of them already commented that it required some time to draw a single curve.

4.4 Curve trends, reported experiences, product evaluation and customer loyalty

We were interested to study how the different Pleasantness curve trends related to the number of reported positive and negative experiences. Non-parametric Mann-Whitney U tests were used since the data was not normally distributed. The test revealed that participants reported significantly more positive items when they drew improving curves ($U = 43.0$, $p < .05$) or curves with a positive end point ($U = 18.0$, $p = .001$). Interestingly, there was no similar connection between the number of negative items and deteriorating or "negative end" curve shapes, as many negative experiences were also reported with improving curves.

Table 4. Attrakdiff results after 3 months of use (scale 1-7).

Product quality (*N*=29)	**Mean (standard deviation)**
Pragmatic quality (PQ)	4.61 (1.25)
Identification (HQI)	4.82 (1.14)
Stimulation (HQS)	4.56 (0.95)
Beauty	5.03 (1.32)
Goodness	4.90 (1.97)

Next, comparisons between the different curve trends considering a) product evaluation (Attrakdiff) and b) customer loyalty (willingness to recommend) were made using Mann-Whitney U tests. Results show 1) significantly higher ratings in HQI ($U = 48.5$, $p < .05$), goodness ($U = 48.5$, $p < .05$) and product recommendation ($U = 44.5$, $p < .05$) for improving curves, and 2) significantly higher ratings in PQ ($U = 23.0$, $p < .005$), HQI ($U = 8.0$, $p < .001$), beauty ($U = 16.0$, $p < .005$), goodness ($U = .000$, $p < .001$), product recommendation ($U = .000$, $p < .001$) and brand recommendation ($U = 18.5$, $p < .005$) for curves with a positive end point (depicting current pleasantness).

Spearman's correlation test was used to measure the relationships between the number of reported positive and negative items per user, evaluation of product qualities and willingness to recommend. Relationships were found between the number of reported positive items with PQ ($r_s(27) = .451$, $p < .05$), HQI ($r_s(27) = .575$, $p = .001$), goodness ($r_s(27) = .639$, $p < .001$), product recommendation ($r_s(27) = .626$, $p < .001$) and brand recommendation ($r_s(27) = .644$, $p < .001$). Furthermore, product recommendation had strong correlation with PQ ($r_s(27) = .660$, $p < .001$), HQI ($r_s(27) = .715$, $p < .001$), beauty ($r_s(27) = .548$, $p < .005$) and goodness ($r_s(27) = .787$, $p < .001$). Similarly, brand recommendation had strong correlations with PQ ($r_s(27) = .705$, $p < .001$), HQI ($r_s(27) = .600$, $p < .005$), beauty ($r_s(27) = .533$, $p < .005$) and goodness ($r_s(27) = .660$, $p < .001$). Interestingly, the number of negative items did not correlate with any of the measured product qualities or recommendation scores.

5. DISCUSSION

The majority of the participants found the UX Curve method easy and clear to use remotely with written instructions, and successfully reported their remembered experiences with pruning shears over the three months. The accuracy of the reported experiences can be questioned, as one third of the respondents had difficulties in remembering the experiences after three months. However, the participants could still provide useful feedback of their most relevant experiences and reasons for the changes in pleasantness. We were able to pinpoint the most meaningful product qualities over time that affected product pleasantness. Our categorization framework was derived from literature, with a focus on the pragmatic and hedonic product aspects. However, alternative frameworks could be more suitable for providing practical information for designers of tangible products, e.g. with a specific focus on ergonomic aspects.

Our results indicate, that the remembered change in product pleasantness (improving/deteriorating curve) over time relates to the current willingness to recommend the product, evaluation of products' overall goodness and hedonic quality identification. It seems, that users, who drew improving curves, were also likely to remember more positive experiences (items), and have higher ratings in product's overall goodness and customer loyalty (willingness to recommend).

The current pleasantness with the product (positive/negative end point) seems to strongly relate to 1) the willingness to recommend the product and brand, and 2) all the evaluated product qualities (Attrakdiff), apart from hedonic quality stimulation. Furthermore, the positive end point (current pleasantness) seems to be related to the number of positive items (experiences) remembered, pruning shears' overall goodness and customer loyalty (willingness to recommend). However, the starting point of the curve did not relate to any of the measured factors. This suggests that users can recall changes in their experiences (related to pleasantness) over three month time, also with a pragmatic, non-interactive tool.

Interestingly, the number of remembered negative experiences (items) did not relate to the deteriorating curve shape (pleasantness over time), negative end point (current pleasantness), product quality evaluations or willingness to recommend. The reason for this could be, as Figure 2 illustrates, that several of the improving curves had "negative start" or were deteriorating at some point. These participants reported negative experiences with the product, especially at the beginning of use, but their overall experience was improving over time. Similarly, some of the deteriorating curves (Figure 3) included also positive experiences.

Our findings add to the previous study by Kujala et al. [16] with mobile phone users, where the shape of the Attractiveness curve was related to user satisfaction and product recommendation. Our study with pruning shears suggests, that the shape of the Pleasantness curve (improving/deteriorating) relates not only to product recommendation, but also overall goodness and hedonic quality identification.

The UX Curve method with Attractiveness curve has been found to elicit more hedonic aspects with interactive products, when compared with interviewing technique CORPUS [16, 17]. However, in our study, the Pleasantness curve did not seem to provide many hedonic experiences with the studied pragmatic, non-interactive product. One reason for this could be the definition of pleasantness that was presented to the participants: "how good the product is or feels". This did not explicitly refer to the aesthetics of the product, e.g. "how good the product looks". Nevertheless, it seems that hedonic aspects truly play a very minor role with the UX of hand tools, such as pruning shears, even over long-term use. However, willingness to recommend the product (and brand) seems to be related to the product's beauty and hedonic quality identification, in addition to pragmatic quality. Therefore, it is possible that hedonic aspects may still affect the customer loyalty with pragmatic, non-interactive tools. However, this was not evident when studied with a retrospective, qualitative method such as UX Curve. It may be that hedonic pleasantness of products which heavily emphasize tangible interfaces could be based on pragmatic aspects of use. Naturally, without a tested prototype of interactive, computerized pruning shears this remains a speculation. Further, satisfaction as a need suggests a semi-hedonic factor underlying tactile interface use in the case of pruning shears. However, more long-term research is required on this topic, also with qualitative repetitive measurement methods.

The Attrakdiff method was found rather suitable for evaluating the qualities of a non-interactive product. From the evaluated product attributes, hedonic quality stimulation seemed to be the least fitting for a pragmatic tool such as pruning shears, since two word-pairs (*lame-exciting* and *easy-challenging*) from this category were removed based on Cronbach's alpha. Also, stimulation was not related to product pleasantness. This seems reasonable for a product such as pruning shears, which does not provide new, stimulating experiences over time, compared with interactive products, such as Facebook [16]. However, stimulating experiences could also be provided through related products or additional services, such as company websites (e.g. product videos containing maintenance instructions) and competitions. For example, there is a yearly wood chopping competition organized in Finland for a specific axe brand.

6. CONCLUSIONS

We have explored the use of two UX evaluation methods, UX Curve and Attrakdiff questionnaire, in a remote, retrospective study of pruning shears. To our knowledge, this is the first time when UX Curve method has been used in a remote study. Furthermore, previous studies focusing on the hedonic aspects of pragmatic working tools are rare.

The main limitation of this study is that no comparison was made with other qualitative methods (e.g. repetitive measurement) for evaluating the long-term UX of pruning shears. Also, UX Curve data was collected only with one curve type (pleasantness).

Based on our experiences, studying retrospective UX remotely with UX Curve is a suitable approach also with non-interactive, pragmatic products. However, we agree with Kujala et al. [16] that the resulting qualitative data might not be as rich as in field studies, considering its usefulness to design work. UX Curve method could be supported with user observations that would provide more precise information on the main challenges already described in the experience narratives. From practitioners' point of view, UX Curve seems to be rather easy method to take into use, although the analysis of the reported experiences and curve shapes can be time consuming. In comparison with user interviews, UX Curve may be more efficient with larger sample sizes. However, remote data collection can still be rather time consuming since paper forms need to be mailed for the participants. Web-based alternatives such as iScale [12] and DrawUX [28] may provide some solutions to these challenges in the future.

7. ACKNOWLEDGMENTS

The authors would like to thank Piia Nurkka and Tiina Koponen for their assistance in data collection and Prof. Timo Saari for his helpful comments. This study was funded by Tekes - the Finnish Funding Agency for Innovation.

8. REFERENCES

[1] Chitturi, R., Raghunathan, R., and Mahajan, V. 2008. Delight by Design: The Role of Hedonic Versus Utilitarian Benefits. J. Mark. 72 (May 2008), 48-63.

[2] Eastman Kodak Company. 1983. Ergonomic design for people at work, (vol. 1), Workplace, Equipment, and Environmental Design and Information Transfer. van Nostrand Reinhold, NY.

[3] Grudin, J. 1992. Utility and Usability: Research Issues and Development Contexts. Interact. Comput. 4, 2 (Aug. 1992), 209-217.

[4] Haapalainen, M., Kivistö-Rahnasto, J., and Mattila. M. 1999. Ergonomic design of non-powered hand tools: An

application of quality function deployment (QFD). Occup. Erg. 2, 3 (1999/2000), IOS Press, 179-189.

[5] Hassenzahl, M. 2004. The Interplay of Beauty, Goodness, and Usability in Interactive Products. Hum-Comput. Interact. 19, 4 (2004), 319-349.

[6] Hassenzahl, M. 2003. The Thing and I: Understanding the Relationship Between User and Product. In Funology: From Usability to Enjoyment, M. Blythe, C. Overbeeke, A.F. Monk and P.C. Wright, Ed. Hum. Com. Int. vol. 3, 31-42.

[7] Hassenzahl, M., Burmester, M., and Koller, F. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Mensch & Computer 2003: Interaktion in Bewegung, J. Ziegler and G. Szwillus, Ed. B.G. Teubner, Stuttgart, Leipzig, 187-196.

[8] ISO/IEC, 9241-11 Ergonomic requirements for office work with visual display terminals (VDT)s - Part 11 Guidance on usability. 1998: ISO/IEC 9241-11: 1998 (e).

[9] Kadefors, R., Areskoug, A., Dahlman, S., Kilbom, Å., Sperling, L., Wikström, L., and Oster, J. 1993. An approach to ergonomics evaluation of hand tools. Appl. Ergon. 24, 3 (Jun 1993), 203-211.

[10] Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., and Stone, A.A. 2004. A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. Science, 306, 1776-1780.

[11] Karapanos, E., Martens J., and Hassenzahl M. 2012. Reconstructing Experiences with iScale. Int. J. Hum-Comput. St. 70, 11 (Nov. 2012), 849-865.

[12] Karapanos, E., Zimmerman, J., Forlizzi, J., and Martens, J.-B. Measuring the dynamics of remembered experience over time. Interact. Comput. 22, 5, (Sep. 2010), 328-335.

[13] Karapanos, E., Zimmerman, J., Forlizzi, J., and Martens, J.-B. 2009. User experience over time: An initial framework. In CHI '09 Proc., ACM, NY, 729-738.

[14] Kilbom, Å., Mäkäräinen, M, Sperling, L., Kadefors, R., and Liedberg, L. 1993. Tool design, user characteristics and performance: a case study on plate-shears. Appl. Ergon. 24, 3 (June 1993), 221-230.

[15] Kujala, S. and Miron-Shatz, T. 2013. Emotions, Experiences and Usability in Real-Life Mobile Phone Use. In CHI'13 Proc., ACM, NY, 1061-1070.

[16] Kujala S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., and Sinnelä, A. 2011. UX Curve: A method for evaluating long-term user experience. Interact. Comput. 23, 5, (Sep. 2011), 473-483.

[17] Kujala, S, Roto, V., Väänänen-Vainio-Mattila, K., and Sinnelä A. 2011. Identifying Hedonic Factors in Long-Term User Experience. In DPPI'11 Proc. ACM, NY, 22-25.

[18] Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S., and Kort, J. 2009. Understanding, scoping and defining user experience: a survey approach. In CHI'09 Proc., ACM, NY, 719-728.

[19] Leppänen, M. and Mattila M. 2000. Designing the Ergonomic Properties of Pruning Shears Using Quality Function Deployment (QFD). In IEA 2000/HFES 2000 Congress Proc., 44, 22, (July 2000), 647-650

[20] Lewis, W. and Narayan, C.V. 1993. Design and sizing of ergonomic handles for hand tools. Appl. Ergon. 24, 5 (Oct. 1993), 351-356.

[21] Moschou, E. and Zaharias, P. 2013. UX-Curve revisited: Assessing long-term user experience of MMOGs. In CHI'13 Proc., April 27 - May 2, (2013), Paris, France.

[22] Norman, D.A. 2009. Memory is More Important than Actuality. Interactions, March + April, (2009), 24-26.

[23] Reichheld, F. F. 2003. The one number you need to grow. Harvard Business Review 81, 12 (Dec. 2003), 46-54.

[24] Rok Chang S., Park, S., and Freivalds, A. 1999. Ergonomic evaluation of the effects of handle types on garden tools, Int. J. Ind. Ergonom. 24, 1, (Apr. 1999), 99-105.

[25] Roto, V., Law, E., Vermeeren, A., and Hoonhout, J. (Ed). 2011. User Experience White Paper. Outcome of the *Dagstuhl Seminar on Demarcating User Experience*, http://www.allaboutux.org/uxwhitepaper

[26] Silverstein, B., Fine, L.J., and Armstrong, T.J. 1987. Occupational factors and carpal tunnel syndrome. Am. J. Ind. Med. 11, 3, (1987), 343-358.

[27] Silverstein, B., Fine, L.J., and Armstrong, T.J. 1986. Hand wrist cumulative trauma disorders in industry. Br. J. Ind. Med. 43, 11 (Nov. 1986), 779-784.

[28] Varsaluoma, J. and Kentta, V. 2012. DrawUX: web-based research tool for long-term user experience evaluation. In NordiCHI'12 Proc., ACM, NY, 769-770.

[29] von Wilamowitz-Moellendorff, M., Hassenzahl, M., and Platz. 2006. A. Dynamics of user experience: How the perceived quality of mobile phones changes over time. In NordiCHI'06, User Experience - Towards a unified view, Workshop, 74-78.

[30] Wakula J. and Landau, K. 2006. Contribution made by ergonomic investigation towards strain-reducing design of pruning shears. In 16th World Congress on Ergonomics (IAE), July 10th - 14th , (2006), Maastricht (Niederlande).

P5

USAGE DATA ANALYTICS FOR HUMAN-MACHINE INTERACTIONS WITH FLEXIBLE MANUFACTURING SYSTEMS: OPPORTUNITIES AND CHALLENGES

by

Varsaluoma, J., Väätäjä, H., Heimonen, T., Tiitinen, K., Hakulinen, J., Turunen, M. and Nieminen, H. 2017.

In Proceedings of 21st International Conference Information Visualisation - IV2017. IEEE. pp. 427-434.

Usage Data Analytics for Human-Machine Interactions with Flexible Manufacturing Systems: Opportunities and Challenges

Jari Varsaluoma, Heli Väätäjä
Tampere University of Technology
Tampere, Finland
e-mail: first.lastname@tut.fi

Katariina Tiitinen, Jaakko Hakulinen, Markku Turunen
University of Tampere
Tampere, Finland
e-mail: first.lastname@sis.uta.fi

Tomi Heimonen
University of Wisconsin-Stevens Point
Stevens Point, Wisconsin, USA
e-mail: Tomi.Heimonen@uwsp.edu

Harri Nieminen
Fastems Ltd
Tampere, Finland

Abstract. *Analyzing data from complex production systems and processes can be used in improving existing products, processes, and services, and innovating novel offerings. We report the findings from a six-month case study with a company developing flexible manufacturing systems. During a collaborative development process of a data analytics and visualization tool, our goal was to identify potential metrics, business opportunities, and challenges when utilizing logged data of end-users' human-machine interactions in development activities. Our key contributions include a characterization of the potential usage data metrics to be logged and visualized, identification of opportunities this data entails for business, and discussion about the challenges related to usage data logging in the studied context. Finally, we propose topics that should be considered in the organization before investing in usage data logging in the context of flexible manufacturing systems.*

Usage data logging; visual data analytics; human-machine interaction; flexible manufacturing systems; multi-dimensional in-depth long-term case study

I. INTRODUCTION

Suppliers of complex industrial systems in metals and engineering industry are increasingly interested in using various types of log data from systems after their deployment on the market. Industrial information systems record and store data about the status and how end-users use and interact with the complex underlying production systems and processes. In the context of this paper, logged usage data means data logged from system use based on end-user interactions. This includes the system features and functionalities used by end-users along with the associated metadata (e.g., time, data input, and automation state). Although some analytics and visualization solutions exist on the market for companies to use on their own manufacturing data, such as Bosch's manufacturing analytics solutions [1], commercial analysis suites for supporting the business operations and development of supplier companies themselves, who provide manufacturing software and solutions for their customers, are rare or non-existent. One of the challenges is that the data to be collected and used has to be identified to be relevant to the product and business development in the supplier company. This calls for understanding the relevant information, usage goals, and the potential users of the analytics and visualization tools to support the development of useful tools.

Little previous research exists specifically in the domain of exploratory user interaction analysis of complex industrial systems, particularly regarding the expectations supplier companies have for using such data. We contribute to this research topic with empirical findings by exploring 1) which usage metrics should be logged and analysed, 2) what potential business opportunities usage data logging enables, and 3) what are the challenges and obstacles related to usage data logging in the context of manufacturing automation? These questions were studied over a case study during which we developed a visual analytics software framework for a company developing industrial manufacturing automation systems for the metal industry. In addition to describing our study method and the visual analytics tool, we will discuss and compare the results with related studies and propose a set of questions to be addressed in companies before investing in usage data logging. Although it presents significant challenges in terms of data acquisition and analysis, utilizing logged usage data was considered a promising avenue for improving both the manufacturing system supplier's and customers' production processes.

II. RELATED WORK

Visualizing temporal patterns is an active area in the domain of visual data analytics. The process encompasses both the extraction of patterns and presentation of the results. For example, Fails *et al.* [2] demonstrated a system that allows the user to create visual pattern queries by interactively defining sequences of an arbitrary number of events and the interleaved timespans. Results are visualized in a two-dimensional table with matches on rows and time on the horizontal axis. Discovered patterns are represented in ball (event) and chain (timespan) fashion. In the LifeFlow interface [3] the focus is more on presenting an effective overview of all possible event sequences and the temporal spacing of events within the sequences. Multiple interactive features allow for a viewing of details to drill into the constituent events. The DataJewel architecture [4] integrates the database component (linking multiple data sources), algorithmic component (provide access to temporal data mining algorithms), and visualization component – a technique called CalendarView, which leverages users' familiarity with common visualizations such as calendars and histograms. With the interface, users can interactively map attributes to colors to explore hypotheses and apply their knowledge of the data domain, or utilize temporal mining algorithms to create the mappings based on discovered event patterns. Frequence [5] seeks to simplify the process of mining temporal patterns from real-world data. A modified Sankey layout is used to represent patterns as a sequence of nodes (events) and edges (subsequences of the events). Thickness of edges is used to represent pattern frequency and color the relationship of the pattern to an outcome variable of interest. Prominent subsequences can be inferred from the thickness of combined multiple edges across events. Patterns of interest can be dynamically defined by clicking on multiple nodes and the patterns can also be filtered and viewed at multiple levels of detail to account for the hierarchical nature of event data.

Relatively few examples exist in research where visual analytics tools are specifically applied to the analysis of logged user interactions in complex industrial manufacturing systems. Holzmann *et al.* [6] studied the acquisition and visualization of user interaction data from a touch screen based robot controller to find cost-efficient solutions for the usability evaluation of handheld terminals in the automation industry. The goal was to help developers to identify possible problems in users' workflow (e.g. navigation problems or unused functions) based on user interface interactions. Navigation path analysis and usage intensity were identified as the most important topics for data logging, based on interviews with a programmer and two project managers in automation industry enterprises.

Grossauer *et al.* [7] created a prototype for automation industry to visualize navigation flows through an application. After applying the visualization tool to multiple datasets, they recommend such tools should include 1) a wide variety of filters and 2) views that show the whole navigation data and allow the inspection of individual sequences. Further, Kandel *et al.* [8] highlight opportunities in supporting visual analytics workflows in enterprise settings, related to the discovery, management and profiling of data. Such opportunities include being able to do early-stage analysis on partially structured data (such as log files), utilizing existing data warehouses without a need for explicit data migration across systems, and providing direct manipulation interfaces for data acquisition and management.

One of the main challenges of data processing and analysis is to facilitate analysts' tasks. Experts can be working on challenging problems to which no direct answer is available, and they participate in inventing, innovating or discovering activities [9]. Tools designed for these activities should support human performance, error-free use, creative exploration, hypothesis building, and history keeping, as well as collaboration and dissemination to others [9]. One of the aims of producing such tools can be thought of as increasing the analytics literacy of the development team [10]. Interactive data analysis should not only help improve the productivity of technically proficient users but also be accessible to users with limited programming skills [10]. Several different types of enterprise data analysts exist, from proficient programmers to users of dedicated analysis applications [11], and these different user groups with varying skills need to be taken into account and identified as potential users of data analytics and visualization tools.

In big data analytics, a domain that our research context resembles, the analysis tasks are exploratory and happen on-demand, the results are aimed at audiences with little background in data science, and with the need to produce reproducible and reliable results [12]. In effect these tasks fall within a continuum that spans activities focusing on producing known, low impact insights (e.g., static reports) to activities that aim to uncover high impact, previously unknown insights (e.g., real-time alerts, predictions, and recognition of patterns) [13].

Work related to data logging in information security context may also provide insights from the product development perspective. For example, SANS [14] and OWASP Foundation [15] have suggested activities to be logged by critical systems from a security standpoint.

A common approach to acquiring usage data is to instrument software applications to log user interaction events. A straightforward but laborious way is to add logging instructions directly into source code, although this approach may introduce new complexity into the system [16]. As an alternative, many instrumentation frameworks aim to reduce the burden on the application developer by semi-automatically logging relevant interactions, e.g., [16, 17] and several Web analytics frameworks such as Google Analytics. The complexity and level of abstraction of the instrumentation varies by its purpose, but when implementing instrumentation and analysis techniques, one should consider the levels of abstraction of the captured events, how higher level interactions are composed of lower level events, and how to capture contextual information [18].

In our case, we focus specifically on the analysis and visualization of end-user interactions with the production system. The use of interactive systems can be logged as a part of usability and user experience (UX) research with the aim of identifying potential problems and issues affecting the experience as in our case. Such logging can address multiple needs within the product development organization. UX measurement can include the use of device functions, access of features by different user groups, or identification of changes in user behavior [19].

Prior research has identified company needs on data analytics and visualization in the domain of metals and engineering industry [20]. The identified needs for data types are: 1) usage combinations, such as the customer's production type and in what mode they use the system, 2) patterns of use, 3) types of user groups and profiles that can be found, 4) summarizations of the system use based on logged data (logs of events, system status, user actions, interactions etc.), 5) identifying problem or fault situations (individual and possible patterns), and 6) changes in use (such as features) over weeks or months.

III. CASE STUDY

Our case study had three interrelated aims that focus on understanding the use of logged usage data. First, we wanted to understand what kind of logged usage data from the customers' systems are relevant to the supplier in manufacturing industry. This information can support the developers of data analytics and visualization tools in what kind of visualizations are required in the context of FMS to present the desired data.

Second, to justify any investment of resources in usage data logging we needed to understand what business opportunities it can provide in the context of FMS. With this, we can better evaluate the value of usage data logging for different stakeholders.

Third, by identifying challenges specific to usage data logging in FMS context, future research can focus on tackling these obstacles.

Based on our experiences we propose a set of questions for stakeholders in manufacturing system supplier companies interested in utilizing usage data logging in practice to support their business.

A. *Method*

The methodological approach of the case study was based on collecting usage data and participant feedback during an agile development process of a data analytics tool in collaboration with a company developing flexible manufacturing systems. The analytics tool development process was inspired by Shneiderman *et al*. [7 9] multi-dimensional in-depth long-term case studies (MILC) approach. The MILC approach combines field studies, with participant observation, interviews, surveys and automated logging of user activity. In this paper, we focus on the qualitative results of user observation and interview sessions. While the exact interview questions developed during the study period, the main themes we discuss in this paper were:

- Which usage data should be logged and/or visualized?
- What are the potential business opportunities enabled by usage data logging?
- What challenges are related to usage data logging in industrial context?

B. *Study context*

Our study took place in the context of industrial manufacturing automation systems called flexible manufacturing systems (FMS). FMS are typically used in metal industry for manufacturing parts by using different metal operations [21]. In FMS, the production is typically conveyed on pallets [22], on which the parts are attached for machining. FMS is used to automate pallet-based machining for manufacturing small batches of different types of products. Goal is to flexibly change the manufactured product without a need for changing the whole factory layout [21]. In FMS, software and hardware are combined to provide a manufacturing company an easily modifiable, dynamic manufacturing system.

FMS is operated via a combination of graphical user interfaces (used e.g. to enter new manufacturing programs, control the program parameters and modify system status) and physical or software coded buttons for pallet control. In this study, our primary focus is on user interactions by the human operators on the workshop floor with the FMS elements of the manufacturing system.

The case study company was interested in collecting and analyzing usage data of their FMS systems after they had been supplied to the customers. Usage data was expected to benefit the company's R&D, customer support, and service business in the future. While the FMS systems already logged data of their behavior, it was mainly used for on-demand maintenance. The log data portraying users' interactions with the system provided a new channel to study the product usage.

C. *Participants*

A purposive sample of six participants working in the case study company participated in the study. Table I presents the participants' roles, responsibilities and familiarity with analytics tools. For data analytics and visualization purposes, four of the participants used MS Excel. One developer also used unspecified data base tools. One developer (ID6) left the study midway, but his responses until then were included in the analysis.

D. *Data analytics and visualization tool*

Next, we describe the developed data analytics tool, "UX-sensors", at the end of the study period. The user front end of UX-sensors is an interactive web application. It consists of a data selection view and the main data browsing view with timelines and analysis tools. The main data browsing view consists of six main elements that are numbered in Fig. 1. The elements are: 1) overview panel, 2) overview timeline, 3) detail timeline, 4) additional filters, 5) tabs, and 6) main filters.

The overview panel contains general information about the selected observation window e.g., the number of found events, most frequent events by value, the average usage session length, the average number of operations per session and the number of error events. Events are split into sessions

TABLE I. THE CASE STUDY PARTICIPANTS' ROLES AND FAMILIARITY WITH DATA ANALYTICS TOOLS.

ID	Role	Main responsibilities	Use frequency of data analytics and visualization tools?
1	Director	Manages research and innovation development	Weekly
5	Product manager	Manages product life cycle service offerings	Monthly
4	Team leader	Manages technical customer support team	Weekly
2	Team leader	Manages software development team	Not at all
3	Software developer	Front-end & back-end development	Less than once a month
6	Software developer	Front-end development	Less than once a month

Figure 1. Main data browsing and analysis view. The elements are: 1) overview panel, 2) overview timeline, 3) detail timeline, 4) additional filters, 5) tabs, and 6) main filters.

based on maximum time duration between successive events. The default value is five minutes and it can be changed in settings. The overview timeline displays the overall number of events by the hour and works as a filter where the user can restrict the further analysis to a shorter observation window. The detailed timeline displays the individual events within the observation window. By hovering over an event, detailed information is displayed. The user can also add additional notes directly to the timeline. The additional filters element can filter the event set further. In our study setup, six main filters were available: factory, observation window (if multiple windows selected), system, user, level and feature.

The tab elements display processed information about the selected events and provide tools for further exploration of the data. Used features tab displays a line graph of feature use over time and a complete list of features and feature-value pairs with count and percentage of total events. The data table can be sorted by each column and filtered by search. Errors and recovery tab provides a list of the most common errors, average recovery time and user interaction sequences during error recovery. Error recovery time is estimated as the duration between the last successive error event and the first following user interaction (i.e., event that is not an error or warning). Frequent sequences tab is used to calculate and display the most frequently occurring sequences. This is done by splitting the events into sessions and then looking for similarities in the event sequences within the sessions. Through the search tab single events or sequences of events can be searched by defining key-value pairs consisting of e.g., feature and value. Data entry tab is for exploring events indicating user data entry and user interaction sequences during data entry. Lastly, the main filters element can contain up to two filter panels on the right side of the timelines.

E. *Procedure*

Before the first group meeting we conducted a background survey, where we asked for what purposes the participants would like to use logged usage data. Five iterations were made to the data analytics tool during the six-month study period. We had two group discussion sessions and three individual observation sessions with each participant. In the observation sessions, participants could freely use the updated tool for exploring the available log data. During the session, the researcher encouraged the participant with questions such as ”what are you thinking now?” The session ended with an interview. Each session was recorded with a video camera and lasted approximately one hour.

F. *Analysis*

The written notes from the sessions were updated based on the videos and transcribed interviews. The interview data were analyzed by coding responses to descriptive categories.

The main categories included benefits, challenges and future opportunities for usage data logging. Finally, the number of comments in total and by separate respondents for each code were summarized to identify the most discussed topics.

IV. Results

The background survey results (n=6) suggest, that the respondents would like to use logged usage data to a) increase knowledge about customers (4 responses), b) increase knowledge about end-users (4), c) improve remote customer support (4), d) support maintenance (3), e) support training and development of training material (3) and f) develop their service business (3). In addition, usage data could be used to compare how customers' different work teams use the product.

A. *Usage metrics to collect and visualize*

Table II lists the usage metrics and related data types that at least two respondents were interested in to collect and/or see visualized. Next, we will discuss the most often mentioned topics in more detail.

Frequency of system feature use seemed to represent a core data for understanding the system use. The director (ID1) commented that feature usage can help in making development decisions: focusing on popular features and studying the reasons behind unused features. There were also some new features in the system that the supplier had no usage information about. Use frequency of features also acted as a stepping stone towards why-questions, that may require direct contact with end-users. For example, one of the developers (ID3) could not figure out a reason for an unusually occasional use of a specific feature. The product manager (ID5) also proposed that the knowledge of feature use could work as a reference for new customers in sales or marketing when showcasing the most popular features. Finally, feature use frequency might inform user interface (UI) developers optimizing the UI.

A possibility to **combine different logs** from various systems on one site or between several factories around the world for visualization purposes was considered valuable for the supplier's management and customers in supporting fleet management. Per the director (ID1), combining usage data with other key indicators, such as **utilization ratio of the system**, could provide more information about events leading to system downtime periods.

User actions before and after events could help customer support to understand what users did before a specific error event and how they tried to solve the problem. A quote from the customer support team leader:

…it would be interesting to know… what kind of operations the user has done to recover from this error… when one needs to use the system in a way that differentiates from everyday use, what do they do then? Do they leave it or will they try something? (ID4)

Inspecting the **long-term statistics** of usage data could reveal repeated usage patterns that are inefficient, suggesting a need for user training or UI redesign. The product manager proposed that a combined long-term data (e.g., over a 3-month period) from several customers showing similar usability issues could convince stakeholders of the need to improve the shop floor UI. Furthermore, **event sequences** that show repeated usage patterns with specific users and systems may reveal possible problems in the UI design, such as navigation issues.

TABLE II. Stakeholders' interests for usage metrics to be collected and/or visualized.

Interests of stakeholders	Respondents
Use frequency of features	5
Combining different logs	5
User actions before / after event	5
System usage ratio	4
Long-term statistics	4
Event sequences	3
Log data from specific user	3
Log data from specific machine / app.	2
Remote customer support activities	2
Use frequency of UI elements (e.g. buttons)	2
User navigation in UI	2
Frequency of same error over time	2
Differences between users	2
Usage of new features	2

Although **collecting log data from a specific user** entails legal and privacy issues that should first be resolved, the respondents considered this as an opportunity to offer tailored training services. One of the software developers (ID2) suggested that if a user seems to be making repeated systematic errors in the handling of the system, training could be offered for this precise use case.

B. *Business opportunities for logged usage data*

Discussions with the stakeholders revealed five business opportunities supported or enabled by usage data logging.

Continuous user interface development. A systematic approach to collect and analyze usage data from the manufacturing systems used by customers could offer developers and UI designers a direct channel to understand how their products are used from the start to the end of their life-cycle. For example, when new systems are installed in a customer's factory, developers may get very limited or even misleading information about the daily usage patterns:

Our knowledge from the users' way of using this is mainly guesswork. Yes, we see it when the system is being implemented, but there is so much fuss going on that it maybe does not represent routine day and routine use. (ID3)

Major software updates were seen as fruitful instances for collecting log data to inform developers:

There is a good chance for learning, since usually something changes in the update. A) does it work although the bugs should have been noticed before… and B) how the users use the new version, are they taking advantage of the new features at all or were they waste of time? Can the users find them, has the training been successful or were they trained at all? (ID3)

Logged usage data can help developers understand how users navigate in the UI, how features and UI elements are used, and what usage patterns emerge. For example, if users repeatedly use long navigation paths or rarely use UI elements that are on the main view, this can decrease use performance and may require changes in the UI. After identifying interesting usage patterns from log data, the emerging why-questions can be studied with qualitative user research methods such as interviews or user observations.

Improve the quality of customer support services. The customer support team uses log data on a daily basis when inspecting what events led to a problem and what events followed. Especially during complex error events, customer support and developers browse and compare text logs from various machines and related services to get the overall picture of the chain of events. Usage data would provide one more source for this investigation process. An easy way to browse and visualize the users' actions and system events on the same timeline could support the inspection process, resulting in more detailed solutions and instructions for the customer to avoid similar problems in the future.

While end-users may have difficulties in recalling the detailed chains of events and exact time when an error occurred, usage data logs show without doubt when and what really happened. This can save time, especially when preparing for maintenance visits at the customer's site.

New opportunities for customer training offers. New service business opportunities can be found by identifying inefficient use or repeated error situations of the customer's system based on usage data. These findings can be used to support e.g. an offer for a tailored training package for the customer's workshop floor personnel or even individuals if usage data allows recognition on this level:

…we could log the usage to see if the user tries to take many actions that are not allowed, so if this happens often then it could indicate that there is a need for additional training. (ID2)

The director (ID1) proposed that training offers could also be planned based on the most and least used features.

Customer reports that provide additional value. Customer reports are currently manually compiled based on ad hoc requests. The product manager suggested that periodic reporting of the status of the customer's systems might be useful to the customer and that the developed analytics tool could provide data for such reports in the future. Reporting logged usage data over longer time periods, such as months or years, could provide value for customers.

The customer support team leader was intrigued about the possibility of creating automatic reports from the logged usage data that could easily be modified and shared with customers. Such reports should be well prepared summaries, since factory managers usually have little time to spare. Clear visualizations of the data should help in skimming through reports intended for executives.

Evidence for accidents. The respondent from customer support (ID4) noted that log data acts as an evidence in case of any accidents at customer's factory when using the FMS. When discussing the liability for damage, log data may provide evidence of user actions and whether the system was working correctly. For example, log data may show that users had turned off some automated safety functions before the accident. As the price of material damages can be very high, the effective use of log data could result in significant monetary savings for the supplier company.

C. Challenges related to usage data logging

In this section, we present the challenges and obstacles related to the usage data logging in FMS context that at least three of the respondents mentioned during the interviews.

Analytics skills and context knowledge. The main challenge that all respondents brought up was that interpreting logged usage data requires specific knowledge of the system and the context at the customer's factory. As one of the developers commented:

…it requires a lot of interpretation and knowledge of how the system works and is built. I assume that the project manager did not get much out from this. I am so much more familiar with this stuff, so that when I see this sequence here, I see those dialogs and buttons in my mind. (ID2)

If usage data is intended to provide additional value for the customer e.g. in the form of periodical reports, the "raw" usage log data should first be analyzed by the supplier:

Perhaps this could be somewhat useful for customer's people in production and maintenance, but the data should first be well refined: what to do and what are the recommended actions. As such, it is not valuable for them. It must be in more natural language. (ID4)

Missing data types. Practitioners in different roles were interested in some data types that the FMS did not currently log or that were not imported to the UX-sensors tool:

It would be nice to have the error code that could be used for searching. It is not always available in our system. (ID4)

…what device were they using at this moment? It should also be somehow added to this data. However, this is probably not available in the logs. (ID3)

One of the developers had the impression that logging is commonly added late when problems start to emerge or when there is a need to know something.

Data quantity. The sheer amount of available log data distributed across log files from several sources underlined the need for efficient searching and filtering functions:

Going through logs is challenging as there are thousands of lines of data and it can easily take half an hour to even find the correct time. We do compare different logs at the same time. (ID4)

A data analytics tool with easy to use timeline visualizations and recognizable codes for different types of events might support the searching of specific events. Over the time, users might learn to read the visualizations and specific patterns that help pinpoint the beginning of interesting events.

Access to log data. The director noted that agreements with customers should be made regarding the use of log data to be able to utilize the data efficiently and avoid any disputes over access in the future. If the data are used for business purposes, customers should receive in exchange something that provides value for their operations:

…if we use it (data) for doing business then primarily the customer wants something in exchange. It could be, for

example, reports or propositions to develop their operations. Even mobile user interfaces. Things that create value. (ID1)

To summarize, manufacturing companies should create systematic processes for accessing, collecting and transferring log data from the customers' systems to better utilize the data for analysis.

V. Discussion

The identified potential usage metrics to be logged and visualized in our study were in line with the overall data and information needs identified in [20]. When logging user interaction with the system in this context, the most relevant themes to start with seem to be finding specific events and user actions around them. A specific need highlighted by our study was the combination of logs from different services in the same analysis tool, for example by visualizing all events created by human-machine interactions and by different digital services on the same timeline. This would reduce the need for alternating between different log files and searching correct timestamps when following the chain of events. Adopting an architecture similar to DataJewel [4] could reduce the complexity of managing separate datasets originating from different modules of the manufacturing system. On the visualization side, an approach similar to LifeFlow [3] could be adapted to interleave events of interest from multiple systems into a single timeline and allow digging deeper into the data on demand.

The proposed business opportunities in our study reflect the expected benefits identified in [20]. Both studies highlight the log data potential to 1) support product development and customer support (teleservice) activities, and 2) provide novel opportunities to offer customized training to customers. Practitioners in supplier companies expect that log data related to tracking the production and efficiency of system use could provide value for customers (ibid.). In our case study, the proposed medium for communicating these data to the customer was a periodical report that the customer could get in exchange for the data or as an additional service.

One interesting finding in our study was using the log data as an evidence during accidents or dangerous situations, where it is important to inspect if the system has functioned correctly, as this can settle who is responsible for the damage. Utilization of information logging standards [e.g., 14] can support both the supplier and customer in preparing for such situations. As a related concern, visualizations that visually encode the temporal properties of the event sequences (e.g., [2, 3]) could assist in assessing the nature of usage patterns preceding faults; for example, whether the pattern is an isolated issue (e.g., erroneous commands carried out by a user) or a more systematic issue that recurs periodically and could indicate a system level problem.

Knowledge of how the FMS is built was considered as a requirement to get full advantage of the logged usage data. Still, an analytics tool should provide an easy access to the key usage metrics, such as the most and least used features, which interest the majority of stakeholders. As the analytics literacy of the team improves over time, more people might take advantage of such tools in the future [8 10].

As suggested by one respondent, logging capabilities may be implemented afterwards only when need arises. Logging requirements should be discussed during the early stages of the industrial system's development process, as this could potentially solve some of the challenges we identified: accessing logged usage data and lacking the log data of specific features. Reflecting on our experiences and the identified challenges during this study, we propose stakeholders in manufacturing companies consider the following questions before embarking on usage data logging:

- **Possibilities**: What type of usage data and related data from the system and context can be logged?
- **Goals**: What do we want to learn from these data? What value can these data provide and for whom?
- **Data analysis**: Who has the requisite skills and context knowledge to analyze the log data?
- **Data access and security**: How can the log data be accessed? Who owns the data? How will the data be transferred and stored? How do we ensure security?
- **Tools and data wrangling**: Which data analytics and visualization tools are suitable for the needs of different stakeholders? How much data wrangling is required to import the raw log data to these tools?

While not an exhaustive list of all questions regarding data logging, we argue that discussing these topics with stakeholders representing different roles in the company will ensure that their needs are acknowledged, leading to the more versatile utilization of logged usage data. Stakeholders from management, product development, design, customer support and marketing should be consulted, as well as customer's representatives, before deciding which data to capture. A systematic process in place for effortlessly logging and importing usage data into an easy to use analytics tool should also motivate those stakeholders who are less familiar with analytics. The proposed questions can serve practitioners working also in other contexts, including different web-based services, mobile systems or industrial systems that enable the logging of end-user interactions.

Limitations and future work. Our case study results reflect only the views of a small sample of different stakeholder roles. Future research should inspect the views from marketing, sales, user training and customers. Another interesting question relates to data ownership and the value proposition for customers in sharing logged usage data with the supplier. Understanding how end-users react towards the logging of their actions will be critical for the viability of data analytics in the industrial context.

VI. Conclusions

In this paper, we present potential usage metrics and identify business opportunities and challenges for usage data logging in the context of flexible manufacturing systems (FMS). We propose a set of questions to support practitioners planning to utilize usage data logging in their product development activities.

Data were gathered during a six-month collaborative development project of a visual analytics tool prototype with six practitioners from a FMS supplier company. We identified a need for combining usage data and system event logs in the same visualization timeline to support the analysis of error situations. Logged usage data were seen beneficial in user interface development, solving error situations and as evidence in accidents. User training offers and periodic reporting based on usage data could provide value for customers. Required analytics skills, data quantity, and accessing data owned by customers were the key challenges in utilizing log data. In future, empirical studies regarding the realization of the opportunities or ways to overcome the identified challenges would be valuable for practitioners in FMS and related industrial contexts.

ACKNOWLEDGMENT

This study was supported by The Finnish Funding Agency for Innovation (TEKES) as a part of User Experience and Usability in Complex Systems (UXUS) research programme.

REFERENCES

[1] Bosch Software Innovations, https://www.bosch-si.com/solutions/manufacturing/data-analytics/manufacturing-analytics.html

[2] J. A. Fails, A. Karslon, L. Shahamat, and B. Shneiderman, "A visual interface for multivariate temporal data: Finding patterns of events across multiple histories," Proc. IEEE Symp. On Visual Analytics Science And Technology (VAST), IEEE Computer Society, 2006, pp. 167-174.

[3] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman, "LifeFlow: visualizing an overview of event sequences," Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 11), ACM, May 2011, pp. 1747-1756, doi:10.1145/1978942.1979196.

[4] M. Ankerst, A. Kao, R. Tjoelker, and C. Wang, "DataJewel: Integrating visualization with temporal data mining," in Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, S. Simoff, M. Böhlen and A. Mazeika, Eds. LNCS 4404, 2008, pp. 312-330.

[5] A. Perer and E. Wang, "Frequence: interactive mining and visualization of temporal frequent event sequences," Proc. 19th international conference on Intelligent User Interfaces (IUI 14), ACM, Feb. 2014, pp. 153-162, doi:10.1145/2557500.2557508.

[6] C. Holzmann, F. Lettner, C. Grossauer, W. Wetzlinger, P. Latzelsperger, and C. Augdopler, "Logging and Visualization of Touch Interactions on Teach Pendants," Proc. 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI 14), ACM, Sept. 2014, pp. 619-624, doi:10.1145/2628363.2628429.

[7] C. Grossauer, C. Holzmann, D. Steiner, and A. Guetz, "Interaction visualization and analysis in automation industry," Proc. 14th International Conference on Mobile and Ubiquitous Multimedia (MUM 15), ACM, 2015, pp. 407-411, doi:10.1145/2836041.2841218.

[8] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer, "Enterprise data analysis and visualization: An interview study," IEEE Transactions on Visualization and Computer Graphics, vol. 18, Dec. 2012, pp. 2917-2926, doi:10.1109/TVCG.2012.219.

[9] B. Shneiderman and C. Plaisant, (2006). "Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies," Proc. 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization, ACM, May 2006, pp. 1-7, doi:10.1145/1168149.1168158.

[10] B. Medler, M. John, and J. Lane, (2011). "Data cracker: developing a visual game analytic tool for analyzing online gameplay," Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 11), ACM, May 2011, pp. 2365-2374, doi:10.1145/1978942.1979288.

[11] J. Heer and S. Kandel, "Interactive analysis of big data," XRDS, vol. 19, Sept. 2012, pp. 50-54, doi:10.1145/2331042.2331058.

[12] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," Interactions, vol. 19, May 2012, pp. 50-59, doi:10.1145/2168931.2168943.

[13] Q. Liu, M. Vorvoreanu, K. P. C. Madhavan, and A. F. McKenna, "Designing discovery experience for big data interaction: a case of web-based knowledge mining and interactive visualization platform," Proc. Second international conference on Design, User Experience, and Usability: web, mobile, and product design - Volume Part IV (DUXU 13), A. Marcus, Ed. Springer-Verlag, vol. Part IV, July 2013, pp. 543-552, doi:10.1007/978-3-642-39253-5_60.

[14] SANS Information Logging Standard, https://www.sans.org/security-resources/policies/server-security/pdf/information-logging-standard

[15] OWASP Foundation, https://www.owasp.org/index.php/Logging_Cheat_Sheet

[16] S. Bateman, C. Gutwin, N. Osgood, and G. McCalla, "Interactive usability instrumentation," Proc. 1st ACM SIGCHI Symp. on Engineering interactive computing systems (EICS 09), ACM Press, 2009, pp. 45-54, doi:10.1145/1570433.1570443.

[17] R. Atterer and A. Schmidt, "Tracking the interaction of users with AJAX applications for usability testing," Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 07), ACM, 2007, pp. 1347-1350, doi:10.1145/1240624.1240828.

[18] D. M. Hilbert and D. F. Redmiles, "Extracting usability information from user interface events," ACM Computing Surveys, vol. 32, Dec. 2000, pp. 384-421, doi:10.1145/371578.371593.

[19] P. Ketola and V. Roto, "On User Experience Measurement Needs - Case Nokia," International Journal on Technology and Human Interaction (IJTHI), vol. 5, 2009, pp.78-89, doi:10.4018/jthi.2009070104.

[20] H. Väätäjä and 9 others, "Opportunities and needs for logged usage data analytics of complex industrial systems," Proc. Twenty-Fourth European Conference on Information Systems (ECIS), AIS, 2016.

[21] T. Sheridan, Human and automation: system design and research issues. New York, NY: John Wiley & Sons, Inc., July 2002.

[22] H. A. ElMaraghy, "Flexible and reconfigurable manufacturing systems paradigms," Int J Flex Manuf Sys, vol. 17, Oct. 2006, pp. 261-276, doi:10.1007/s10696-006-9028-7

P6

GUIDELINES FOR DEVELOPMENT AND EVALUATION OF USAGE DATA ANALYTICS TOOLS FOR HUMAN-MACHINE INTERACTIONS WITH INDUSTRIAL MANUFACTURING SYSTEMS

by

Varsaluoma, J., Väätäjä, H., Heimonen, T., Tiitinen, K., Hakulinen, J., Turunen, M. and Nieminen, H.

Accepted for publication in Proceedings of the 22th International Academic MindTrek Conference (Academic MindTrek'18). ACM, New York.

Guidelines for Development and Evaluation of Usage Data Analytics Tools for Human-Machine Interactions with Industrial Manufacturing Systems

Jari Varsaluoma,
Heli Väätäjä
Tampere University of Technology
Tampere, Finland
first.lastname@tut.fi

Tomi Heimonen
University of Wisconsin-Stevens Point
Stevens Point, Wisconsin, USA
Tomi.Heimonen@uwsp.edu

Katariina Tiitinen, Jaakko Hakulinen, Markku Turunen
University of Tampere
Tampere, Finland
first.lastname@uta.fi

Harri Nieminen
Fastems Ltd
Tampere, Finland

ABSTRACT

We present the lessons learned during the development and evaluation process for UX-sensors, a visual data analytics tool for inspecting logged usage data from flexible manufacturing systems (FMS). Based on the experiences during a collaborative development process with practitioners from one FMS supplier company, we propose guidelines to support other developers of visual data analytics tools for usage data logging in context of complex industrial systems. For instance, involving stakeholders with different roles can help to identify user requirements and generate valuable development ideas. Tool developers should confirm early access to real usage data from customers' systems and familiarize themselves with the log data structure. We argue that combining expert evaluations with field study methods can provide a more diverse set of usability issues to address. For future research, we encourage studies on insights emerging from usage data analytics and their impact on the viewpoints of the supplier and customer.

CCS CONCEPTS

• **Human-centered computing → Visualization design and evaluation methods**; **Empirical studies in HCI**; Field studies

KEYWORDS

Usage data logging; visual data analytics; human-machine interaction; flexible manufacturing systems; guidelines; lessons learned.

ACM Reference format:

J. Varsaluoma, H. Väätäjä, T. Heimonen, K. Tiitinen, J. Hakulinen, M. Turunen, and H. Nieminen. 2018. Guidelines for Development and Evaluation of Usage Data Analytics Tools for Human-Machine Interactions with Industrial Manufacturing Systems. In *Proceedings of the 22th International Academic MindTrek Conference, Tampere, Finland*, October 2018 (Academic Mindtrek'18). 10 pages. https://doi.org/10.1145/3275116.3275138

ACM ISBN 978-1-4503-6589-5/18/10…$15.00
https://doi.org/10.1145/3275116.3275138

1 INTRODUCTION

Visual data analytics tools that are easy to use and support the goals of their users can help practitioners working in companies when inspecting collected usage data. However, a common challenge in studying visualization tools is how to evaluate the tools and their effectiveness [1]. Several studies have emphasized that when developing visualization tools, they should be evaluated in real work context with actual end users, instead of short-term laboratory experiments [e.g., 1, 2, 3]. The review by Isenberg et al. [4] shows that the number of the published evaluation studies of information visualization tools done in real use context has increased. However, the number of studies related to the UX of visualization tools was considered surprisingly low.

This work is motivated by the call for more empirical research aiming to understand and design for the user experience (UX) of visualization tools [4]. Furthermore, little empirical research is available regarding the development of visual data analytics tools together with users in the context of industrial manufacturing systems. In practical terms, our research aimed to support a manufacturing automation systems company by developing and evaluating a prototype tool that enables the systematic use of the logged usage data to support product and service development and innovation activities through the gained insights. We describe a case study with an industrial manufacturing company, during which we developed UX-sensors, a visual data analytics tool for inspecting usage data. This work presents the tool itself, but focuses on the evaluation process and lessons learned during its development. Guidelines are provided to support the collaborative development process of similar visual data analytics tools for logged usage data, derived from our findings. In addition, we present our development process and discuss how it could be extended in future studies where visual data analytics tools are developed together with practitioners from companies.

In the following, we first review related work regarding information visualization tool evaluation approaches with users and usage data analytics in related industrial contexts. Then, we provide an overview of the study context and describe the UX-sensors tool. Next, we describe the iterative development and evaluation process for the UX-sensors tool, the used evaluation methods, and the relevant findings. Following, we present the proposed guidelines, discuss our study in contrast to previous research, and conclude with the limitations of the current study and topics for future research.

2 RELATED WORK

2.1 Information Visualization Tool Development and Evaluation with Users

Various approaches and guidelines have been proposed that can support the development and evaluation process of information visualization tools with users. An overview of relevant evaluation methods for visualizations is presented in [2], while advice to when to use which method is provided in [5]. The use of qualitative evaluation methods such as observations and interviews can help achieving a richer understanding of the factors that influence visualization development and usage [2, 6]. Carpendale [2] encouraged that more studies in the information visualization field should utilize such methods. However, as in our case, field studies can take advantage of both qualitative and quantitative methods for information visualization evaluation [1].

Lam et al. [7] state that visualization evaluation approach should be based on evaluation goals and questions instead of methods. They provide seven types of evaluation scenarios in the information visualization domain, based on the overview of 850 papers from information visualization literature. Sedlmair et al. [8] propose a nine-stage design study methodology (DSM) and practical guidance for designing visualization systems in collaboration with domain experts. Based on their own experiences and literature review in the fields of human-computer interaction (HCI) and social science, they summarize 32 design study pitfalls to guide the whole process from learning and designing to reporting design studies. Recently, Crisan et al. [9] proposed additions to DSM and practical guidelines for evaluating external constraints, regulatory and organizational, that can affect visualization evaluation with companies.

Several studies report experiences from visualization evaluation in specific context. For instance, Sedlmair et al. [10] list challenges and recommendations for information visualization evaluation based on their experiences from a variety of studies in a large company setting. Saraiya et al. [11] report a long-term study with bioinformaticians to analyze how visualizations are used to gain insights into the data. They emphasize 1) the users' natural motivation to do data analysis and 2) the evaluation of the significance of insights as two essential reasons to evaluate long-term visualization tool usage in a real-world setting. Longitudinal studies enable the inspection of long-term insight generation process and identifying long-term usability problems with data visualization tools [11]. Medler et al. [3] presented their insights from the development of Data Cracker, a visual game analytics tool for supporting game designers. Authors argue that it is beneficial to develop analytic tools in parallel with product development and produce visual prototypes to help the team understand how the tool could be beneficial for them [3]. Additionally, it may be necessary to create functionality for addressing legal issues, such as privacy controls. The team integration insights include the necessity of involving interdisciplinary teams and the possibility of encountering prejudices towards analytic tools from the product developers. Finally, the role of communication is important in order to anticipate how product changes affect the interpretation of data and to update the developers on the progress of tool development [3].

We applied the Multi-dimensional In-depth Long-term Case Study (MILC) [1] approach suggested for the studies of creativity support tools in [12]. MILC was used to guide the visual data analytics tool development in a long-term study in real use context. Shneiderman and Plaisant [1] propose the use of MILCs to study the efficacy of novel visualization tools not only in terms of their strengths but also to refine the tool iteratively with end users and to produce sufficient evidence to warrant further development. The MILC approach and its derivatives [13] have been used to develop visualization tools for event sequence analysis [14] and in the evaluation of a visual analytics tool in the domain of electronic medical records analysis [15]. MILC was also identified as a relevant approach for the long-term analysis of domain expert use of visual analytics [16]. The evaluations by Wongsuphasawat and colleagues [14] show that the periodic meetings with a domain expert allowed for both the generation of insights and additional questions by the domain expert and guidance for tool development. The study by Stolper and colleagues [15] demonstrated the benefit of the case study approach in documenting insights generated during the long-term use of a visual analytics tool.

2.2 Usage Data Analytics in Manufacturing and Automation Industry

Data analytics and visualization solutions exist on the market for companies to use on their own manufacturing data, such as Bosch's manufacturing analytics solutions [17]. However, little previous research exists in the domain of exploratory user interaction analysis of complex industrial systems, particularly regarding the development of usage data analytics tools. Where many consumer applications are reasonably simple in their operating logic, manufacturing systems have a large number of processes and rules that govern the functioning of the whole. Unlike many consumer-oriented systems, the data that are collected from industrial applications can have significant business value to its producer (i.e., the clients of the system supplier), which puts the onus on developing tools that can generate added value for all stakeholders.

Holzmann et al. [18] studied the acquisition and visualization of user interaction data from a touch screen based robot controller to find cost-efficient solutions for the usability evaluation of handheld terminals in the automation industry. The goal was to help developers to identify possible problems in users' workflow (e.g., navigation problems or unused functions) based on user interface interactions. Navigation path analysis and usage intensity were identified as the most important topics for data logging, based on interviews with a programmer and two project managers in automation industry enterprises.

In another example, Grossauer et al. [19] created a prototype for automation industry to visualize navigation flows through an application. Based on their experiences with applying the visualization tool to multiple datasets, they recommend such tools should include 1) a wide variety of filters and 2) views that show the whole navigation data and allow the inspection of individual sequences.

We need to learn more about the benefits and challenges related to usage data analytics in manufacturing automation and related industrial contexts. Our study adds to the body of knowledge in this domain with new empirical research done in the context of flexible manufacturing systems.

3 STUDY CONTEXT: FLEXIBLE MANUFACTURING SYSTEMS

Our study was conducted in the context of industrial manufacturing automation systems called flexible manufacturing systems (FMS). Metal industry uses FMS for manufacturing parts by using different metal operations, such as cutting operations (e.g. milling, drilling and boring), metal-forming operations (e.g., rolling, stamping, and welding), and surface-finish operations (e.g., grinding and painting) [20]. In FMS, the production is typically conveyed on pallets [21], on which the parts are attached for machining. FMS enables the automation of the pallet-based machining for manufacturing small batches of different types of products while providing flexibility, as the manufactured product can be changed without changing the whole factory layout [20]. FMS combines software and hardware in order to provide a manufacturing company an easily modifiable, dynamic manufacturing system. Hardware is controlled by software, which in our study manages the production and provides different production optimizations tools, such as fine scheduling.

FMS is normally operated by human operators although robots are also used in some factories. Today, FMS is operated via a combination of graphical user interfaces (e.g., used to enter new manufacturing programs, control the program parameters and modify system status) and physical or software coded buttons on the user interface for pallet control. In our study, the main focus is on user interactions by the human operators on the workshop floor with the FMS elements of the manufacturing system.

The company participating in this study was interested in collecting and analyzing the usage data of their FMS systems after they had been supplied to the customers. Usage data was expected to benefit the company's R&D, customer support, and service business in the future. While the FMS already logged the data of their behavior, it was mainly used for on-demand maintenance. The usage data portraying users' interactions with the system provided an entirely new channel to study the product usage.

4 UX-SENSORS – THE USAGE DATA ANALYTICS FRAMEWORK

In this section, we describe the usage data analytics framework of the UX-sensors tool at the end of the collaborative development period with the FMS company. The framework consists of three components: the data model, which provides an abstraction of case-specific log data; the analytics software framework, which facilitates the storage and analysis of the logged usage data; and the analytics user interface, which allows the interactive exploration of the dataset.

4.1 Data Model

The data model utilized by the framework is based on typical events recorded of human-machine interactions, such as button clicks, data entry, and view changes. The fundamental item of the data model is an event consisting of a timestamp, a feature and a value attributes. Additionally, events can have parameters and context information. Finally, each event has a level specifying whether it is a regular occurrence, a note added via the analysis tool or an error of some level.

The model aims to be generic so that data from different processes and user interfaces can be analyzed. Most of the existing log files can be converted into events in a straightforward manner. In the company data used in our development and evaluations of the tool, a feature refers to parts of the user interface and value to the action executed. Parameters encode additional operation related parameters and context tells about the identity and generic state of the system and user interface at the time.

4.2 Software Architecture

The software architecture of the analytics framework consists of an interactive web application and a set of server components (see Fig. 1). The framework includes a data-import tool for uploading log files to the system, visualization front-end for exploring log data, and server-side components for importing and analyzing the usage log data. The log data and configuration information is stored in CouchDB [22], a NoSQL database that provides an efficient model to handle and query the subsets of massive log data sets. CouchDB is a document storage type database, where each data item, in our case each event, is a JSON document. CouchDB views and lists are used to query the data e.g. by the factory. In addition, a configuration document is used to define tool instance specific properties, such as additional filters, features of interest, user interface structure settings, and color-coding rules for events.

All requests from the web application front end are directed through a proxy server to appropriate back-end components. The proxy takes care of security-related aspects and provides a single server address and port for the web application. It also provides access to a logging service, which stores log data from the web application in a format directly compatible with the tool itself. This logging was used during evaluations.

The server side is mainly built on Node.js [23]. Python is used for the analysis modules and data import. In addition, the analysis modules use R statistical computing framework for extracting frequently occurring sequences from the event data. The front end is built on the Bootstrap front-end framework [24]. Visualizations are built using D3.js [25], a JavaScript library, which allows binding data to the DOM of an HTML document. Crossfilter [26] is utilized for filtering the event data.

The tool was deployed on an external server separate from the customers' flexible manufacturing systems. Log data was uploaded to the analytics tool only by on-demand basis. In the future, it could be beneficial to directly connect the tool to the logging components of the customers' systems.

4.3 UX-Sensors User Interface

The user front end of the developed data analytics tool is an interactive web application. It consists of a data selection view

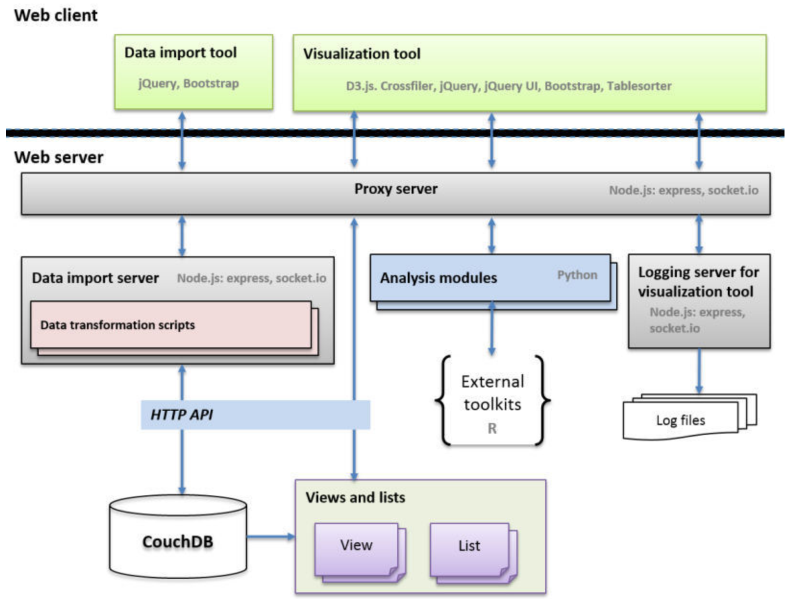

Figure 1: Software architecture of the developed UX-sensors tool.

and the main data browsing view with timelines and analysis tools. The main data browsing view consists of six main elements that are highlighted and numbered in Fig. 2. The elements are: 1) overview panel, 2) overview timeline, 3) detail timeline, 4) additional filters, 5) tabs, and 6) the main filters.

The overview panel contains general information about the selected observation window, e.g. the number of found events, most frequent events by value, the average usage session length, the average number of operations per session, and the number of error events. Events are split into sessions based on the maximum time duration between successive events. The default value is five minutes and it can be changed in settings. The overview timeline displays the overall number of events by the hour and works as a filter where the user can restrict the further analysis to a shorter observation window. The detailed timeline displays the individual events within the observation window. By hovering over an event, detailed information is displayed. The user can also add additional notes directly to the timeline, for example, to record and share insights gained from the data. The additional filters element can filter the event set further. In the setup used in the field study, six main filters were available: factory, observation window (if multiple windows selected), system, user, level, and feature.

The tab elements display processed information about the selected events and provide tools for further exploration of the data. The used features tab displays a line graph of feature use over time and a complete list of features and feature-value pairs with a count and percentage of the total events. A feature represents a part of the FMS software and a value the action user has performed, for example, "Inventory - Release Pallet". The data table can be sorted by each column and filtered by search. The errors and recovery tab provides a list of the most common errors, average recovery time and user interaction sequences during error recovery. Error recovery time is estimated as the duration between the last successive error event and the first following user interaction (i.e., event that is not an error or warning). The frequent sequences tab is used to calculate and display the most frequently occurring sequences. This is done by splitting the events into sessions and then looking for similarities in the event sequences within the sessions. Through the search tab, single events or sequences of events can be searched by defining key-value pairs consisting of e.g., feature and value. The data entry tab is for exploring events indicating user data entry and user interaction sequences during data entry. Lastly, the main filters element can contain up to two filter panels on the right side of the timelines.

5 UX-SENSORS – DEVELOPMENT AND EVALUATION

In this section, the process for developing and evaluating the UX-sensors tool in collaboration with stakeholders from the FMS supplier company is presented. Then the used evaluation methods and their relevant results are described.

5.1 UX-Sensors Development and Evaluation Process

Fig. 3 illustrates the development process for UX-sensors, consisting of requirements gathering phase and iterative development and evaluation phase. Next, the development process is described in more detail.

Workshop to identify company needs. As a part of a larger academic research project with three companies from metals and engineering industry, we held a workshop to identify the company needs on usage data analytics and visualization in this domain. The identified requirements for data types were: 1) usage combinations, such as the customer's production type and in what mode they use the system, 2) patterns of use, 3) types of user groups and profiles that can be found, 4) summarizations of the system use based on logged data (the logs of events, system status, user actions, interactions etc.), 5) identifying problem or fault situations (individual and possible patterns), and 6) changes in use (such as features) over weeks or months such as how the taking of the system into use and learning to use is progressing,

Figure 2: The main data browsing and analysis view of the UX-sensors tool. The elements are: 1) overview panel, 2) overview timeline, 3) detail timeline, 4) additional filters, 5) tabs, and 6) main filters

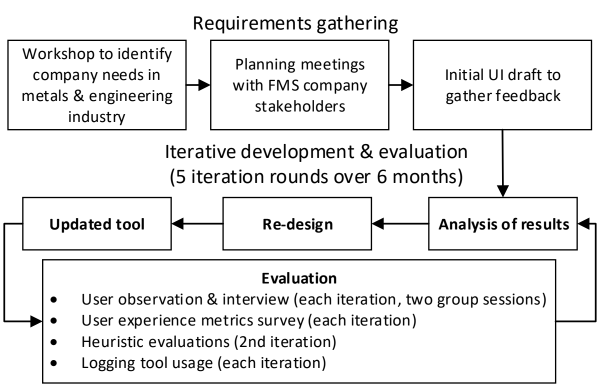

Figure 3: Summary of the development process for the UX-sensors tool.

to identify issues needing support or training and whether problems or faults appear over time.

Planning meetings with FMS company. After the workshop, planning continued with stakeholders from one flexible manufacturing systems (FMS) supplier company on the development of a usage data analytics tool. In discussions with stakeholders from R&D, software development and customer support, requirements for the data analytics tool and its features were gathered. The same people participated in the iterative development and evaluation process (described later) of the tool. Instead of utilizing an off-the-shelf data analytics tool, we decided to develop our own visual analytics framework. Given the varying needs of the stakeholder companies, a custom framework was expected to expedite the development process, over learning and customizing an existing tool, and provide the development team experience in the design and development of visual analytics tools for supporting other projects.

Initial UI draft. Next, an initial user interface draft of UX-sensors was presented to stakeholders to spark more conversations and to gather feedback on the proposed UI design. This feedback was utilized in designing the first interactive version of the tool to be used in the iterative development and evaluation cycles.

Iterative development and evaluation. During the following six-month study period, including a one month holiday season, we iteratively developed and evaluated the tool in collaboration with practitioners from the FMS supplier company. Two of the authors were responsible for the user studies and reporting their findings to three researchers responsible for the software development of the UX-sensors tool.

Our evaluation approach to UX-sensors was based on the MILC approach [1], which has inspired the implementation of several long-term studies where the use of data analytics and visualization tools have been evaluated in real use context [13, 14, 15, 16]. The MILC approach requires, at a minimum, 3-5 domain experts who are available for a period of weeks to months, and a tool that provides sufficient, problem-free basic functionality for users [1]. The long-term process requires the documentation of current tools and practices, establishing evaluation criteria, a schedule of user research, instrumenting the tool for collecting log data, providing training, and modifying the tool as needed [1].

Six practitioners participated in the evaluations: three software developers (including a team leader), a leader of a remote customer support team, a product manager for product life cycle services and a director managing research and innovation development. One developer working for a subcontractor left the study after the second iteration round. From these participants, the developers and the leader of the customer support team were identified as the lead users of the UX-sensors tool as they had experience of inspecting log data from their customers' FMS when challenging error events occurred. However, including the product manager and the director in the development process was expected to generate more diverse set of ideas for utilization of the tool and logged usage data.

The iterative development and evaluation phase started when the first interactive prototype of the UX-sensors tool was introduced to the company personnel in a training workshop. In the workshop, all six participants could inspect a usage data set with the tool and give feedback from the user interface. During the following six months, four more iterations of the tool were introduced to the participants. User feedback was collected after each iteration of the UX-sensors tool. Email reminders were sent after each update to encourage participants to try out the tool. We organized two group meetings, including the first training workshop, and three sets of single user observation sessions. After each meeting, a link to a web survey was sent to the participants in an email. After the second iteration, five external evaluators conducted heuristic evaluations of the prototype. Finally, log data was collected from the UX-sensors tool for following its usage over the whole development period. The next section summarizes the used methods and the main findings.

At the end of the development process, a data import tool was implemented to allow company practitioners to import usage data logs to the UX-sensors tool. We anticipated that practitioners would use UX-sensors tool to inspect logged usage data in the near future. However, when inquired after six months, we learned that the stakeholders still worked on challenges related to the legal issues considering the data ownership, privacy and security. When collecting data from customers' employees working with the system in different countries, the supplier has to carefully follow the local data collection policies and make agreements with each customer regarding data usage.

5.2 Used Evaluation Methods and Main Findings

User observation and interviews. In the observation sessions, participants could freely use the tool for exploring the available usage data and try any new features, while the researcher encouraged the participant to think aloud with questions such as "what do you think of this feature?" The session ended with an interview, where researchers could ask feedback from specific features, confirm their observations during the session, and inquire if participants had received any insights from the data. Each session was recorded with a video camera and lasted approximately one hour. All sessions were arranged in the company's meeting rooms. The observation and interview data were analyzed by coding similar findings or responses to descriptive categories, with comments related to the development needs of the data analytics tool separated from other topics. The comments were grouped based on the features or aspects of the tool that they referred and then reported to the analytics tool developers.

The observation sessions provided information regarding usability issues, suggestions for new features and insights that participants got from the usage data. For example, the customer support representative asked for adding references to the system generated error codes in the log data events and support for exporting the tables or lists of the analyzed data for modifying the data with other tools for creating reports. Developers were interested in acquiring more details regarding logged error events such as a direct reference to the line in the code. One of the developers also proposed how the future version of the tool could function for importing log data files collected from different customers. The concept of event sequences was challenging for most participants and therefore tooltip help texts and an explanation of how the sequences are calculated were added to the UI. Furthermore, the content of the original sequences tab was divided into frequent sequences tab and search tab to clarify the UI.

One insight from the usage data that generated much discussion among the participants related to how the autopilot feature in the FMS was operated. The usage data proposed that users did not follow the shortest route in the UI to activate the feature, suggesting a need for changes in the UI and/or in the user training process. Over the following interviews, we learned that developers had considered whether certain actions should not require that the autopilot is activated, as changing its state very often requires user effort. Therefore, this insight from the usage data may result in some changes in the UI in the future.

As a general observation, we learned that developers rarely had possibilities for gathering feedback on how the FMS are used on a daily basis after they have been installed in the customer's factory. Logged usage data was seen as a relevant channel for supporting developers' understanding about the end-users and their ways of using the system, especially over longer periods. Preliminary knowledge of the customer's ways of using the system and the common errors could significantly help focusing site visits on customer's factory, where support can be provided and more qualitative data gathered to understand the reasons behind the findings from log data.

User experience metrics survey. Repeated web surveys aimed to capture whether the user experience (UX) with the tool changed over time. Aiming to make the repeated evaluations less taxing for the participants, we limited the number of different measured UX factors and used only a 4-point Likert scale. Table 1 presents the questions and results from the first and the last (fifth) UX survey, including three questions adapted from the After Scenario Questionnaire (ASQ) by Lewis [27]. The director and one developer did not answer the last survey, hence only the

Table 1: Evaluations of the first and fifth iteration of the UX-sensors tool (n=4). Scale from 1 to 4, where 1 = Completely disagree, 2 = Somewhat disagree, 3 = Somewhat agree and 4 = Completely agree.

This tool is…	Easy to learn		Useable		Flexible in its interaction		Pleasant to use		Useful	
SURVEY	1st	5th	1st	5th	1st	5th	1st	5th	1st	5th
MEAN	3	2,75	2,75	2,25	2,75	2,75	2,25	2,25	3	2,50
SD	0,82	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,82	0,58

ASQ: I am satisfied with the tool's…	Ease of use		Amount of time it takes to complete tasks		Support information (help, documentation) during usage	
SURVEY	1st	5th	1st	5th	1st	5th
MEAN	2,25	2,50	2,50	2,50	2,50	3,00
SD	0,50	0,58	0,58	0,58	0,58	0,00

other four respondents are included (n=4). Interestingly, while the respondents got more satisfied with the ease of use and support information with the tool, they considered it less useful and harder to learn. One possible reason for this is that each iteration, also the fifth, presented UI changes or new features, and participants had to familiarize themselves with these new features. Also, the novelty value of the tool may have decreased over time.

Heuristic evaluations (HE). We utilized the top ten heuristics for information visualization with the widest explanatory coverage proposed by the study by Forssell et al. [28]. The HE tasks consisted of 1) exploring the 2nd iteration of the UX-sensors tool and its features, 2) identifying usability issues and describing them in free text, 3) identifying the heuristics that were violated, 4) assessing the severity rating of the finding based on Nielsen's rating scale [29], and 5) assessment of how well the heuristic explained the finding [30].

Five external evaluators (three female) took part in the HEs. Evaluators' experience in the field of HCI (either studies or research work) varied from 1 month to 2.5 years. Two had no work experience in the field of data visualization while others' experience varied from 3 months to 2.5 years. None of the evaluators had experience in the application domain of the FMS. Three evaluators had previous experience from expert evaluations of interactive software.

The HEs resulted in 99 different problems or suggestions for improvements that at least one of the evaluators reported. These findings were reported to the developers of the UX-sensors tool and used in updating the tool for the following iterations. "Information coding" (30 references) and "orientation and help" (22) were two of the most often violated heuristics. Since the evaluators were not familiar what the actual usage data represented, they focused on the UI and visualization related issues. For this purpose, the information visualization heuristics [28] seemed to be well-suited, as several comments related to the used graphs and charts, such as color coding, axis information, and zoom functions. From the observation sessions done during the same iteration, we identified 68 different problems or suggestions for improvements in total. Interestingly, only 14 problems were identified with both methods, for example, lack of help texts for features and unfamiliar terms, color coding issues, not listing events in a table format, and saving the previously conducted searches. In contrast to the HEs, the findings from the observation sessions reflected the requirements that the employees had for doing their work tasks, including specific types of data, features, and visualizations.

Logging tool usage. Log data from UX-sensors was used for following how actively participants used the tool between the observation sessions. This prodded us to discuss with the less active participants what could motivate them to utilize the tool more often. While access to more real usage data from different customers' systems was hoped, it was also evident that the learning curve was steeper for those participants who were not accustomed to working with "raw" log data from FMS.

In conclusion, considering all the methods we utilized, user observations, interviews and heuristic evaluations provided the most useful feedback for improving the tool. Repeated survey questions provided feedback of the tool's UX over time, but discussions with the participants resulted in more practical insights regarding how the tool was used. Finally, logging the usage of the UX-sensors tool itself was an easy way in inspecting how the tool was used outside the observation sessions.

6 GUIDELINES FOR DEVELOPING AND EVALUATING USAGE DATA ANALYTICS TOOLS

In the following, nine guidelines are presented based on the insights during the development of the UX-sensors tool in collaboration with practitioners from the FMS supplier company.

1. Gather an Interdisciplinary Team to Support the Development Process. We confirm the experiences from other domains [3] in that an interdisciplinary team can greatly support the development process of the analytics tool in the context of automated manufacturing industry. The company employees who participated in the development project had different analytic needs and requirements regarding the collected usage data. For example, developers and customer support personnel were more interested in details related to specific error situations, while manager-level personnel often discussed the more general usage data types of their customers' systems. We presume that including stakeholders from marketing, sales, and user training as well as customers' representatives could provide even more insights from the possibilities and challenges of usage data logging, as these roles came up in discussions with the current participants.

Although gathering representatives from various areas in the company requires effort, it can pay off in the ideas of new features for the developed tool or new ways to utilize the gathered log data to provide value for the company and the customer. Furthermore, participating in the development and evaluation of the data analytics tool can benefit the collaborating company by improving the basic data literacy skills of the employees [3]. For instance, during our group discussion sessions, the participants became more aware of the analytics needs of their colleagues and the way the tool should be developed was discussed based on these user requirements.

2. Ensure Early Access to Real Logged Usage Data. One of the key issues in the development process was acquiring sample log files to support the testing and demonstration of the tool. Data gathering is prone to delays that may jeopardize the whole project [8]. One option is to use synthetic data that at least allows the inspection of the functionality of the analytics tool and concrete discussions with stakeholders while tool developers are waiting for access to real data [9]. In our case, we utilized data from a local FMS training environment. Since customer data is usually confidential, getting access to the log files that are gathered from customers located in different countries requires familiarization with the local rules regarding data logging and usage. However, we learned that these challenges generated beneficial discussions among stakeholders, such as how to provide value for the customer in exchange for the data collected from their factory.

We emphasize that access to real logged usage data should be secured as early as possible in the analytics tool development process. Resources should be allocated to building trust and showing the value that the customer can get from sharing the logged usage data with the supplier. This could include reports on how customer's different teams use the system over time and suggestions for additional user training. However, privacy, security, and intellectual property issues need to be solved [31] before accessing usage data. When access to real usage data is negotiated in advance, the developed tool can be tested more efficiently and any disputes over the data access are minimized in the future. Finally, as the goal is to generate new insights from the data, it is vital that the data are meaningful to the users inspecting them. Lack of interesting data can affect stakeholders' motivation to explore the tool on their own and participate in the evaluation activities.

3. Identify Other Data Types That Can Support Usage Data Analytics. During the requirements gathering process, analytics tool developers should identify what other contextual data could support users in analyzing logged usage data and consider whether these data can be visualized with the same tool. For example, we learned that the developers and remote customer support personnel occasionally had to inspect several log files when sourcing error situations between different log data sources. We, therefore, identified the need for viewing human-machine interaction events and events generated by different digital system services on the same timeline to support the sourcing of error events. Although not implemented in the current version of the UX-sensors, this feature could improve the current process of searching correct timestamps and manually switching between different text log files when following the chain of events.

4. Allocate Resources to Explore the Log Data Structure Prior to Data Wrangling. An important aspect of the development process is clearly communicating the structure of the logged data if different teams work on the logging services and the analytics tool development. The best-case scenario would be working together with the programmers of the automation system and agree on what should be logged after deciding which data is needed. In our case, the analytics tool developers were not familiar with the logging services. It took us considerable effort to understand the structure and meaning of the log data and map it to meaningful labels and functionality in the visualization. The familiarization process required close collaboration with an FMS development team member familiar with the logging procedure.

We also recommend mapping out potential 'edge cases' (e.g., log file types or log entries that differ in formatting from others) to avoid unnecessary troubleshooting. Moreover, subsequent changes to logging services should be made in a way that does not change the log format, to avoid additional work on data wrangling. If changes are unavoidable, care should be taken to work with the analytics tool developers to limit the scope of required changes.

5. Establish Coverage of Logging and Compatibility with the Visualization Tool. With complex industrial systems, there can be multiple ways to log events at the system and organization level, and such log files can be stored in different locations. In our case, we utilized log data from one part of the FMS system, which was not enough to implement all the planned features for the visualization tool. For example, it was discussed that teleservice log files should be incorporated into the data visualization tool, but these files were not made available during the study. Thus, in addition to understanding what the log files contain (guideline 4), it is important to establish which log files are required to fully address the design requirements.

Customers' manufacturing systems that are older may log data differently. This means that not all desired log data may be available from all systems and in the same format. Therefore, support for all different kinds of log events may be difficult to implement and it should be decided what kind of logging the developed visualization tool should primarily support.

6. Combine Expert Evaluation and Field Study Methods to Include Different Viewpoints. User studies with practitioners helped us to understand their goals and requirements regarding the usage data. In addition, heuristic evaluations [28] by external evaluators supplemented user observations in the early stages of the iterative development process by providing a good summary of general usability problems related to user interface and data visualizations. Although MILC approach [1] does not mention expert evaluations, previous studies [10, 32] have found heuristic evaluations done by external HCI experts to be useful in identifying usability issues when developing data visualization tools. While it can be challenging to find HCI experts who are also familiar with the specific domain, such as manufacturing automation, hiring students with HCI or visualization background can be a viable option [32]. Stakeholders from the company could also act as evaluators, but it may be challenging to motivate them to invest time in learning the evaluation process and conducting the evaluations.

7. Collect Log Data of the Analytics Tool to Follow Its Usage. It can be convenient if tool developers use the tool itself to analyze log data collected from its usage. In UX-sensors, the logging mechanism was designed to store data in a format compatible with the system, so that we could use it internally to support the evaluation activities. The log data reveals how actively the participants really use the analytics tool, without a need to disturb company employees with questions regarding the tool usage. This information can be used to motivate the participants and plan interventions if needed. Logged usage data from the tool provides information on how different features are used over time, especially outside observation sessions. Finally, log data provides information about how the tool is used after the collaborative development period, revealing its real applicability to the company over time.

8. Provide Support for Users with Varying Analytics Skills. Interactive data analytics tools should support users who are less familiar with programming and analytics [33]. Although the need for help texts and support for the learning process was highlighted in our heuristic evaluation results, we argue that the company personnel with less experience in data analytics will also benefit from these improvements. Furthermore, stakeholders who do not actively participate in the development process of the tool, but who might use it in the future, are likely to find instructions designed for novice users helpful. Finally, presenting the generally most interesting data first in the UI is recommended. In our case, this meant the frequencies of used features and error events, which were the first tabs in the main data browsing and analysis view of the tool.

9. Support the Sharing of Insights. The key principles of developing creativity support tools include support for collaboration and open interchange [12]. While stakeholders have their own channels for communication, visualization tool developers can support the sharing of insights during group discussions and by implementing features into the tool that support information sharing. For example, we allowed users to add notes to the usage data timeline, with the aim that others inspecting the same data could view these comments. The sharing of findings can also be supported by allowing an easy exporting of data tables and visualization images from the data analytics tool.

7 DISCUSSION

We developed UX-sensors, a visual data analytics tool for logged usage data, in collaboration with a FMS supplier company, aiming to support their R&D, customer support and service business activities in the future. Based on the lessons learned during this study, we provided guidelines to support other researchers and practitioners developing visual data analytics tools in similar contexts.

We learned that developers in the collaborating company rarely had opportunities for gathering feedback on how the FMS are used on a daily basis after they have been installed in the customer's factory. UX-sensors was seen as a potentially helpful tool in accessing the more systematic data of FMS usage over a longer time, such as months or years. Usage data was expected to provide insights on how users are using the FMS and then guide the qualitative research and on-site interventions to study why users use FMS in a specific way.

Developing data analytics and visualization tools can be a challenging process when the tool development is done separately from the development of the underlying automation system and its data logging services, as in our case. Without a deep understanding of the logging process, close collaboration was required with the developers of the manufacturing system during data wrangling. On one occasion, the collaborating company had to update the logging capabilities of their FMS, which required us to update the UX-sensors tool as well. Similar findings were reported during the development of a visual game analytics tool [3], where anticipating the effects of game design changes was presented as one guideline. Steady communication, such as participating in the development team's weekly meetings, could keep analytics tool developers informed if any changes are planned to the logging capabilities of the manufacturing system.

Our experiences from the iterative development process where we applied and extended the MILC [1] approach with expert evaluations were generally positive. The used evaluation methods provided us with meaningful data to support the UI development of UX-sensors. Interestingly, while the user observations and interviews provided the most important findings for the tool development, the HEs in the early phases of the development suggested various improvement ideas for the tool UI that were not identified with the participants working in the FMS context. In comparison to [3], we did not face significant prejudice against usage data logging among the stakeholders. The participants appeared to be genuinely interested in the possibilities of usage data logging, and during the long study period, they actively participated in the organized sessions and answered repeated web surveys. The user experience metrics survey provided useful comparison points over time regarding how the UX of the tool was evaluated. However, choosing a four-point Likert scale for the sake of effortlessness for the participants seemed unnecessary, as scale with e.g. five or seven points would have provided more detailed results regarding how the participants evaluated the UX of the tool over time.

The main insight that participants gained from the logged usage data of the FMS revealed unexpected navigation paths used in the UI when users activated the autopilot feature, suggesting a need for updating the UI or offering training for users. While such insights can benefit UI designers and end-user instructors, the actual benefits of usage data logging for other stakeholders remained unexplored. For example, would the logged usage data help customer support and developers in working out complex error events, and would customers benefit from annual reports regarding their FMS usage? Furthermore, before new services such as training offerings on a personal level can be realized, legal issues related to user privacy and data ownership need to be carefully settled with each customer.

Two insights gained during the development and evaluation process are likely to be of interest beyond the complex industrial systems context. First, the potential business value of the captured usage data for the end-user organizations of the system incorporating logging could be a way to create buy-in with customers. Second, the captured usage data could be used to develop additional services that enhance the end-user organization's use of the system, such as targeted training.

Limitations and future research. Our results are limited due to the lack coverage of different practitioner roles from one company. This is a common limitation to evaluation studies on the information visualization domain, where long-term involvement and motivation are required from participants. Future studies should focus on what kind of insights or benefits real logged usage data can provide for stakeholders also in other roles such as marketing, sales, and user training, and study the customers' viewpoints regarding the value gained from usage data logging.

Lam et al. [7] emphasize the growing need, referring to [34], for studying the design context for visualization tools, including work environments, users' tasks, and work practices. In the spirit of this notion and as a continuation of the work started here, we propose that future studies explore how utilizing usage data analytics tools can support current work practices in manufacturing automation organizations. The following research questions for future studies are proposed: What kind of and how significant insights can usage data logging provide in manufacturing automation context, and for whom? How are these insights shared in the organization and what is their impact over time, for example, resulting in changes in the manufacturing system UI or innovations that support product development, customer support or service business?

Given the limitations on access to both participants and real-world log data, collaborative case studies in the real context of use are the preferred way to provide a better understanding about the benefits of usage data logging in practice. This knowledge can help researchers and tool developers in designing data analytics and visualization tools with positive user experience and providing instructions that can support practitioners in utilizing insights from logged usage data in their work.

8 CONCLUSIONS

We have presented our iterative development and evaluation process for developing UX-sensors, a visual data analytics tool for logged usage data, in collaboration with a flexible manufacturing systems supplier company. We have summarized our experiences from the development and evaluation process as guidelines to support other researchers and practitioners developing usage data analytics tools for complex industrial systems. Finally, we have proposed research questions to further study the insights from logged usage data and their utilization in manufacturing automation context.

Our goal was that the developed tool would support stakeholders in the company with the generation of new insights from the logged usage data of their customer's FMS. The insights from usage data and discussions with stakeholders proposed that logged usage data analysis can potentially support UI and service development. Logged usage data was expected to provide insights on how users are using the FMS and guide the more qualitative research to study why users use FMS in a specific way.

ACKNOWLEDGMENTS

This study was supported by The Finnish Funding Agency for Innovation (TEKES), currently known as Business Finland, as a part of User Experience and Usability in Complex Systems (UXUS) research programme.

REFERENCES

[1] B. Shneiderman and C. Plaisant. 2006. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization (BELIV '06). ACM, New York, NY, USA, 1–7.

[2] S. Carpendale. 2008. Evaluating Information Visualizations. Information Visualization: Human-Centered Issues and Perspectives, 4950/2008, 19–45.

[3] B. Medler, M. John, and J. Lane. 2011. Data cracker: developing a visual game analytic tool for analyzing online gameplay. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 2365–2374.

[4] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller. 2013. A Systematic Review on the Practice of Evaluating Visualizations. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2818–2827. http://doi.org/10.1109/TVCG.2013.126

[5] T. Munzner. 2009. A Nested Model for Visualization Design and Validation. IEEE Transactions on Visualiza-tion and Computer Graphics, 15(6), 921–928.

[6] M. Q. Patton. 2001. Qualitative Research and Evaluation Methods, 3rd edn. Sage Publications, London.

[7] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. 2012. Empirical Studies in Information Visuali-zation: Seven Scenarios. IEEE Transactions on Visualization and Computer Graphics, Institute of Electrical and Electronics Engineers, 2012, 18 (9), 1520–1536.

[8] M. Sedlmair, M. Meyer, and T. Munzner. 2012. Design study methodology: Reflections from the trenches and the stacks. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2431–2440. http://doi.org/10.1109/TVCG.2012.213

[9] Crisan, J.L. Gardy, and T. Munzner. 2016. On Regulatory and Organizational Constraints in Visualization Design and Evaluation. In M. Sedlmair, P. Isenberg, T. Isenberg, N. Mahyar, H. Lam, (Eds.) Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16). ACM, New York, NY, USA, 1–9.

[10] M. Sedlmair, P. Isemberg, D. Baur, and A. Butz. 2011. Information Visualization Evaluation in Large Compa-nies : Challenges, Experiences and Recommendations. Information Visualization Journal, 10(3), 248-266. http://doi.org/10.1177/1473871611413099

[11] P. Saraiya, C. North, V. Lam, and K. Duca. 2006. An Insight-Based Longitudinal Study of Visual Analytics. IEEE Trans. Visualization and Computer Graphics, 12(6), 1511–1522.

[12] B. Shneiderman, G. Fischer, M, Czerwinski, M, Resnick, B, Myers and 13 others. 2006. Creativity Support Tools: Report from a U.S. National Science Foundation Sponsored Workshop, International Journal of Human–Computer Interaction 20(2), 61–77.

[13] Perer and B. Shneiderman. 2008. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). ACM, New York, NY, USA, 265–274.

[14] K. Wongsuphasawat, J.A. Guerra Gómez, C. Plaisant, T.D. Wang, M. Taieb-Maimon, and B. Shneiderman. 2011. LifeFlow: visualizing an overview of event sequences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 1747–1756.

[15] C.D. Stolper, A. Perer, and D. Gotz. 2014. Progressive visual analytics: User-driven visual exploration of in-progress analytics. IEEE Transactions on Visualization and Computer Graphics, 20(12), 1653–1662.

[16] D. Gotz and H. Stavropoulos. 2014. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. IEEE Transactions on Visualization and Computer Graphics, 20(12), 1783–1792.

[17] Bosch Software Innovations, https://www.bosch-si.com/solutions/manufacturing/data-analytics/manufacturing-analytics.html

[18] C. Holzmann, F. Lettner, C. Grossauer, W. Wetzlinger, P. Latzelsperger, C. Augdopler. 2014. Logging and Visualization of Touch Interactions on Teach Pendants. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14). ACM, New York, NY, USA, 619–624.

[19] C. Grossauer, C. Holzmann, D. Steiner, and A. Guetz. 2015. Interaction visualization and analysis in automa-tion industry. In Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15). ACM, New York, NY, USA, 407–411.

[20] T. Sheridan. 2002. Human and automation: system design and research issues. John Wiley & Sons, Inc., publication, Santa Monica, USA.

[21] H.A. ElMaraghy. 2006. Flexible and reconfigurable manufacturing systems paradigms. Int J Flex Manuf Sys, 17(4), 261–276.

[22] CouchDB, http://couchdb.apache.org/

[23] Node.js, https://nodejs.org

[24] Bootstrap, http://getbootstrap.com/

[25] D3 Data-Driven Documents, http://d3js.org/

[26] Crossfilter, http://square.github.io/crossfilter/

[27] J.R. Lewis. 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instruc-tions for use. Int. J. Hum.-Comput. Interact., 7(1), 57–78.

[28] C. Forsell and J. Johansson. 2010. An heuristic set for evaluation in information visualization. In Santucci, G. (Ed.) Proceedings of the International Conference on Advanced Visual Interfaces (AVI '10). ACM, New York, NY, USA, 199–206.

[29] J. Nielsen. 1993. Usability Engineering. Academic Press, USA.

[30] J. Nielsen. 1994. Heuristic evaluation. In Nielsen, J., Mack, R.L. (Eds.). Usability Inspection Methods. John Wiley & Sons, NY, USA, 25–61.

[31] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A.H. Byers. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

[32] M. Tory, and T. Möller. 2005. Evaluating visualizations: Do expert reviews work? IEEE Computer Graphics and Applications, 25(5), 8–11. http://doi.org/10.1109/MCG.2005.102

[33] J. Heer, and S. Kandel. 2012. Interactive analysis of big data. XRDS, 19(1), 50–54.

[34] C. Plaisant. 2004. The challenge of information visualization evaluation. In Proceedings of the Working Conference on Advanced Visual Interfaces (AVI). ACM, New York, USA, 109–116.

[35] E. Carroll, C. Latulipe, R. Fung, M. Terry, and D.R. Cheriton. 2009. Creativity Factor Evaluation: Towards a Standardized Survey Metric for Creativity Support. Proceedings of the Seventh ACM Conference on Creativity and Cognition (C&C '09). 127–136. doi:10.1145/1640233.1640255

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-4226-8

ISSN 1459-2045