PhD-FLSHASE-2015-14
The Faculty of Language and Literature, Humanities, Arts and Education

DISSERTATION

Defense held on 12/05/2015 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN PSYCHOLOGIE

by

Carine LALLEMAND
Born on 19 June 1986 in Thionville (France)

TOWARDS CONSOLIDATED METHODS FOR THE DESIGN AND EVALUATION OF USER EXPERIENCE

Dissertation defense committee

Dr. Vincent Koenig, dissertation supervisor
Senior Research Scientist, Université du Luxembourg

Dr. Kasper Hornbæk
Professor, University of Copenhagen

Dr. Romain Martin, Chairman
Professor, Université du Luxembourg

Dr. Regina Bernhaupt
Associated Researcher, IRIT, Université Paul Sabatier

Dr. Guillaume Gronier, Vice Chairman
Senior Research Scientist, Luxembourg Institute of Science & Technology

Towards Consolidated Methods for the Design and Evaluation of User Experience

Carine Lallemand

Lallemand, C. (2015). *Towards Consolidated Methods for the Design and Evaluation of User Experience.* (Doctoral dissertation). University of Luxembourg.

ABSTRACT

In the "third wave" of Human-Computer Interaction (HCI), the emergence of User Experience (UX) as a key concept has opened up both exciting perspectives and hard challenges. The conceptual shift to a more comprehensive and emotional view of human-computer interactions has been accompanied by the development of numerous methods and tools for the design and evaluation of interactive systems. UX research has thus been mainly driven by novelty and innovation and to date a majority of the developed tools lack validation and consolidation. UX research undoubtedly raises new concerns and challenges common conceptual and methodological practice. Thus the primary objective of this thesis is to contribute to UX consolidation.

We addressed this objective by relying on a mixed-methods approach for the empirical part of this thesis, involving comparatively large and representative samples. This part encompasses six studies, representing a variety of perspectives related to UX research consolidation. More specifically, this dissertation includes a replication study (Paper A, N = 758), the translation and validation of a tool (Paper B, N = 381), the development of new design and evaluation methods (Paper C and D, N = 137 and 33), the empirical assessment of the relevance of established HCI methods for the evaluation of UX (Paper E, N = 103) and finally an investigation on how to bridge UX research and practice through a design approach (Paper F).

The contributions of this thesis to UX research and practice regard both UX as a concept and its methodologies. Main findings inform about the benefits, challenges, and limitations of UX consolidation strategies as derived from our respective studies (papers A to F). Each study provides advances to both research and practice, while the combination of our studies pushes forward consolidation of UX. This is an essential step with regards to an emerging concept and informs an overarching research agenda aiming at a continuous interdisciplinary fostering of the UX field.

Keywords: user experience, evaluation, design, methods, research consolidation.

ACKNOWLEDGMENTS

Working through a PhD project is a demanding job and sometimes might seem a lonely task. When I think however about all the people who have contributed to bringing it to life, I realize that this PhD is much more than only the result of my work, perseverance and desire. This thesis is of course the result of three years of effort, but it is also much more than that.

First of all, this PhD stems from the opportunity that was offered to me to apply as a PhD candidate. My work would have been very different and it would not have felt so easy and rewarding if I could not count on the unfailing trust and support of my supervisor and advisors. My very first thanks therefore go to Dr. Vincent Koenig, Dr. Guillaume Gronier and Prof. Romain Martin for their patience, open-mindedness and watchful supervision. I was very lucky to be supported by this brilliant and complementary team.
The Luxembourg Institute of Science and Technology and the University of Luxembourg both provided a great environment to work in as a PhD student and I would like to thank these institutions for having welcomed me so warmly and having provided me with all necessary resources to succeed in the achievement of this work. Of course, I am also grateful to the Fonds National de la Recherche (FNR) for funding this PhD and supporting PhD students in all stages of their project.
Finally, thanks to Ghislain Sillaume and Lars Wienecke; I also had the great opportunity to be part of the CUbRIK FP7 project, where I collaborated with a great interdisciplinary team and tested some preliminary methodological approaches. It was also very instructive and inspiring to see how such a huge research project was managed.

Second, this PhD is also the result of passion and inspiration. The passion for scientific research that I developed during my studies at the University of Metz, as well as the passion for the topic of UX that came later on as I started diving deeply into the scientific literature. Much inspiration, and more, came from talented researchers who opened the way in the field of User Experience. I had the great pleasure to meet some of them and to exchange views on the subject; their talent and humility impressed me.
I would also like to thank all anonymous reviewers for the time spent on the assessment of my work and the positive changes that their constructive criticisms have led to.

Third, this PhD is the result of collaboration. A thesis is definitely not a lonely path. It is a great collaborative adventure and a unique opportunity to meet amazing people. Some of them, my colleagues, were everyday companions: each of them contributed in one and many ways, from discussing ideas to sharing my frustrations or successes. Daily crosswords challenges, jokes, funny lunches or meetings are unforgettable memories shared at the CRP with Pierre-Jean, Karen, Thierry, Michaël, François, Christina, Olivier, Lidia, Sabina, Corinne, Daniel and Mélanie.

At the university, the HCI research group was undoubtedly a precious source of inspiration and valuable advice, but also represented shared moments of conviviality. More than anywhere else, this is the place where I could appreciate the value of interdisciplinarity. A special thanks goes to my colleagues Robert Reuter, Sophie Doublet, Katja Weinerth, Nicolas Louveton, Gilbert Busana, Salvador Rivas, Martin Kracheel, Claire Müller, Roderick McCall, Eric François, Jean-Louis Huynen for their great support at several stages of this work.
I also thank my dear friends Ioana Ocnarescu, Kerstin Bongard-Blanchy and Jean-Baptiste Labrune for all studious and friendly design moments spent together. I truly enjoyed their enthusiasm and the work we have achieved on "extra-terrestrial objects"!

The experience of this thesis and its content would also have been very different without several years of rewarding involvement in the association FLUPA, as a council member and Vice-President. Regular contacts with UX practitioners helped me to understand UX in practice and to always wonder how to effectively transfer the results of UX research into the "real world". Moreover, several colleagues whom I met in FLUPA also spent time to participate in my experiments or to reflect on my research work and its potential to be used in their daily job. Some of them are already mentioned above; in addition, thank you Julien, Valériane, Daphné, Raphaël, Corinne, Franck, Joffrey, Christophe, Nabil, Maxime, and Patrice.

I am also very much indebted to those with whom I had lengthy discussions or e-mail exchanges, or who read parts of this thesis and whose comments were invaluable; as well as those who accepted to spend their time to participate in my experiments. Those who were not already mentioned will recognized themselves.

Finally, as I think about the deepest roots of this work, about the original motivation and strength I had for achieving this thesis, I would say that it is undoubtedly the result of my education. I was incredibly lucky to be raised by an extraordinary mom, with a gentle heart and brilliant mind. Based on the way she led her own life, she taught me how to make my own decisions and how to become an independent and persevering woman. Warm thanks to my family members and relatives as well.

Last but not least, I would have done this work with much less pleasure and daily happiness without the love, presence and support of Clément.

Merci à tous.

LIST OF JOURNAL PUBLICATIONS

This Ph.D. dissertation is based on the following six manuscripts, each reporting on one of the studies conducted over the course of the present research. Except for paper F (published in a refereed international conference), all papers have been submitted to high-ranked international journals, some of them being already published or accepted for publication.

The present manuscript starts with an introductory section, which aims at introducing the research topic, the research objectives and the contributions of the thesis to the research area. Following this introduction, the manuscript includes the papers listed below. Finally, we present a general discussion, conclusion and perspectives for future work.

In the present manuscript, chapters are labelled with numbers and papers with letters. For the sake of homogeneity, the papers have been slightly reformatted.

A. 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. http://dx.doi.org/10.1016/j.chb.2014.10.048

B. Lallemand, C., Koenig, V., Gronier, G., & Martin, R. Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. Submitted for journal publication.

C. Lallemand, C., Koenig, V., & Gronier, G. The contextual dimension of user experience: development and application of the user experience context scale (UXCS). Submitted for journal publication.

D. Lallemand, C. UX Cards – Experience design through psychological needs. Submitted for journal publication

E. Lallemand, C., Gronier, G., Koenig, V. How does UX alter established HCI evaluation methods? Expert versus laboratory evaluation. Submitted for journal publication.

F. Lallemand, C., Bongard-Blanchy, K., & Ocnarescu, I. (2014). Enhancing the design process by embedding HCI research into Experience Triggers. In Proceedings of ErgoIA 2014, Biarritz, France. http://dx.doi.org/10.1145/2671470.2671476

OTHER PUBLICATIONS RELATED TO THIS THESIS

In addition to the main research papers included in the present manuscript, the following workshop, conference or magazine publications are related to this research work.

Lallemand, C., Koenig, V., & Gronier, G. (2014). How Relevant is an Expert Evaluation of User Experience based on a Psychological Needs-Driven Approach? *Proceedings of the 8th Nordic Conference on Human-Computer Interaction NORDICHI'14*. New York, NY: ACM Press (pp. 11-20). doi:10.1145/2639189.2639214

Lallemand, C., Koenig, V., & Gronier, G. (2013). Replicating an International Survey on User Experience: Challenges, Successes and Limitations. *Proceedings of the CHI'13 Workshop on the Replication of HCI Research*, April 27-28, 2013, Paris, France.

Lallemand, C., Gronier, G., & Koenig, V. (2013). L'expérience utilisateur : un concept sans consensus ? Enquête sur le point de vue des professionnels. *Actes de la 8e conférence de Psychologie Ergonomique EPIQUE 2013*, Bruxelles, Belgique.

Lallemand, C. (2012). Using diaries to study UX: Explication, Application, *User Experience*, *11* (3).

TABLE OF CONTENTS

PART II - STUDIES

APPENDICES

Part I

RESEARCH SUMMARY

INTRODUCTION: USER EXPERIENCE DESIGN AND EVALUATION

1

"*User Experience is not about good industrial design, multi-touch, or fancy interfaces. It is about transcending the material. It is about creating an experience through a device.*"
- Marc Hassenzahl (2013)

1. THE ADDED VALUE OF EXPERIENCE DESIGN

"*Welcome to the experience economy*"
In 1998, Pine and Gilmore were pioneers in predicting the emergence of an "experience economy" where experiences are described as a real offering that businesses should stage and sell. At the same period in the field of Human-Computer Interaction (HCI) the notions of emotion and pleasure with products started to be acknowledged by authors (Jordan, 1998, 2000; Norman, 2004), thereby emphasising a major conceptual and methodological shift in the study of human-computer interactions. Following the "era" of usability and user-centred design, the HCI field entered the "era" of user experience (UX) and experience design (Hassenzahl, 2010).

The fact that UX has rapidly become a major concern for both researchers and practitioners in HCI (Hassenzahl & Tractinski, 2006) might be explained in several ways. According to Hassenzahl (2010, 2013), modern societies are moving from a materialistic view to an ever-growing experiential perspective. Studies in positive psychology on what makes people happy indeed tend to confirm that experiences make people happier than material possessions (Van Boven & Gilovitch, 2003). But what are the implications of such findings on the design of interactive products and systems, given that we are still predominantly designing technological artifacts (which are actually "material possessions") rather than experiences? Thankfully, artefacts and experiences are no contradiction, and many material belongings actually support positive experiences (Van Boven & Gilovitch, 2003; Hassenzahl, 2013), the values and meanings associated with products being more important to the users than the products per se (Vyas & van der Veer, 2006).

While people are becoming more connected and dependent on technology (Rogers, 2009), McCarthy & Wright (2004) explain that "technology is deeply embedded in everyday experience" and that we therefore need to understand human-computer interactions by looking at technology as an experience. Because technology has the potential to change people's lives, we should strive for experience-centred technological products, having the power to support "pleasurable, meaningful and even treasured moments" (Hassenzahl, 2013).

As stated by van Boven & Gilovitch, "experiences cannot be acquired if they are not available" (p. 1201). Communities should therefore invest in

making experiences available to people. And so should the HCI community. Hassenzahl (2013) argues to put experiences before things and to make experience an "explicit objective for design" (p. 2). To design for experience however is a challenging objective and we will describe within this chapter how the field of UX works towards that goal.

2. USER EXPERIENCE AS A CORE CONCEPT IN HCI

In the 1990s, while HCI research was mainly focused on the topic of usability, Donald Norman was amongst the first authors to use the term "User Experience" in order to describe all aspects of a person's experience with a system (Norman, Miller, & Henderson, 1995). Norman explained he introduced the term UX because he believed "usability" to be too narrow to represent a holistic vision of human–computer interactions. Nearly at the same period, Alben's paper entitled "Quality of experience" (Alben, 1996) put the focus on users sensations, their understanding of how things work, their feelings during usage, the achievement of their goals and also on the overall interaction context. Since then, User Experience has progressively grown into a core concept of Human–Computer Interaction (Dix, 2010). This conceptual shift to a more comprehensive and emotional view of human-computer interactions has opened up both exciting perspectives and hard challenges.

Initially considered as a buzzword or umbrella term (Roto, Law, Vermeeren, & Hoonhout, 2011), the first challenge faced by the UX research field was to understand, define and scope UX (Law et al., 2009). One of the three pillars of the UX Manifesto, published in 2007 (Law et al., 2007) therefore consisted in answering the question "What is UX?", in particular by studying the basic concepts and assumptions related to UX. Across the numerous definitions that have been proposed (Desmet & Hekkert, 2007; Hassenzahl & Tractinsky, 2006; Law et al., 2009), researchers and practitioners agree that UX is the result of the interaction between three elements: the user, the system and the context (Roto, Law, Vermeeren, & Hoonhout, 2011). Following this common vision in the field of HCI, 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" (p. 95).

Despite sharing many common grounds with the concept of usability (ISO, 1998, 2010), UX spans further by also including emotional, subjective and temporal aspects involved in the interaction. As stated by McCarthy and Wright (2004): "Experience of technology refers to something larger than usability or one of its dimensions such as satisfaction or attitude." (p. 6) While the concept of usability mainly focuses on an objective approach of the interaction, UX is also exploring subjective factors characterizing the experience between the human and technology.

The analysis framework for UX studies (Blythe, Hassenzahl, Law, & Vermeeren, 2007) exemplifies how extended the scope of UX research actually is by classifying UX studies according to 5 aspects. As shown in Table 1, UX studies can extend from reductive to holistic approaches, can

have evaluation or development purposes, adopt quantitative or qualitative methods, be focused on work- or leisure-based domains, and finally have either a personal or social unit of analysis. In this framework, each aspect is represented by a continuum instead of a dichotomy opposing the two underlying dimensions.

Table 1. The analysis framework for UX studies (Blythe et al., 2007)

Aspect	Representative dimension
Theory	Reductive - - - - - Holistic
Purpose	Evaluation - - - - - Development
Method	Quantitative - - - - - Qualitative
Domain	Work based - - - - - Leisure based
Application	Personal - - - - - Social

In 2010, the broadness and lack of sharpness of UX are reflected by the definition proposed by the ISO 9241-210 standard, describing UX as "a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service" (p. 3).

2.1. THE CHALLENGES OF UX DESIGN AND EVALUATION

Concurrently with numerous attempts to scope and define UX (Law et al., 2009), a broad discussion on UX evaluation and measure rapidly appeared on the research agenda (Hassenzahl & Tractinsky, 2006; Law et al., 2007, 2008). The issue of "How to design and evaluate UX" was defined as the third pillar of the UX Manifesto (Law, Vermeeren, Hassenzahl, & Blythe, 2007), focusing on the following objectives (p. 2):

- to develop theoretically sound methodologies for analyzing, designing, engineering and evaluating UX;
- to understand UX in practice through case studies, thereby identifying factors that may facilitate or hinder the incorporation of UX into interactive products.

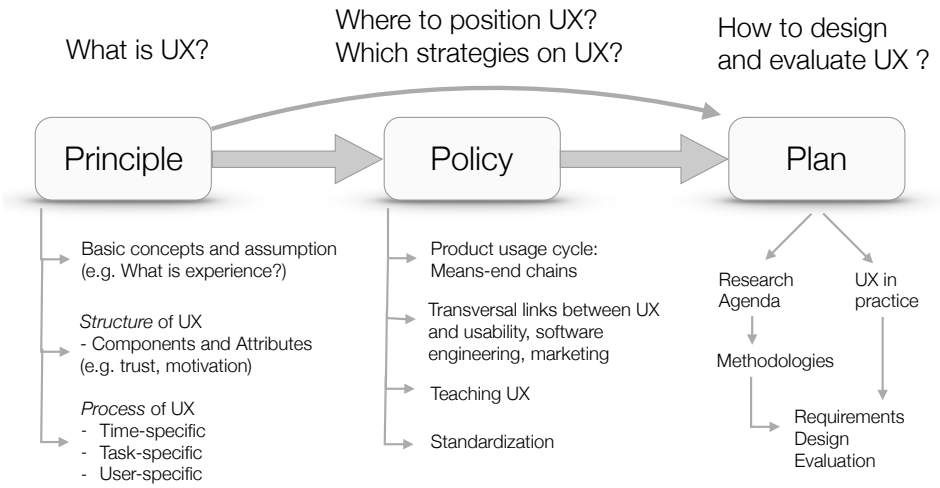


Fig. 1. The three pillars of the UX Manifesto (adapted from Law et al., 2007)

However, the diversity of definitions and interpretations of what constitutes UX along with the complexity of UX dimensions challenge UX design and evaluation to an extreme (Roto et al., 2011). It also makes it difficult to select appropriate UX evaluation methods (Bevan, 2008). Compared to usability, UX is more holistic, comprising aspects that are less tangible, and thus more complex to assess. As we have seen in the previous section with Blythe et al.'s analysis framework for UX studies (2007), the main focus is no longer on task performance in a controlled work setting (Grudin, 2006; Lazar, Feng, & Hochheiser, 2010). Traditional usability metrics such as task completion time do not reliably inform about the felt experience (Rogers et al., 2007). In 2006, Bødker described emotions and experiences as keywords in the third wave of HCI, but also as a challenge the field is facing.

Researchers generally agree that UX is subjective, holistic, situated, temporal, and has a strong focus on design (Bargas-Avila & Hornbæk, 2011; Roto et al., 2011). Each of these characteristics involves specific requirements in terms of design and evaluation.
Based on the literature (Hassenzahl, 2008; Law et al., 2009; Roto et al., 2011), we assume UX to be unique to an individual and influenced by several factors, encompassing prior experiences and expectations based on those experiences. As UX is related to users' perceptions, objective measures (e.g., task execution time, number of errors) extensively used to evaluate the usability of interactive systems therefore appears as insufficient for UX evaluation. The collection of subjective and emotional data on the felt experience is a necessary step to understand UX.
The social and cultural context of the interaction also plays an important role by impacting the felt experience: "UX may change when the context changes, even if the system does not change." (Roto et al., 2011, p. 10). However, as technology evolves, contexts of use and application types are "broadened and intermixed" (Bødker, 2006). The highly contextual nature of UX thus challenges evaluation, as it ideally requires a holistic assessment of the interaction and questions the evaluation in artificial settings. Finally, the temporal dynamics of UX is also highlighted in prominent publications in the field of HCI (Karapanos et al., 2010; Roto et al., 2011), which showed that UX actually precedes the direct interaction with a product (e.g., through advertising, experience of related technologies or reading of online user reviews) and does not stop just after usage. Time spans of UX have therefore been identified to help thinking about the dynamics of UX (Karapanos et al., 2010; Roto et al., 2011). As UX varies ceaselessly across time, its assessment requires longitudinal methods able to capture this dynamics or a constraint on the time span of UX one wants to apprehend.

3. UX EVALUATION METHODS: AN OVERVIEW

To account for the richness and complexity of experiences, UX research attempts to produce viable alternatives to traditional HCI methods. Researchers have responded to the challenges underlying UX by developing new methods or by adapting existing ones to the characteristics of UX. While there is no generally accepted overall measure of UX (Roto et al., 2011), many methods and tools have been proposed to assess UX or some of its dimensions. As they rely on different conceptualisations of UX and

have their roots in a diversity of disciplines (Stolterman, 2008), these methods show a high degree of variety.

3.1. CLASSIFICATION OF UX EVALUATION METHODS

Nearly ninety UX evaluation methods have been identified and categorized by researchers around 2010 (Roto, Obrist, & Väänänen-Vainio-Mattila, 2009; Vermeeren et al., 2010) according to the four following criteria (Figure 2):

- the type of method,
- the development phase,
- the period of experience under investigation,
- the evaluator / information provider.

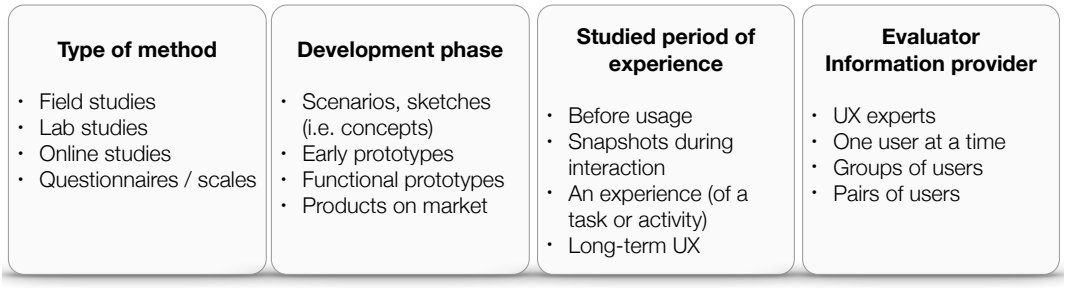


Fig. 2. Categorization of UX evaluation methods (adapted from www.allaboutux.org)

Such a categorization of UX evaluation methods provides benefits to both practitioners and researchers. From an industrial viewpoint, it offers a method repository and allows practitioners to select the most suitable method according to the requirements of their project and the available resources. From an academic viewpoint, such overviews allow for taking stock of the state of UX research by reviewing what categories of methods are well researched and what categories still need further scientific development.

3.1.1. TYPE OF METHOD

Method types are: field studies, lab studies or online studies.
A laboratory study refers to the evaluation of human-computer interactions in a controlled environment where the evaluator monitors the use of a system, observes the users' actions and reactions, and assesses the users' feelings about the quality of the interaction.
Laboratory evaluations are generally opposed to field studies (also called "in-situ" or "in-the-wild"), which involve assessing the interaction in its real context of use. Both types of methods might involve the use of a combination of methods (aka. mixed-methods) and often include questionnaires or scales that the participants are invited to fill out (Lazar, Feng, & Hochheiser, 2010).
Some methods are location independent and might therefore be used either in the lab, the field or online (Vermeeren et al., 2010).

3.1.2. DEVELOPMENT PHASE

With progressing development, the available materials vary and evolve. In Vermeeren et al.'s classification (2010), UX evaluation methods are organized according to four development phases, depending on the material available to conduct the evaluation. In this category, methods are classified according to the product development phase where the use of the method fits best. The first phase is the evaluation of concepts under the form of scenarios or sketches; followed by the evaluation of early prototypes, functional prototypes and, finally, fully fledged products on market.

3.1.3. STUDIED PERIOD OF EXPERIENCE

The studied period of experience is of particular importance for the assessment of UX, as users' experiences are strongly dependent on temporal aspects. Being a "continuous stream", UX is highly dynamic and will vary ceaselessly across time (Law et al., 2009; Roto et al., 2011). When assessing UX, one should therefore be aware that each method generally assesses only a single time span of UX. It is therefore necessary to choose which time span fits best the objective of a specific study.

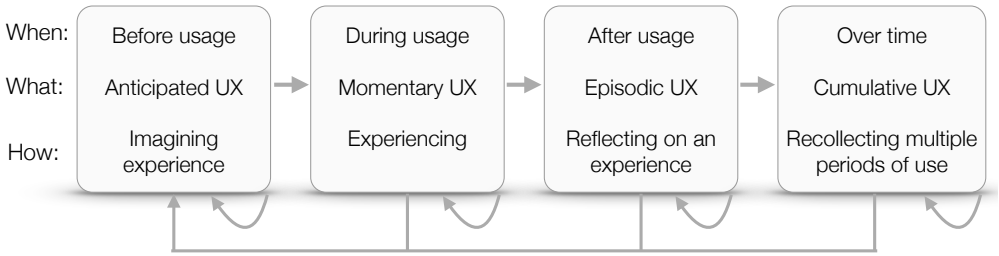


Fig. 3. Time spans of user experience. (adapted from Roto et al., 2011)

The UX White Paper (Roto et al., 2011) introduces four time spans of UX (Figure 3). Anticipated UX refers to the experience as the users imagine it before usage. The assessment of anticipated UX is generally conducted by presenting concepts to the users. The Anticipated eXperience Evaluation method (AXE; Gegner & Runonen, 2012) assesses anticipated UX in a qualitative manner by presenting a product concept in an interview setting and using visual stimuli to gather users' feelings and attitudes towards the concept.

During usage, the stream of felt experience is called momentary UX. Psychophysiological measurement devices (Mandryk, Inkpen, & Calvert, 2006) or think-aloud protocols (Aranyi, van Shaick, & Barker, 2012) are examples of momentary UX evaluation tools. Moment-to-moment self-reporting tools might also be used for that purpose (Lottridge & Chignell, 2009).

After usage, users will reflect on their experience: this is episodic UX. Numerous evaluation methods and tools are focused specifically on the episodic UX time span.

Finally, through the recollection of multiple periods of use over time, users will also develop a cumulative UX. The latter can be assessed through dedicated long-term UX evaluation methods such as UX evaluation curves (Karapanos, Martens, & Hassenzahl, 2012; Kujala et al., 2011).
While some methods support the assessment of a single UX time span, others are less specific and might allow collecting data on several time spans. This is the case of many questionnaires, reported by their authors as being relevant for assessing several time spans of UX (Vermeeren et al., 2010). However, empirical studies are needed to check whether these claims are justified.

3.1.4. INFORMATION PROVIDER

This last category is focused on who provides the information collected by the method: is the evaluation conducted by a UX expert or by one or several users (either through individual or group sessions)?
Out of their collection of 96 UX evaluation methods, Vermeeren et al. (2010) identified only 13 expert methods of which six require users or groups of users in addition to the expert. Only seven methods are therefore described as purely expert-based. Amongst these, some methods are derived from expert evaluation, like Jordan's property checklists (2000) described as a structured way to conduct an expert evaluation; or Jordan's concept of immersion where the investigator himself uses the system in real contexts and evaluates it. Following the evolutions of the HCI field, several sets of UX heuristics have also been developed (Arhippainen, 2013; Colombo & Pasch, 2012; Väänänen-Vainio-Mattila & Wäljas, 2009).
As the involvement of users in design and evaluation processes is relentlessly stressed to be an essential requirement for UX evaluation, methods involving users outnumber the expert-based methods. The majority of UX evaluation methods involve an evaluation by users in individual sessions, whereas methods relying on pairs of users or groups of users are a less common information source.

3.1.5. ADDITIONAL CLASSIFICATION CRITERIA

In addition to the four classification criteria introduced by Vermeeren et al. (2010), further categories might be used to categorize UX evaluation methods. Relevant categories might include the type of stimulus used to conduct the evaluation, objectivity vs. subjectivity of the data collected or the dimension of UX being assessed.

The most common types of stimuli are verbal (e.g., using questions, sentences or words) or visual (e.g., using pictures or characters). Non-verbal methods are often used whenever the language is likely to constitute a barrier to the evaluation. Evaluations of children's experiences are therefore often conducted using visual scales such as SAM (Bradley & Lang, 1994) or PrEmo (Desmet, 2003). These tools are also commonly used for the evaluation of users' emotional responses, which might be hard to express verbally. Recent approaches however try to extend the scope of UX evaluation methods by involving other basic senses such as the touch (Isbister, Höök, Laaksolahti, & Sharp, 2007). Just as Obrist, Tuch, and Hornbæk (2014) envisage using smells to explore novel design

opportunities, we could also imagine future approaches using smell or taste in support of UX evaluation (e.g., asking users to associate a given smell or taste with certain aspects of a felt experience). Finally, ambiguous stimuli are sometimes used to stimulate users' interpretation and provoke their responses. This is the rationale behind constructive or projective evaluation techniques such as the sentence completion technique (Kujala et al., 2013) or the AXE method (Gegner & Runonen, 2012). This category also encompasses the analysis of UX through users' collages (McKay, Cunningham, & Thomson, 2006) or drawings (Xu et al., 2009).

For evaluation methods involving final users, the collected information might be either objective (e.g. using psychophysiological or performance measures) or subjective (collected by an evaluator, e.g. during an interview or self-reported by a user).

Finally, UX evaluation methods might also be classified according to the dimension or attribute they assess. Only few methods pertain to a holistic UX evaluation, depending on which UX model or theory they rely on. Examples of holistic evaluation tools encompass the AttrakDiff scale (Hassenzahl, Burmester, & Koller, 2003), which provides a global subjective evaluation of both pragmatic and hedonic qualities of a system; or the meCUE scale based on Thüring and Mahlke's UX model (2007), which provides a self-reported assessment of product perceptions, users' emotions, consequences of use and an overall evaluation.
Regarding the evaluation of more specific aspects of UX, numerous methods and tools support the evaluation of users' affects (Russell, Weiss, & Mendelsohn, 1989; Watson, Clark, & Tellegen, 1988), emotions (Bradley & Lang, 1994; Izard, 1977; Scherer, 2005), intrinsic motivation (Ryan, 1982), values (Rokeach, 1973), or psychological needs (Hassenzahl, Diefenbach, & Göritz, 2010; Sheldon et al., 2001).

3.2. WHAT ARE THE NEEDS FOR METHODOLOGICAL DEVELOPMENTS?

By providing an overview of existing UX evaluation methods, Vermeeren et al.'s study allowed taking stock of the state of UX research and highlighting what additional types of methods and tools are needed.
In the category "studied period of experience", the authors point out the dearth of early-stage methods to evaluate anticipated UX, which refers to the time spans where users imagine their experience before usage. In the category "information provider", the same authors point to the lack of group methods to assess the UX of users groups.

More recently, Karapanos, Martens, and Hassenzahl (2012) highlighted the need for lightweight long-term evaluation methods. Although UX temporal dynamics is recognized in UX research as an essential factor (Karapanos, Zimmerman, Forlizzi, & Martens, 2010; Roto et al., 2011), methods supporting the capture of experience over time are scarce. The use of longitudinal research methods, often borrowed from social sciences, appears as a relevant option to provide insights into long-term UX (Lallemand, 2012). However, the cost and time associated with these methods seem to constrain their widespread use. The recent development

of retrospective elicitation methods such as the CORPUS interview (von Wilamowitz-Moellendorff, Hassenzahl, & Platz, 2006), the UX Curve (Kujala et al., 2011) or iScale (Karapanos et al., 2012), is a first answer to cope with this issue.

Some authors call for the development of more specific methods, targeted at particular application domains in addition to the more generic evaluation tools that already exist (Bernhaupt & Pirker, 2013). Indeed, generic methods do not always fit particular contexts and do not inform effectively on the impact of some characteristics of the system or product on the felt experience. While specific tools have been developed to study for instance the UX of interactive TV (Bernhaupt & Pirker, 2013), Web 2.0 services (Väänänen-Vainio-Mattila & Segerståhl, 2009; Väänänen-Vainio-Mattila & Wäljas, 2009), virtual environments (Chertoff, Goldiez, & LaViola, 2010), or games (Brockmyer et al., 2009), this seems insufficient with regards to the industrial needs.

Finally, while self-reported UX evaluation methods mainly rely on verbal or visual stimuli, some researchers are recently exploring alternative approaches such as tactile UX evaluation methods (Isbister, Höök, Laaksolahti, & Sharp, 2007; Lottridge & Chignell, 2009; Regal et al., 2014; Tscheligi et al., 2014). Only few research works however explored these approaches so far and there is a global need for further development of the field of tactile UX evaluation.

Beyond the development of novel methods, Vermeeren and colleagues (2010) also highlighted that three major concerns have to be addressed with regards to UX evaluation methods at a general level: their scientific quality, their practicability and their cost-effectiveness. The limitations of current UX methods are at the heart of this dissertation and will be discussed in details in Sections 4 and 5. Before focusing on this essential issue, we will first introduce the methodological matters under debate in the UX research community.

3.3. METHODOLOGICAL DEBATES

During the third wave of HCI (Bødker, 2006), new topics such as UX or ubiquitous computing have shaken up established design and evaluation methods. While researchers agree that the broadened scope of UX (as compared to usability) is inevitably changing the way we evaluate the quality of an interaction, debates on how to evaluate UX are dividing the research community.

3.3.1. FIELD VS. CONTROLLED STUDIES?

While controlled experiments used to be the scientific gold standard in many disciplines, a recent trend in our field claims for more naturalistic evaluation approaches (Rogers, 2011; Schneiderman, 2008; see also Crabtree and al., 2013 "the turn to the wild" TOCHI special issue). A passionate debate animated the Ubicomp community following the publication of Kjeldskov et al.'s intentionally provocative paper "Is it worth the hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware

Mobile Systems in the Field" (2004), where the authors claim that field studies do not bring much added value to the usability evaluation process.

In the field of UX, the laboratory setting has been described as less effective for evaluating UX than it is for evaluating usability (Benedek & Miner, 2002). With the acknowledgment of the temporal and contextual factors underlying UX, the "turn to the wild" movement has gained influence in research (Rogers, 2011). As stated by Vermeeren and colleagues (2010), "evaluation studies conducted in a field setting provide a much more realistic context to obtain reliable UX data, compared to a laboratory environment" (p. 529). However, field studies are labor-intensive, costly and require functional prototypes for realistic use (Rogers et al., 2007). Building these prototypes requires "considerable technical expertise in many different areas" (Carter, Mankoff, Klemmer, & Matthews, 2008). Real settings also challenge observation and data collection as one should try to observe and record interactions without interfering too much in the situation. Additionally, there is less control over confounding variables that might bias the results.

To balance the tension between field and controlled studies, some authors have proposed adding contextual features to laboratory setups in order to improve the realism of the laboratory setting (Kjeldskov & Skov, 2003; Kjeldskov et al., 2004). This can go from adding specific furniture, triggering specific situations through role-playing (Simsarian, 2003) to recreating a complete use context in a laboratory (Kjeldskov et al., 2004). Kjeldskov and Skov (2007) refers to controlled high fidelity simulations of the real world by using the term "in-sitro studies", as a combination between "in-situ" evaluations in the real world and "in-vitro" evaluations in controlled laboratory environments. In the case of UX however, recreating a meaningful setting in a controlled environment seems challenging as UX is very often embedded in - and influenced by – daily routines and usual social interactions.

3.3.2. QUALITATIVE VERSUS QUANTITATIVE?

Another methodological debate in the field of UX opposes qualitative versus quantitative evaluation approaches (Law, 2011). The choice between qualitative and quantitative research approaches generally implies collecting richer but less validated data on the one hand or more standardised but limited data on the other hand. In the case of UX, the choice of using quantitative or qualitative approaches often depends on the conceptual and theoretical positioning of UX researchers, as well as on their research background. Whereas some researchers argue for the need to measure UX, others question this quantitative perspective, unconvinced that experiential or emotional dimensions might be reduced to numbers (Law et al., 2007; Law, 2011). In her paper on the measurability and predictability of UX, Law (2011) opposes the "design-based UX research camp" and the "model-based UX research camp".
On the one hand, the design-based camp argues that UX should not be reduced to a number of factors by pinpointing that, "such approaches… can miss some of the insights available in accounts that resist such reduction… qualitative data provides a richness and detail that may be

absent from quantitative measures" (Swallow, Blythe, & Wright, 2005, pp. 91-92). Moreover, they also suggest that the predefined properties of UX explored through quantitative tools might not be relevant and meaningful to all users (Kujala, Walsh, Nurkka, & Crisan, 2013). The main advantage of this qualitative perspective is that rich description of experiences might be a source of inspiration for designers (Law et al., 2007).
On the other hand, the model-based camp argues for the measurability of UX attributes and processes by claiming that, "there is nothing that cannot be measured" (Law, 2011, p. 4). This approach, inspired by experimental psychology, attempts to study the interrelationships between UX components in order to identify general UX principles and mechanisms (Law et al., 2007). According to Law and van Schaik (2010), UX measures provide benefits to UX practitioners by allowing them to compare several artefacts or design options. The proponents of this camp highlight however that the validity, usefulness and meaningfulness of quantitative measures need to be carefully considered (Law & van Schaik, 2010). Both approaches being far from mutually exclusive (Law et al., 2007), they also do not deny the appropriateness and usefulness of qualitative approaches in understanding UX.

Beyond this conceptual division, choosing between quantitative and qualitative might also result from a trade-off between the richness of the collected data and data analysis constraints. As noted by Vermeeren et al.'s (2010), "many UX researchers are passionate about having open, qualitative evaluation methods, as predefined metrics may reveal just small parts of the whole UX. […] However, the practicability of methods without predefined measures is lower, since data analysis is harder with qualitative data" (p. 528). To balance the tension, many academics make use of multiple methods (i.e., mixed methods), thereby combining qualitative and quantitative perspectives, which is also the approach taken by us. Both perspectives appear complementary in accounting for the complexity of UX, qualitative findings often being used to understand the reasons behind quantitative ratings (Dix, 2010; Kujala et al., 2013). As stated by Law, van Schaik, and Roto (2014), "to be fruitful in the field of UX, it is a must to have both…. integrating the advantages of both approaches rather than treating them as mutually exclusive is the challenge of future UX theorizing" (p. 3)

Interesting research papers give us some clues to understand the reality of UX research and practice with regards to this qualitative vs. quantitative debate. In their attempt to categorize UX methods, Vermeeren et al. (2010) reviewed 96 methods and found a third being quantitative, a third qualitative and the last third using a mixed-methods approach. In their review on empirical research on UX, Bargas-Avila and Hornbæk (2011) showed that among 66 empirical studies, half were qualitative, while researchers used quantitative methods in 33% of the cases and combined approaches in 17% of the cases. By comparing their results to Barkhuus and Rode's review (2007), they also show an apparent shift in academia from quantitative to qualitative methods. More recently, Law et al. (2014) conducted both interviews and a survey to collect attitudes toward UX measurement. The outcomes of their study allowed them to build a synthesis of arguments for and against UX measurement. Here again, no clear consensus emerged and respondents mentioned distinct pros and cons

for both qualitative and quantitative UX assessment methods. Some participants described quantitative measures as useful to convince decision makers to modify a problematic design, whereas others highlighted the ease to derive alternative design ideas from qualitative UX feedbacks.

After having reviewed existing UX evaluation methods from a research perspective, we will describe UX evaluation practices in the next section.

3.4. UX EVALUATION METHODS IN PRACTICE

How is UX actually addressed in current evaluation practices? A recent survey on the state of UX evaluation practice paints "a picture of the "UX jungle" in what concerns UX evaluations", which remain fuzzy and often subjective (Alves, Valente, & Nunes, 2014, p. 101). In addition, novel UX evaluation methods from academia are transferred into practice slowly (Odom & Lim, 2008). One can even observe that methods developed in academia are rarely applied in software development practices (Ardito et al., 2014). This is partly due to the fact that UX methods are demanding and still need to be adapted to the requirements of evaluation in an industrial setting (Väänänen-Vainio-Mattila, Roto, & Hassenzahl, 2008). The cost and challenges associated with some methods preclude their frequent and widespread use (Gray, Stolterman, & Siegel, 2014; Rogers, 2004). Some practitioners also complain about the fact that the developed methods would not lead to the expected results and globally lack relevance to design practice (Stolterman, 2008).

While experience design requires novel approaches to cope with the complexity of the object under study, the analysis of both research and design practice show that many professionals have not yet changed their design and evaluation methods. "Strangely, while I find the proposition to consider the experience before the thing quite a radical change, many practitioners and academics of HCI happily embrace experience – however, without changing much in their approach." (Hassenzahl, 2013) UX being commonly understood by practitioners as an extension of usability (Lallemand et al., 2015), established usability evaluation methods remain standard practice for the evaluation of UX (Alves et al., 2014). In Academia, Bargas-Avila and Hornbæk (2011) reviewed 66 empirical studies on UX and concluded that the most frequent UX evaluation pattern is a combination of "during and after measurements – similar to traditional usability metrics, where users are observed when interacting and satisfaction is measured afterwards" (p. 2694).

Beside this observation, studies on UX practice point out that practitioners are using a myriad of evaluation tools and methods that are not backed up by research (Alves et al., 2014; Gray et al., 2014; Green, Dunn, & Hoohout, 2008). Informal low-cost methods are widely used (Venturi & Troost, 2004; Vredenburg et al., 2002), notwithstanding their inherent limitations and the ever-growing development of alternative methods involving users. Time and budget constraints are often mentioned as main explanations for the non "proper" application of research-based methods and the global adoption of low-cost - so called "guerilla" - design and evaluation methods

(Gray et al., 2014; Hartson & Pyla, 2012). Homegrown methods and associated ressources are commonplace in evaluation practices (Følstad, Law, & Hornbæk, 2012).
Even though the involvement of users in design and evaluation processes is relentlessly stressed as essential, Alves et al. (2014) also showed by surveying practitioners that final users are involved in less than 50% of reported UX evaluation cases only. This basically suggests that in half of the cases, UX evaluations solely rely on experts. These results are consistent with previous findings on UCD practices, where users' involvement was often described as selective and punctual rather than systematic and where heuristic evaluations were reported as frequent (Venturi & Troost, 2004; Vredenburg et al., 2002). However, as mentioned in section 3.1.4, a majority of novel UX evaluation methods requires the involvement of final users (Vermeeren et al., 2010) and might therefore not be used by practitioners. Similarly, surveys on UX practice show that field studies are considered the most important practice, even though they are not widely used (Vredenburg et al., 2002).

The analysis of UX evaluation in practice thus calls for the need to reflect upon more valid and reliable evaluation methods, while at the same time considering their relevance to design practice, their cost-effectiveness and their applicability in industrial settings. These are prerequisite for a better transfer of findings from research to practice.

Beyond a limited transferability to practice, UX methods suffer from additional limitations. In the next sections, we will focus on the limitations of current UX methods and on the efforts already made by the research community to tackle these challenges.

4. LIMITATIONS RELATED TO UX DESIGN AND EVALUATION METHODS

In the preceding sections, we have presented an overview of UX evaluation methods in research and practice, as well as the main methodological debates animating the UX research community. We will now focus on the limitations and challenges related to current UX methods.

4.1. SCIENTIFIC RIGOR AND RESEARCH CONSOLIDATION IN HCI

More than 20 years ago, Greenberg and Thimbleby (1992) published an article entitled 'The weak science of HCI'. In this position paper, the authors argued, "the way that science is undertaken –or purported to be undertaken– in HCI is inadequate" (p. 1). According to them, most of the scientific papers published in HCI fail to present an underlying theory supporting their hypotheses, thereby weakening the claimed generalization of experimental results. Furthermore, this lack of theoretical integration also aggravates the issue of replicability, as an experiment without underlying theory requires a direct replication (i.e., the experiment being exactly repeated using the same format, tools, and experimental protocol), which is particularly hard to achieve in the fast evolving context of HCI. Regrettably,

some of the criticisms addressed to the field at that time seem to still apply to UX research nowadays, including research into design and evaluation methods.

Despite the recommendations of the APA Publication Manual (6th edition, 2010), the publication standards of both conferences and journal articles are often lacking scientific rigor when it comes to describing the methods used to conduct the study as well as presenting the results obtained (Cairns, 2007). By surveying 80 studies from the BCS HCI conference and two leading HCI journals, Cairns (2007) revealed a global issue related to the inappropriate use of inferential statistics. A large majority of the papers reviewed failed in meeting statistical standards, thereby considerably weakening their contribution to the body of knowledge in HCI. Main problems identified were related to results reporting, checking of assumptions, over-testing and use of inappropriate tests. According to Cairns, this lack of scientific rigor is equally attributable to both authors and reviewers/editors: "not only are authors producing weak statistical analysis but this is being accepted by a large number of their peers." (p. 5). Additionally, the relatively constrained maximum page number of HCI publications was also mentioned as a cause of weak results reporting since it prevents authors to report statistical analysis in detail.

Another generic topic associated with scientific research consolidation is the issue of replicability, which is the "touchstone of common sense philosophy of science" (Collins, 1985). Greenberg and Thimbleby (1992) explained, "the most important feature of a scientific idea is that it is prepared for, indeed encourages, its own criticism, testing, refutation, and eventual replacement." Potential repeatability is therefore a way for authors to defend the validity of a claim; and replication a mean to increase confidence in- and generalization of- scientific findings.
According to Greiffenhagen and Reeves (2013), replication recently emerged as a key concern in some areas of HCI research because of "the feeling that HCI emphasizes novelty over consolidation of research; consolidation that can be achieved via replication" (p. 1). In practice, replication studies in the field of HCI are relatively scarce (Hornbæk, Sander, Bargas-Avila, & Simonsen, 2014). In a recent study, Hornbæk and colleagues (2014) found out that only 3% out of 891 examined HCI papers attempted replication of an earlier finding. Amongst these, many studies were not intended to be replications.
Reasons explaining the low priority of replications in the research agenda are manifold. As highlighted by Wilson (2013), "Our community is driven to publish novel results in novel spaces using novel designs, and to keep up with evolving technology". In that sense, there is no prestige associated with replication or incentives to replicate earlier findings. Replications are often considered by reviewers as "just replications" (John, 2013) and harder to publish than novel results or research designs. Additionally, as we noted earlier in this section, insufficient or invalid documentation of scientific experiments considerably reduce potential repeatability of results. Finally, several authors also cite the multi-disciplinary nature of HCI as another explanation for the lack of concern for replication studies (Greiffenhagen & Reeves, 2013; Hornbæk et al., 2014). While psychology values experiments and considers replication as the gold scientific standard, this does not hold true for every discipline HCI is drawing on.

4.2. SCIENTIFIC QUALITY OF UX METHODS

In the third section of this chapter, we saw that researchers have responded to the challenges underlying UX by developing new methods or by adapting existing ones to the characteristics of UX. Those methods are very diverse as they draw on a multiplicity of perspectives and backgrounds associated with UX research, i.e., the fields of design, psychology, or social sciences to name the most prominent ones. It has however been noted that a majority of these new methods lacks validity and/or reliability (Vermeeren et al., 2010). While the standards and common practice of "good research" partly depend on the researcher's perspective and background (Dix, 2010), it is nevertheless necessary to safeguard a general scientific quality of UX methods. This is a strong prerequisite in academia, where the main purpose of UX methods is to acquire sound knowledge about UX phenomena, but it should also be more of a priority in industry. Based on the fact that those methods are the key elements of product development, a better scientific grounding appears as a necessity. This is the "moral need for science" described by Greenberg and Thimbleby (1992), who claim that users "have a right to expect that the best methods have been used to ensure that the product purchased will behave well" (p. 2).

Validity and reliability are key aspects of scientific quality. To be valid, a method should measure what it is intended to measure. When different measurements using the same method provide the same result, the method is considered reliable (Spector, 1992). Even though the notions of validity and reliability are commonly associated with quantitative research, validity concerns is also applicable to qualitative studies and methods (Maxwell, 1992). This is important to emphasize because the proportion of qualitative methods in UX is high (Bargas-Avila & Hornbæk, 2011; Vermeeren et al. 2010).

In some cases, UX methods are still in early phases of development and their validity has not yet been fully examined (Vermeeren et al., 2010). Those methods need to be further developed before any claim on their scientific quality can be done. This is a demanding process requiring many efforts to test a method in several contexts in order to check if the results produced are valid and reliable. This long-term effort should ideally be distributed over the community since the scientific process implies building on previous research to produce cumulative knowledge. However, UX researchers often reinvent items instead of using existing validated instruments (Hornbæk, 2006). The issue of validity with ad-hoc measures is regrettably not limited to practitioners. The use of homegrown and short-lived instruments does indeed hinder the consolidation of previous research tools and might cause additional problems in academia: (1) they are often created without a careful consideration for validity issues, which represents a threat to the validity of the findings (2) they do not allow for a proper comparison with other studies, which weakens their overall interest for the research community.
In other cases, the nature of the method itself jeopardizes its validity. Expert-based methods, widely used by practitioners (Alves et al., 2014), rely on the prediction of the felt experience by experts. Pretending to assess the users' experiences solely from the point of view of an expert seems hence

arguable (Lallemand, Koenig, & Gronier, 2014). Frequently, the development and use of methods whose validity is doubtful might be explained by a focus on practicability over scientificity (Kujala et al., 2013; Vermeeren et al., 2010).
Another common issue related to UX methods is that they are not always based on any sound theoretical knowledge or model. Before developing a new assessment tool, it is essential to clearly and precisely define the construct of interest (Spector, 1992; Law & Van Schaik, 2010).

A specific challenge is the assessment of the validity of a method as it implies having access to all necessary documentation. The comprehensive material composing UX methods is not systematically accessible: only some items of questionnaires are provided, wordings of questions or interview guides are absent, and only a snapshot of visual material is provided. This leaves an uncertainty as to whether the method might not be valid or whether the incompleteness of documentation does not allow concluding on the validity. While the AXE method (Anticipated Experience Evaluation) is for instance quite well described in Gegner and Runonen's paper (2012), only 2 visual items out of 12 are provided as an illustration. Similarly, Turunen et al.'s paper (2009) on the SUXES method provides the readers with only 1 statement out of the four multiple-items questionnaires used in the method. Obviously, the conference paper format is an obstacle to the publication of full methodological material and the authors might intend to publish the material as Appendix in a journal article. However, we argue that this is an obstacle to research, as the method cannot be used by any other research team before full access to the items is provided: this can take several years or might even never happen. Likewise, authors also often forget to clearly indicate the instructions given to participants or to specify if and how demonstration or practice was used. These issues are related to publication standards and have negative implications on the overall quality of UX research. As we have seen for the AXE or SUXES methods, a lack of documentation might prevent the research community from building on previous work. In other cases, poor documentation criteria might lead to a misimpression of validated tools, as it has been the case for the Attrakdiff scale (Hassenzahl, Burmester, & Koller, 2003): it has been considered as valid and has been re-used in several studies, even tough the initial validation study (Hassenzahl et al., 2003; Hassenzahl, 2004) was poor and required further validation steps. Standards on validation processes and reporting might be improved to avoid these pitfalls and could be made available for easy reference to the community.

UX methods thus suffer from several limitations in terms of scientific rigour. As we will see in the next subsection, language and culture exacerbate these issues.

4.3. TRANSLATION AND CROSS-CULTURAL VALIDATION

Another problem related to UX methods is the general lack of proper translation and cross-cultural validation studies. Conversely to the field of psychology, UX research papers that focus on the translation or validation of methods are scarce. Instead of being the main goal of a study, the translation of a tool is often a mean to another end, thus only briefly

mentioned in the method section. An illustration of this is provided by Raita and Oulasvirta (2011) whose description of their translation process is limited to the following sentence: "The questionnaires [NASA-TLX, PANAS, SUS and AttrakDiff scales] were translated into Finnish by the first author in several iterations where colleagues at HIIT were presented with the English original and asked to evaluate a translation."

Rigorously translating an assessment instrument is laborious time-consuming work, involving numerous steps. However it is a scientific prerequisite to ensure the equivalence of the translated version to the initial tool (Brislin, 1986). Non-rigourous scale translations constitute a threat to the validity of scientific findings (Vallerand, 1989) and other disciplines have proposed rigorous principles for the adaptation and translation of evaluation tools (Vallerand, 1989; van de Vijver & Hambleton, 1996; Gudmundsson, 2009; International Test Commission, 2010). Yet "quick and dirty" translations are common in UX research and tolerated by the community.
The example of the Attrakdiff 2 scale (Hassenzahl, Burmester, & Koller, 2003), one of the most widely used tools amongst researchers to assess the global perceived quality of interactive systems, is a good illustration of these problems. Initially developed in German, the scale was then freely translated into English without any indication on the quality of this translation (Hassenzahl, 2004). The scale has then been translated into several Europeans languages: Icelandic (Isleifsdottir & Larusdottir, 2008), Estonian (Peedu, 2011) and Finnish (Raita & Oulasvirta, 2011). These translations have in common the lack of a rigorous scientific process ensuring the validity of the translated scale. The translation is often done by the authors of the papers: "The AttrakDiff 2 questionnaire was translated from English to Icelandic by Marta Larusdottir and this was the first time it was used in the Icelandic format." (Isleifsdottir & Larusdottir, 2008); and statistical analyses on the factorial structure or internal consistency of the translated scales are insufficient, if not totally absent. The soundness of the results provided in the aforementioned studies is therefore uncertain, and the comparison with other studies using the Attrakdiff appears problematic.
Another issue exists when an English scale is applied to respondents who are not native English speakers. Here again, this practice can lead to problematic situations, as the one faced by Holm & Lehtiniemi (2011), who tried to use the English AttrakDiff scale on a Finnish sample and "found it quite tedious and unpractical to translate the terms used in the tool to those participants who were not fluent with English." (p.331). This creates an awkwardness for both the experimenter and participants; what is worse, a misunderstanding of items or questions might seriously bias the results of the evaluation.

With regards to publication practice in HCI, it is not usual for authors to inform about the language in which the material has been used in a particular study. Christou's study on game experience (2010) is one of the many examples of papers illustrating this statement. The methodological section informs us that the study was conducted in Cyprus (where the official languages are Greek and Turkish) and that the participants had to fill out the Attrakdiff scale to report their experience with the game. However, there is no indication of the language in which the scale was used.

When the authors report the translation of a tool, it is likewise not always stated how this translation has been done and by whom, nor if the translated version has been checked for validity before being used. The material composing a particular method or tool (e.g., items composing a scale) is generally disseminated in English, even though it might have been developed in another language and used on a non-English speaking sample of participants. While this practice supports dissemination of research findings, we argue that the original version of the tool should also be provided and that authors should bring to attention that the English items are only provided for illustration purposes. The English version published in journals or conference proceedings is often a free translation done by the authors, without any guarantee of equivalence with the initial material. The risk at hand in this case is that other researchers will use the English version in their experiments, without checking for validity and applicability in another cultural background.

Finally, generalizability of UX evaluation methods is also easily presumed, when methods are transferred from one cultural environment to another without consideration of potential cultural differences impacting UX. One should however remind that the theoretical models behind UX tools have so far mainly been developed in western countries and their applicability and validity in other parts of the world cannot be taken for granted (Frandsen-Thorlacius et al., 2009). Differences might even exist in two neighbour countries because of varying cultures or values (Hofstede, Hofstede, & Minkov, 2010). As done by Liu et al. (2012), it is necessary when using an instrument developed in another country to investigate whether the UX assumptions underlying the scales are applicable to the population under study. Claims on the generalizability of studies and findings across different cultural backgrounds should only be done with greatest care.

4.4. LACK OF RELEVANCE TO UX PRACTICE

Although the HCI field has always tried to integrate scientific concerns with design or engineering goals (Carroll, 1997), current research is frequently criticized for lacking relevance to UX practice (Roedl & Stolterman, 2013; Stolterman, 2008; Woolrych et al., 2011). As mentioned in previous sections, UX evaluation methods are rarely transferred to practitioners, and surveys on UX practice show that professionals favour the use of low-cost in-house evaluation methods (Alves et al., 2014; Gray et al., 2014). Practitioners also tend to desert scientific conferences they used to attend in favour of practitioner-focused events, "justifying the expenditure to attend these conferences through real and tangible gains in applicable methods" (Gray et al., 2014). Temporal and financial constraints are mentioned as the main reasons for the non-use or non-proper application of research-driven approaches (Gray et al., 2014). To address the requirements of product development in a timely manner, industry increasingly develops its own UX methods and tools, or adapts current methods under the attractive label of "guerilla" or "Do It Yourself". The majority of these industrial UX methods remains unpublished (Vermeeren et al., 2010) and will never be checked for validity, thereby increasing the gap with academia.

According to Stolterman (2008), research frequently fails at supporting design practice because "it has *not* been grounded in and guided by a sufficient understanding and acceptance of the *nature of design practice*." So, despite the disposition of UX to be an applied field of research (i.e., to generate information that can be directly applied to real-world problems; Bordens & Abbott, 2011), the methods brought by academia do not satisfactorily address the needs of industry.
Roto et al. (2013) summarized the main differences between UX evaluations in academia vs. industry (Figure 5). While academia uses evaluation methods to understand phenomena, industry uses them as a mean to the improvement of systems and products.

Academia	**Industry**
Understand the phenomenon	Improve product UX
Validate(d) methods	Cost-efficient methods
Big, well-planned study	Frequent and small evaluations
Controlled variables	Real life contexts
Project team of UX experts	UX people a minority in the project
Skilled researchers	UX evaluation often outsourced

Fig. 5. UX evaluation in academia vs. industry (Roto et al., 2013)

To address the needs of UX practitioners, methods and tools should be flexible, applicable, cost-effective, and lightweight (Roto, Ketola, & Huotari, 2008). However, reaching a compromise between scientificity and practicability is a challenging objective. In their survey on UX evaluation methods, Vermereen et al. (2010) showed that the cluster of UX evaluation methods described as having strong scientific qualities (e.g., in terms of validity) was associated with issues of practicability (e.g., specific equipment or expertise required, difficult data analysis, time-consuming method). Providing practitioners with valid and reliable methods while at the same time ensuring their cost-effectiveness and practicability is an additional challenge to be faced by the UX field.

In the current section, we showed that the limitations and challenges related to current UX methods are variegated and numerous. These issues have major implications for both research and practice and it is therefore pressing to tackle them. In the next section, we will investigate the efforts of the research community to address these problems. While UX research currently struggles with meeting practitioners' needs, we will also see within the next section how the research community has attempted so far to meet those needs and build bridges between research and practice.

5. TOWARDS CONSOLIDATION OF UX RESEARCH

To cope with the aforementioned limitations and challenges, several research efforts have already been made towards the consolidation of UX research. In this section, we will first describe how the research community has gone into action around the topics of UX design and evaluation methods. Then, we will explore the initiatives taken to raise awareness on validity issues and the need to consolidate UX research; before illustrating

this concern by describing empirical research consolidation practices. Finally, we will describe how the research community tries to reduce the gap between research and practice.

5.1. MOBILISATION OF THE RESEARCH COMMUNITY AROUND THE TOPICS OF UX DESIGN AND EVALUATION METHODS

First, a high number of workshops have been organized within the research community in order to reflect on UX design and evaluation challenges. In 2006, the workshop "UX – Towards a unified view" examined the potential and pitfalls of UX evaluation methodologies (Hassenzahl, Law, & Hvannberg, 2006). In 2007, during the "Towards a UX manifesto" workshop, the issue of "How to design and evaluate UX" was defined as one of the top-3 priorities in the UX research agenda. (Law et al., 2007). The following year, an international workshop on Valid Useful User Experience Measurement (VUUM) was organized in Reykjavik (Law et al., 2008). The CHI'08 conference additionally welcomed a workshop on UX Evaluation Methods for Product Development (UXEM; Väänänen-Vainio-Mattila et al., 2008). In 2010, the Dagstuhl Seminar on Demarcating User Experience resulted in the production of a UX White Paper (Roto et al., 2011). By bringing clarity to the concept of UX, this document has laid the basis for further reflection on UX design and evaluation. In 2012, the International Workshop on the Interplay between User Experience (UX) Evaluation and System Development (I-UxSED; Law et al., 2012) focused more specifically on the relevance to practice. Similarly, a workshop on UX and user-centered development processes will be held this year during the INTERACT'15 conference (Winckler, Bernhaupt, Forbrig, & Sauer, 2015). Recently, several workshops were also focused on specific UX evaluation methods, such as tactile UX evaluation methods (Tscheligi et al., 2014). Furthermore, international conferences such as CHI, HCI, MobileHCI, NordiCHI, INTERACT, or DIS – to only name a few – do welcom regular panel discussions on the topic of UX methodologies.

Additionally, UX research agendas have been published (Hassenzahl & Tractinsky, 2006; Law & van Schaik, 2010), often under the form of introductory papers to journal special issues on UX. By drawing a state of UX research and proposing directions for further research, these initiatives have laid the foundation for further advancements of the UX field.

5.2. RAISING AWARENESS ON VALIDITY ISSUES AND THE NEED TO CONSOLIDATE UX RESEARCH

Several recent initiatives aimed at raising awareness on validity issues and the need to consolidate UX research.

In 2008, the workshop on Valid Useful User Experience Measurement (Law et al.) welcomed submission on the validity and reliability of UX measurement. Several papers aimed at providing guidance on good research and evaluation practices: the Scale Adoption Framework for Evaluation (Green, Dunn, & Hoonhout, 2008) for instance aimed at supporting the selection of UX measurement scales by describing appropriate psychometrics properties.

The RepliCHI workshop (Wilson, Resnick, Coyle, & Chi, 2013), as its name suggests, has been focused on the issue of replication in HCI. The main idea behind this initiative was to facilitate the publication of replication studies, to raise awareness on the benefits of replication to both the research community and practitioners, and, to illustrate the potential of replication as a teaching method. Several experience reports presented at the workshop were related to UX (Carlson, Paget, & McCollum, 2013; Lallemand, Koenig, & Gronier, 2013; Tractinsky, 2013). Directions for future actions were discussed, such as educating reviewers to understand the value of replication and accept valuable replication studies (John, 2013). Amongst other relevant conclusions, the workshop highlighted the need for authors to document their experiments in sufficient details to allow for replication. RepliCHI workshop organizers also suggested creating a special venue for replication studies during the main international HCI conferences, such as the CHI conference. However, the ultimate goal of such an initiative is not to compartmentalize replication studies as a particular type of submission but to progressively encourage their full integration in the regular track of conferences programs.

In 2014, Hornbæk and colleagues raised additional awareness on the topic of replication by systematically investigating replication studies in HCI. The authors show that replication studies in our field are rare (around 3% of 891 browsed papers) and that the reporting of replications might be improved in several ways. Their paper highlights the benefits of replication studies and provides recommendations to improve reporting of experiments. With regards to the publication of replications, the authors also suggest revising the guides for submission, which work against the publication of replications by focusing too much on novelty and originality of findings.

Empirical research consolidation[1] practices per se also contribute to raise awareness on the issue of consolidation and encourage the community by setting good examples. As we have seen, one aspect of consolidation is replication. Before a finding can be accepted as well-established within any discipline, research has to be repeated, that is to say replicated. As mentioned previously, replication studies in the field of HCI are scarce (Hornbæk et al., 2014) but we can still mention some, such as Tractinsky, Katz, and Ikar (2000), Van Schaik and Ling (2008) or Tuch, Trusell, and Hornbæk (2013). Another way to consolidate research findings is to conduct meta-analyses on previously published data. By carefully aggregating studies, meta-analyses create more comprehensive and rigorous syntheses of research. While a multitude of focused studies are necessary to produce knowledge and support relevant findings, we often lack overview, therefore risking to "not seeing the forest for the trees" (Noblit & Hare, 1988). To date, empirical findings are rarely synthesized (Bargas-Avila & Hornbæk, 2011). In 2010, Hassenzahl for instance provided an overview of studies reporting a relationship between beauty and usability. With regards to design and evaluation methods, papers reviewing methods help to identify shortcomings and gaps and to propose directions for future research. They also constitute repositories that might support the selection of a method. To name only a few, we can mention Kjeldskov and Graham's

[1] Please note that we use the term "research consolidation" to refer to a large range of practices and strategies having in common to strongly build on- and improve the validity of previous research.

(2003) review of mobile HCI research methods, Vermeeren et al.'s (2010) study on UX methods or Wölfel and Merritt's (2013) survey of card-based design tools.

5.3. REDUCING THE GAP BETWEEN RESEARCH AND PRACTICE

As demonstrated in section 4.4, the lack of relevance to practice appears as a critical issue for an applied field such as UX. Of course, research methods and instruments are not always intended to be used by industry, and their primary objective is to allow for a valid and reliable assessment of the phenomenon under study. Nevertheless, as members of a research community, it is worthwhile wondering how we might provide practitioners with methods combining scientificity and practicability?

Efforts to reduce the gap between research and practice have already been made by the UX research community.

As stated by Ardito et al. (2014), "Several authors acknowledge that, in order to understand the reasons of the limited impact of usability engineering and UX methods, and to try to modify this situation, it is fundamental to thoroughly analyze current software development practices, involving practitioners and possibly working from inside the companies". Several scientific papers have thus raised awareness on the nature of design practice (Cross, 2001; Roedl & Stolterman, 2013; Stolterman, 2008). Similarly, studies are conducted to investigate practitioners' viewpoints, practices and needs (Følstad et al., 2012; Ketola & Roto, 2008; Roto et al., 2008). Ketola and Roto (2008) for instance explored UX measurement needs during product development at Nokia. Requirements for UX evaluation methods to be applicable in industry have thus been suggested (Roto, Ketola, & Huotari, 2008; Väänänen-Vainio-Mattila, Roto, & Hassenzahl, 2008). According to Roto and colleagues (2008), to be applicable in practice, methods need to: (1) be lightweight, (2) be applicable for various types of products and prototypes, (3) inform about the pros and cons of the evaluated system in order to derive suggestions for improvement, (4) produce a fair UX score, allowing comparison between different products. Väänänen-Vainio-Mattila et al. (2008) completed this list with the requirements of methods being valid, reliable and repeatable, being suitable for different target user groups and product lifecycles phases, and requiring a low level of expertise.

Other authors have approached the issue of relevance to practice by analyzing the way we actually conduct research on methods in HCI (Folstad, Law & Hornbæk, 2006; Woolrych et al., 2011). Focusing on the components of methods rather than on the method as an "indivisible whole" (Woolrych et al., 2011) would help coping with the situatedness of UX design and evaluation. Additionally, it would be worth studying the effects of some variables that researchers often attempt to reduce. Woolrych et al. (2011) argue for instance that one should not try to reduce the evaluator effect, as it is perfectly relevant to design practice for knowing more about which evaluator-dependent variables provide the best results in a specified context.

With regards to editors and publishers, scientific journals such as *Design studies* (Elsevier) specifically focus on the analysis and development of design activity and emphasise the relevance to the practice and teaching of design. In addition to the contribution of such journals, it is worth mentioning the importance of the "implications for design" section required in many scientific papers, even though some researchers argue that this requirement does not fit to all kinds of research works (Dourish, 2006). Papers providing guidance on how to select a UX method (Bevan, 2008; Vermeeren et al., 2010) or how to judge whether a scale fulfils scientific requirements (Green, Dunn, & Hoonhout, 2008) are also intended to support UX practitioners.

Finally, researchers also concretely attempt to develop practical lightweight methods in order to support product development (Karapanos et al., 2012; Kujala et al., 2013). UX retrospective evaluation tools such as the UX Curve (Kujala et al., 2011) or iScale (Karapanos et al., 2012) are for instance presented as alternatives to demanding longitudinal methods. The downside of these practices however is the risk of a decreased care for scientific rigor, as exemplified by these statements on the sentence completion method:

> "Our goal was to develop a practical technique for gathering qualitative user feedback for product development purposes rather than to develop a strict measurement tool. Sentence completion was selected as it appears practical to use in product development contexts." (Kujala et al., 2013, p. 12)
>
> "Systematic research is therefore needed to ensure the validity and reliability of the sentence completion approach…. However, as the goal is to develop a practical evaluation method for design purposes, we can question to what extent these criteria are applicable." (Kujala et al., 2013, p. 14)

Methods dedicated to specific application domains are also developed to address the needs of industry, where generic evaluation methods often fail at providing precise insights able to support product development (Bernhaupt & Pirker, 2013; Väänänen-Vainio-Mattila & Wäljas, 2009).

Globally, UX researchers seem to agree that understanding design practice and being able to use science-based UX methods in the design process is of major importance. It is nonetheless essential to beware of immediate practical usefulness of the methods developed (Kuutti, 2010) and to always follow good scientific practices. Finally, we agree with Law (2011) in saying that UX "practice and research should go hand-in-hand" (p.7). Design practice and UX research could undeniably benefit from a closer integration. HCI research can support designers in the rigorous development or selection of new methods (Green et al., 2008), while designers can provide researchers with industrial use cases to test and enhance design methods. As a young field that is still developing, UX needs more efforts in documenting the relevance and validity of design and evaluation methods. As we have seen in this chapter, there are many ways of contributing to the consolidation of UX research. The present work contributes to this objective through six studies. In the second chapter of this thesis, we will introduce our research objectives, the methodologies used to address these objectives, as well as an outline of the six research papers included in this manuscript.

REFERENCES FOR CHAPTER 1

Alben, L. (1996). Quality of Experience: Defining the Criteria for Effective Interaction Design. *Interactions*, *3*, 11. doi: 10.1145/235008.235010

Alves, R., Valente, P., & Nunes, N. J. (2014). The state of user experience evaluation practice. *In Proceedings of the 8th Nordic Conference on Human-Computer Interaction NordiCHI'14*. New York, NY: ACM Press (pp. 93-102). doi:10.1145/2639189.2641208

American Psychological Association. (2010). *Publication manual of the American Psychological Association. - 6th ed.* Washington, DC: American Psychological Association.

Aranyi, G., van Schaik, P., & Barker, P. (2012). Using think-aloud and psychometrics to explore users' experience with a news Web site. *Interacting with Computers*, *24*(2), 69-77. doi: 10.1016/j.intcom.2012.01.001

Ardito, C., Buono, P., Caivano, D., Costabile, M. F., & Lanzilotti, R. (2014). Investigating and promoting UX practice in industry: An experimental study. *International Journal of Human-Computer Studies, 72*, 542-551.

Arhippainen, L. (2013). A Tutorial of Ten User Experience Heuristics. *Proceedings of International Conference on Making Sense of Converging Media* (AcademicMindTrek '13). New York, NY: ACM Press.

Bargas-Avila, J.A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. In Proceedings of the ACM conference on human factors in computing systems (CHI 2011) (pp. 2689–2698). New York, NY: ACM Press. doi: 10.1145/1978942.1979336.

Barkhuus, L., & Rode, J. (2007). From Mice to Men - 24 years of Evaluation in CHI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'07*. New York, NY: ACM Press.

Benedek, J., & Miner, T. (2002). Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting. *Proceedings of Usability Professionals' Association Conference UPA'02*, Orlando, FL.

Bernhaupt, R., & Pirker, M. (2013). Evaluating User Experience for Interactive Television: Towards the Development of a Domain-Specific User Experience Questionnaire. IFIP TC13 Conference on Human-Computer Interaction INTERACT'13. Lecture Notes in Computer Science 8118. p:642-659. Berlin/Heidelberg: Springer.

Bevan, N. (2008). Classifying and selecting UX and usability measures. In *Proceedings of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM). Reykjavik, Island.*

Blythe, M, Hassenzahl, M & Vermeeren, APOS (2007). An analysis framework for user experience (UX) studies: a green paper. In E Law, A Vermeeren, M Hassenzahl & M Blythe (Eds.), *Towards a UX manifesto. COST294-MAUSE affiliated workshop* (pp. 1-5). Lancaster: COST.

Bødker, S. (2006). When second wave HCI meets third wave challenges. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction*. New York, NY: ACM Press (pp. 1-8). doi:10.1145/1182475.1182476

Bordens, K. S., & Abbott, B. B. (2011). *Research Design and Methods*. A process approach (8th edition). New-York, NY, USA: McGraw-Hill.

Bradley, M., & Lang, P. (1994). Measuring emotion: The self-Assessment Manikin and the semantic differential. *Journal of Behavioral Therapy & Experimental Psychiatry*, *1*, 49-59.

Brislin, R. W. (1986). *The wording and translation of research instruments*. In W. J. Lonner, & J. W. Berry (Eds.). *Field methods in cross-cultural research* (pp. 137-164). Newbury Park, CA: Sage.Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods*. New York: Wiley.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroon, E., Burkhart, K. M., Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, *45*, 624-634.

Cairns, P. (2007). HCI... Not As It Should Be: Inferential Statistics in HCI Research. *Proceedings of the 21st British HCI Group Annual Conference on People and Computers BCS-HCI '07*. Swinton, UK: British Computer Society.

Carlson, J. L., Paget, M., McCollum, T. (2013). Replication two telepresence camera depth-of-field settings in one user experience study. *Proceedings of the CHI2013 Workshop on the Replication of HCI Research*. Paris, France, April 27-28, 2013.

Carroll, J. (1997). Human-computer interaction: Psychology as a science of design. *Annual Review of Psychology, 48*, 61-83.

Carter, S., Mankoff, J., Klemmer, S. R., & Matthews, T. (2008). Exiting the Cleanroom: On Ecological Validity and Ubiquitous Computing. *Human-Computer Interaction*, *23*(1), 47-99. doi:10.1080/07370020701851086

Chertoff, D.B., Goldiez, B., LaViola, J. J. (2010). Virtual Experience Test: A virtual environment evaluation questionnaire. *Proceedings of the 2010 IEEE Virtual Reality Conference VR'10*. Washington, DC, USA: IEEE Computer Society.

Christou, G. (2010). Exploring player perceptions that contribute to the appeal of World of Warcraft. In *Proceedings of the 4th International Conference on Fun and Games FnG'12*. New York, NY: ACM Press. doi: 10.1145/2367616.2367630

Collins, H. M., (1985) *Changing Order: Replication and Induction in Scientific Practice*, Beverley Hills & London: Sage. [2nd edition 1992, Chicago: University of Chicago Press]

Colombo, L., & Pasch, M. (2012). 10 Heuristics for an Optimal User Experience. *Proceedings of alt.chi 2012, Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM Press.

Crabtree, A., Chamberlain, A., Grinter, R., Jones, M., Rodden, T. and Rogers, Y. (2013). *Special Issue of "The Turn to The Wild", ACM Transactions on Computer-Human Interaction (TOCHI*), 20(3). doi:10.1145/2491500.2491501

Cross, N. (2001). Designerly ways of knowing: design discipline versus design science. *Design Issues*, *17*(3), 49–55.

Desmet, P.M.A. (2003). Measuring emotion; development and application of an instrument to measure emotional responses to products. In M.A. Blythe, A.F. Monk, K. Overbeeke, & P.C. Wright (Eds.), *Funology: from usability to enjoyment* (pp. 111-123). Dordrecht: Kluwer Academic Publishers.

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

Dix, A. (2010). Human-Computer Interaction: a stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers, 22*(1), 13-27.

Dourish, P. 2006. Implications for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'06* (pp. 541-550), New York: ACM.

Følstad, A., Law, E. L.-C., & Hornbæk, K. (2012). Analysis in Practical Usability Evaluation: A Survey Study. *ACM Conference on Human Factors in Computing Systems CHI'12*, 2127-2136.

Frandsen-Thorlacius, O., Hornbæk, K., Hertzum, M., & Clemmensen, T. (2009). Non-Universal Usability? A Survey of How Usability Is Understood by Chinese and Danish Users. *Proceedings of ACM Conference on Human Factors in Computing Systems* (CHI 2009), 41-50.

Gegner, L. & Runonen, M. (2012). For What It Is Worth: Anticipated eXperience Evaluation. In *Proceedings of the 8th International Conference on Design and Emotion*, London, UK: The Design and Emotion Society.

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, *61*(2), 29-45. doi:10.1027/1901-2276.61.2.29

Gray, C. M., Stolterman, E., & Siegel, M. A. (2014). Reprioritizing the Relationship Between HCI Research and Practice: Bubble-Up and Trickle-Down Effects. *Proceedings of the 2014 conference on Designing interactive systems DIS'14*. New York, NY: ACM Press. doi: 10.1145/2598510.2598595

Green, W., Dunn, G., & Hoonhout, J. (2008). Developing the Scale Adoption Framework for Evaluation. In *Proceedings of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM). Reykjavik, Island*, (2008) pp. 49-55.

Greenberg, S., & Thimbleby, H. (1992). The weak science of human-computer interaction. *Research Symposium on Human Computer Interaction CHI '92*, Monterey, California, May.

Greiffenhagen, C., & Reeves, S. (2013). Is replication important for HCI? In: M. L. Wilson et al. (Eds.). Replichi 2013, *Proceedings of the CHI2013 workshop on the replication of HCI research,* Paris, France, 27th-28th april 2013. CEUR Workshop Proceedings 976, pp.8-13.

Grudin, J. (2006). Is HCI homeless ? In search of inter-disciplinary status. *Interactions*, *13*(1), 54-59.

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

Hassenzahl, M. (2008). User Experience (UX): Towards an experiential perspective on product quality. *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine* (pp. 11-15). New-York, NY: ACM. doi: 10.1145/1512714.1512717

Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*, *3*(1), 1-95. doi: 10.2200/S00261ED1V01Y201003HCI008

Hassenzahl, M. (2013). Experiences Before Things: a Primer for the (Yet) Unconvinced. *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2059-2068). New York: ACM. doi:10.1145/2468356.2468724

Hassenzahl, M., Burmester, M., & Koller, F. (2003). *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*. In J. Ziegler, & G. Szwillus (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung* (pp. 187–196). Stuttgart, Germany: B.G. Teubner. doi: 10.1007/978-3-322-80058-9_19

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers*, 22(5), 353-362.

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behavior & Information Technology*, *25*, 91-97. doi: 10.1080/01449290500330331

Hofstede G., Hofstede G. J., & Minkov M. (2010). *Cultures and organizations: software of the mind: intercultural cooperation and its importance for survival* (3rd ed.). New York, USA: McGraw Hill

Holm, J., & Lehtiniemi, A. (2011). A Virtual World Prototype for Interacting with a Music Collection. In A. Ozok & P. Zaphiris (Eds.), *Lecture Notes in Computer Science*: *Vol. 6778. Online Communities and Social Computing.* (pp. 326-335). doi: 10.1007/978-3-642-21796-8_35

Hornbæk, K., Sander, S., Bargas-Avila, J. & Simonsen, J. (2014). Is Once Enough? On the extent and Content of Replications in Human-Computer Interaction. *Proceedings of ACM Conference on Human-Factors in Computer Systems CHI'14*. New York: ACM. doi: 10.1145/2556288.2557004

Hornbæk, K. (2006). Current Practice in Measuring Usability: Challenges to Usability Studies and Research, *International Journal of Human-Computer Studies*, *64*(2), 79-102. doi: 10.1016/j.ijhcs.2005.06.002

Isbister, K., Höök, K., Laaksolahti, J., and Sharp, M. (2007). The sensual evaluation instrument: Developing a trans-cultural self-report measure of affect. *International Journal of Human-Computer Studies, 65*(4), 315–328.

International Organization for Standardization. (1998). ISO 9241-11:1998, *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.* Geneva, Switzerland: International Organization for Standardization.

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

International Test Commission (2010). *Guidelines for Translating and Adapting Tests.* [http://www.intestcom.org]

Isleifsdottir, J., & Larusdottir, M. (2008). Measuring the User Experience of a Task Oriented Software. In E. Law et al. (Eds.), *Proceedings of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement* (pp. 97-101), Reykjavik, Iceland.

Izard, C.E. (1977). *Human emotions*. New York: Plenum Press.

John, B. E. (2013). Avoiding "It's JUST a replication". *Proceedings of the CHI2013 Workshop on the Replication of HCI Research*. Paris, France, April 27-28, 2013.

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

Jordan, P. W. (2000). *Designing pleasurable products: an introduction to the new human factors*. London: Taylor and Francis.

Karapanos, E., Martens, J.-B., Hassenzahl, M. (2012). Reconstructing Experiences with iScale. *International Journal of Human-Computer Studies*, *70*(11), 849-865. doi:10.1016/j.ijhcs.2012.06.004

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J.-B. (2010). Measuring the dynamics of remembered experience over time. *Interacting with Computers*, 22(5), 238-335. doi:10.1016/j.intcom.2010.04.003

Ketola, P., Roto, V. (2008). Exploring User Experience Measurement Needs. *5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement* (VUUM). Reykjavik, Iceland.

Kjeldskov, J., & Graham, C. (2003). A Review of Mobile HCI Research Methods. In Human-Computer Interaction with Mobile Devices, the 5th International Symposium Mobile HCI'03 (LNCS, pp. 317-335). Berlin: Springer Verlag.

Kjeldskov, J., & Skov, M. B. (2003). Creating a Realistic Laboratory Setting: A Comparative Study of Three Think-Aloud Usability Evaluations of a Mobile System. In *Proceedings of the 9th IFIP TC13 International Conference on Human-Computer Interaction (Interact 2003)*. IOS Press, 663-670.

Kjeldskov, J., & Skov, M. B. (2007). Studying Usability In Sitro: Simulating Real World Phenomena in Controlled Environments, *International Journal of Human-Computer Interaction, 22* (1-2). doi: 10.1080/10447310709336953

Kjeldskov, J., Skov, M. B., Als, B. S., and Høegh, R. T. (2004). Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In *Proceedings MobileHCI 2004*, Berlin, Germany: Springer-Verlag, 529-535. doi: 10.1007/978-3-540-28637-0_6

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers, 23*(5), 473-483. doi:10.1016/j.intcom.2011.06.005

Kujala, S., Walsh, T., Nurkka, P., & Crisan, M. (2013). Sentence Completion for Understanding Users and Evaluating User Experience. *Interacting with Computers*, *26*(3), 238-255. doi:10.1093/iwc/iwt036

Kuutti, K. (2010). Where are the Ionians of user experience research ? *Proceedings of the 6th Nordic Conference on Human-Computer Interaction NordiCHI '10: Extending Boundaries*, October 16–20, 2010, Reykjavik, Iceland (pp. 715-718). New-York, USA: ACM.

Lallemand, C. (2012). Using diaries to study UX: Explication, Application, *User Experience*, *11* (3).

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

Lallemand, C., Koenig, V., & Gronier, G. (2013). Replicating an International Survey on User Experience: Challenges, Successes and Limitations. *Proceedings of the CHI2013 Workshop on the Replication of HCI Research*. Paris, France, April 27-28, 2013.

Law, E. C. (2011). The measurability and predictability of user experience. In *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems EICS'11*. New York, NY: ACM Press.

Law, E., Bevan, N., Christou, G., Springett, M., & Lárusdóttir, M. (2008). *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland.

Law, E., Roto, V., Hassenzahl, M., Vermeeren, A. & Kort, J. (2009). Understanding, scoping and defining UX: a survey approach. In *Proceedings of the ACM conference on human factors in computing systems CHI'09*. New York, NY: ACM Press. doi: http://dx.doi.org/10.1145/1518701.1518813.

Law, E., & Van Schaik, P. (2010). Modelling User Experience – An agenda for Research and Practice. *Interacting with computers, 22 (5),* 313-323. doi: 10.1016/j.intcom.2010.04.006

Law, E. L.-C., van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies, 72*(6), 526-541.

Law, E., Vermeeren, A., Hassenzahl, M., & Blythe, M. (2007). Towards a UX Manifesto. COST294-MAUSE affiliated workshop, Lancaster, UK.

Law, E., Abrahão, S., Vermeeren, A., & Hvannberg, E. (2012). Interplay between User Experience Evaluation and System Development: State of the Art. International Workshop on the Interplay between User Experience (UX) Evaluation and System Development (I-UxSED 2012), Copenhagen, Denmark, 14-17 October 2012.

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in Human-Computer Interaction*, London, UK: John Wiley & Sons.

Liu, S., Zheng, X.S., Liu, G., Jian, J., Liu, Z., & Peng, K. (2012). Assessing user experience of interactive products: a Chinese questionnaire. In *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction VINCI'12*, New York, NY, USA: ACM Press. doi: 10.1145/2397696.2397711

Lottridge, D., & Chignell, M. (2009). Emotrace: Tracing Emotions through Human-Product Interaction. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting HFES'09*.

Mandryk, R.L., Inkpen, K., & Calvert, T. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour and Information Technology*, *25*(2), 141-158. doi: 10.1080/01449290500331156.

Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, Fall 1992, 62* (3), 279-300.

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

McKay, D., Cunningham, S., & Thomson, K. (2006). Exploring the user experience through collage. In *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction CHINZ '06*, 109–115. New York, NY: ACM Press.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography:synthesising qualitative studies*. London: Sage.

Norman, D. (2004). *Emotional Design: Why we love (or hate) everyday things*. New York: Basic Books.

Norman, D., Miller, J., & Henderson, A. (1995). What You See, Some of What's in the Future, And How We Go About Doing It: HI at Apple Computer. *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI'95*, Denver, Colorado, USA.

Obrist, M., Tuch, A., & Hornbæk, K. (2014). Opportunities for Odor: Experiences with Smell and Implications for Technology. *Proceedings of ACM Conference on Human-Factors in Computer Systems CHI '14*. New York, NY: ACM Press.

Odom, W., & Lim, K.-Y. (2008). A Practical Framework for Supporting "Third-Wave" Interaction Design. Proceedings of alt.chi 2008, *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM Press.

Peedu, G. (2011). *Enhancing Public Service User Experience in Information Society*. (master's thesis). Tallinn University, Institute of Informatics: Estonia. Retrieved from: www.cs.tlu.ee/teemaderegister/get_file.php?id=108.

Pine, B.J., & Gilmore, J.H. (1998). Welcome to the Experience Economy: work is theatre & every business a stage. *Harvard Business Review*, July-August 1998.

Raita, E., & Oulasvirta, A. (2011). Too good to be bad: Favorable product expectations boost subjective usability ratings, *Interacting with Computers*, *23*, 363-371. doi: 10.1016/j.intcom.2011.04.002

Regal, G., Busch, M., Hochleitner, C., Wokerstofer, P., & Tscheligi, M. (2014). TACTUX: a tactile user experience assessment board. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction NordiCHI'14*, New York, NY: ACM Press. doi: 10.1145/2639189.2670256

Roedl, D. J., Stolterman, E. (2013). Design research at CHI and its applicability to design practice. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'13*, New York, NY: ACM Press. doi: 10.1145/2470654.2466257

Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.

Rogers, Y. (2004). New theoretical approaches for human-computer interaction. In B. Cronin (Vol. Ed.), *Annual review of information, science and technology: Vol. 38* (pp. 87-143). Medford, NJ: Information Today.

Rogers, Y. (2009). The Changing Face of Human-Computer Interaction in the Age of Ubiquitous Computing. *Proceedings of USAB'09,* 1-19. doi: 10.1007/978-3-642-10308-7_1

Rogers, Y. (2011). Interaction design gone wild: striving for wild theory. *Interactions*, *18*(4), 58–62.

Rogers, Y., Connelly, K., Tedesco, L., Hazlewood, W., Kurtz, A., Hall, R.E., Hursey, J., & Toscos, T. (2007). Why it's worth the hassle: the value of in-situ studies when designing Ubicomp. In *Proceedings of the 9th international conference on Ubiquitous computing UbiComp'07*. Berlin, Heidelberg: Springer-Verlag, 336-353.

Roto, V., Ketola, P., Huotari, S. (2008). User Experience Evaluation in Nokia. Now Let's Do It in Practice. *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. New York, NY: ACM Press. doi: 10.1145/1358628.1358967

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (2011). User Experience White Paper: Bringing clarity to the concept of user experience. *Result from Dagstuhl Seminar on Demarcating User Experience*, Sept. 15-18, 2010, Finland. http://www.allaboutux.org/uxwhitepaper

Roto, V., Obrist, M., Väänänen-Vainio-Mattila, K. (2009). *User Experience Evaluation Methods in Academic and Industrial Contexts.* In User Experience Evaluation Methods in Product Development (UXEM'09). Workshop in Interact'09, Sweden.

Russell, J. A., Weiss, A., & Mendelsohn, G. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*(3), 493–502.

Ryan, R. (1982). Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. *Journal of personality and social psychology 43*(3), 450–461.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 693-727. doi: 10.1177/0539018405058216

Schneiderman, B. (2008). Science 2.0. *Science*, *319*, 1349-1350.

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.

Simsarian, K. T. (2003). Take it to the next stage: the roles of role-playing in the design process. *Extended Abstracts on Human Factors in Computing Systems (CHI EA'03).* New York, NY, USA: ACM Press. doi:10.1145/765891.766123

Spector, P.E. (1992). *Summated rating scale construction. An Introduction*. Newbury Park: Sage.

Stolterman, E. (2008). The nature of design practice and implications for interaction design research. *International Journal of Design, 2*(1), 55-65.

Swallow, D., M. 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 EACE* '05, 91–98. New York, NY: ACM Press.

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

Tractinsky, N. (2013). Replicating and extending research on relations between visual aesthetics and usability. *Proceedings of the CHI2013 Workshop on the Replication of HCI Research*. Paris, France, April 27-28, 2013.

Tractinsky, N., Katz, A., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*(2), 127-145. doi: 10.1016/S0953-5438(00)00031X

Tscheligi, M., Isbister, K., Höök, K., Obrist, M., Busch, M., & Hochleitner, C. (2014). "Touch me" – Workshop on Tactile User Experience Evaluation Methods. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, New York, NY: ACM Press. doi: 10.1145/2559206.2560471

Tuch, A., Trusell, R., & Hornbæk, K. (2013). Analyzing users' narratives to understand experience with interactive products. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2079-2088). Paris, France: ACM. doi:10.1145/2470654.2481285

Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., & Hella, J. (2009). SUXES – User Experience evaluation method for spoken and multimodal interaction. In *Proceedings of Interspeech 2009*.

Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008). Now let's do it in practice: user experience evaluation methods for product development. *CHI '08 Extended Abstracts on Human Factors in Computing Systems,* New York, NY: ACM Press.

Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008). Towards practical user experience evaluation methods. *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland

Väänänen-Vainio-Mattila, K., & Segerståhl, K. (2009). A Tool for Evaluating Service User eXperience (ServUX): Development of a Modular Questionnaire. User Experience Evaluation Methods, *UXEM'09 Workshop at Interact'09*, Uppsala, Sweden.

Väänänen-Vainio-Mattila, K., & Wäljas, M. (2009). Development of Evaluation Heuristics for Web Service User Experience. *Extended Abstracts on Human Factors in Computing Systems (CHI EA'09)*. New York, NY, USA: ACM Press. doi: 10.1145/1520340.1520554

Vallerand, R.J. (1989). Vers une méthodologie de validation transculturelle de questionnaires psychologiques : implications pour la recherche en langue française. *Psychologie Canadienne*, *30*(4), 662-689.

Van Boven, L., & Gilovich, T. (2003). To do or to have? That is the question. *Journal of Personality and Social Psychology*, *85*, 1193-1202. doi:10.1037/0022-3514.85.6.1193

Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89-99.

Van Schaik, P., & Ling, J. (2008). Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. *Interacting with computers, 20*, 419-432. doi: 10.1016/j.intcom.2008.03.001

Venturi, G., & Troost, J. (2004). Survey on the UCD integration in the industry. In *Proceedings of the 3rd Nordic Conference on Human-Computer Interaction (NordiCHI'04)*. New York, NY: ACM Press, 449-452. doi:10.1145/1028014.1028092

Vermeeren, A, Law, E., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User Experience Evaluation Methods: Current State and Development Needs. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, New York, NY: ACM Press. doi :10.1145/1868914.1868973

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 Fourth Nordic Conference on Human–Computer Interaction, pp. 74–78.

Vredenburg, K., Mao, J.-Y., Smith, P.W., and Carey, T. (2002). A Survey of User-Centered Design Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'02)*. New York, NY, USA: ACM Press, 471-478. doi:10.1145/503376.503460

Vyas, D., & van der Veer, G. C. (2006). Experience as meaning: some underlying concepts and implications for design. *Proceedings of the 13th European conference on Cognitive ergonomics ECCE'06*. Zurich, Switzerland: ACM Press.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070.

Wilson, M.L.L., Resnick, P., Coyle, D., & Chi, E.H. (2013). RepliCHI: The Workshop. *CHI '13 Extended Abstracts on Human Factors in Computing Systems (*pp. 3159–3162), New York, NY, USA: ACM Press.

Winckler, M., Bernhaupt, R., Forbrig, P., & Sauer, S. (2015). IFIP WG 13.2 Workshop on User Experience and User-Centered Development Processes. To be held in September 14th 2015, Bamberg, Germany - in conjunction with IFIP TC3 INTERACT conference.

Wölfel, C., & Merritt, T. (2013). Method card design dimensions: a survey of card-based design tools. In P. Kotzé et al. (Eds.) *Lecture Notes in Computer Science: Vol. 8117*. INTERACT 2013, Part I (pp. 479-486). doi: 10.1007/978-3-642-40483-2_34

Woolrych, A., Hornbæk, K., Frøkjær, E., & Cockton, G. (2011). Ingredients and meals rather than recipes: a proposal for research that does not treat usability evaluation methods as indivisible wholes. *International Journal of Human-Computer Interaction*, *27*(10), 940-970. doi: 10.1080/10447318.2011.555314

Xu, D., Read, J., Sim, G., McManus, B., & Qualter, P. (2009). Children and 'smart' technologies: can children's experiences be interpreted and coded? In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers BCS HCI'09*, 224–231. Swinton, UK: British Computer Society.

2 RESEARCH APPROACH

« *Our community is driven to publish novel results in novel spaces using novel designs, and to keep up with evolving technology* »
- Max L. L. Wilson (2013)

UX has first been challenged and thought of as a temporary buzzword or trend, before being eventually considered by the community as a concept and field worth of sound research. Recent years have witnessed the flourishing and fruitful development of new methods and tools (Vermeeren et al., 2010). However, as shown in our introductory chapter, UX research on evaluation and design methods suffers from several limitations. We believe that one way for the research community to overcome these challenges is to foster and strengthen research consolidation practices.

1. BACKGROUND AND MOTIVATION

When we started working on this dissertation and digging into the intriguing topic of UX, we soon realized that based on existing litterature we would face several methodological challenges along the research process. We were initially interested in consolidating research drawing on the links between psychological needs and UX. We therefore envisioned conducting experiments to test hypotheses about the impact of psychological needs fulfilment on UX. However, being able to conduct this work required addressing a series of preliminary methodological issues, which we explain in the following.

First, we needed to better understand UX as a concept and how it was used in both research and practice. To do so, we wanted to rely on an existing study in order to be able to analyze potential evolutions over time in the understanding of UX. We achieved this objective by replicating an international survey on UX, reported in study A, safeguarding that our methods are grounded on actual practitioners' viewpoints.
The second methodological issue we had to face regarded the availability of adequate tools and scales to gauge the perceived quality of UX and to correlate this evaluation with an assessment of psychological needs. While the AttrakDiff scale seemed to fulfil our requirements and was already used in several studies on the topic, we encountered three main obstacles: (1) the scale was not available in French, which is the main language of the sample we wanted to study (2) the scale's validity (both in its original German language or in English) was not rigorously reported in the publications (3) the scale did not provide an explicit measure of contextual factors, which we were additionally interested in. We therefore decided to conduct two studies. On the one hand we translated the AttrakDiff scale into French, following a thorough cross-cultural validation methodology. This is the focus of Study B. On the other hand, we designed a summated rating scale – the UX context scale (UXCS) - to collect subjective data on contextual factors. The scale validation process is presented in Study C. In both cases, we complied with highest scientific standards in order to avoid

methodological limitations imposed on numerous published studies (Chapter 1, section 4).
As many researchers in the field, we then also faced the issue of selecting relevant UX evaluation methods. We wondered at this point if established HCI methods – that were also proven to be majorly used in practice - would support a valid assessment of UX. We therefore chose the two most widely used methods (i.e., expert evaluation and user testing in a laboratory setting) and applied them to the evaluation of UX and the impact of need fulfilment on this evaluation (Study E). We used the French version of the AttrakDiff scale and our UX context scale to investigate UX-related factors. However, with regards to psychological needs, we lacked two research tools: a pragmatic tool used by experts to conduct their expert evaluation, and a self-reported questionnaire on needs fulfilment to support users' evaluations after the testing session. The latter already existed in English, initially developed by Sheldon et al. (2001) within the field of positive psychology. We developed the former under the form of UX Cards (Study D).

The more we progressed with this project, methodological consolidation appeared as a necessity to inform studies on UX. Replication studies, statistical validation of quantitative measures, and rigorous translations of evaluation tools; all these essential contributions are too rarely addressed by the UX research community. A main question thus guided our work: What processes and practices we can provide might contribute to the consolidation of UX research on methods?

2. RESEARCH OBJECTIVES

The research papers included in this dissertation serve the common goal of contributing to consolidating UX research, with a particular focus on UX design and evaluation methods. To address this goal, six main methodological studies were conducted. Using different approaches and methods, these studies represent a variety of perspectives informing UX research consolidation. This dissertation includes a replication study (Paper A), the translation and validation of a tool (Paper B), the development of new design and evaluation methods (Paper C and D), the empirical assessment of the relevance of established HCI methods for the evaluation of UX (Paper E) and finally an investigation on how to bridge UX research and practice through a design approach (Paper F).

As can be seen from the background and motivations behind our work, we followed an iterative and incremental research process, where each step contributes to methodologically consolidating UX. Consolidation thus represents an overall research objective to our project. While each study informs more precise research questions, their overarching common denominator is informing a strive for consolidation. For a first overview of research questions, refer to section 4 *Research outline*.
We also intend to support the development of consolidation practices by raising awareness for this crucial need.

3. RESEARCH CONTRIBUTIONS

Considered one by one, each study contributes to the advancement of the body-of-knowledge on UX.

Our research contributions at the theoretical level thereby encompass:

- an analysis of practitioners' perspectives on UX and the evolutions over time in the understanding and practice regarding this concept (study A)
- a cross-cultural validation of Hassenzahl's UX model (study B)
- first insights into how contextual dimensions impact UX (study C)
- relevant findings on the links between psychological needs and UX (studies D and E)

Direct methodological outputs include:

- a French version of the AttrakDiff scale, supporting an assessment of the perceived pragmatic quality, hedonic quality and attractiveness of an interactive product (study B)
- the design of the UX context scale (UXCS), a self-reported questionnaire for the evaluation of the contextual dimension impacting user experience (study C)
- the design of the UX Cards, a pragmatic tool able to support psychological needs-driven UX evaluation and design (study D)

When considered altogether, the papers additionally inform research consolidation strategies that the research community might learn from. To do so, we discuss the benefits, challenges and limitations of the consolidation strategies used in this thesis. Beyond the contributions of our six studies, this work thereby allows for a critical discussion on the necessity of consolidation in UX research.

The originality of this thesis stems from three main points: (1) adopting a high-level perspective on UX design and evaluation methods (2) exploring several UX research consolidation strategies, thereby (3) enabling a critical discussion on the necessity of consolidation in UX research (4) informing consolidation in an unbiased way. Our work also often reflects our concern of being relevant to UX practice.

4. RESEARCH FRAMEWORK AND METHODOLOGIES

This dissertation falls within the scope of psychology and Human-Computer Interaction. Our research follows a user-centered approach, defined by Forlizzi and Battarbee (2004) as a way to "understand people's actions, and aspects of experience that people will find relevant when interacting with a product" (p. 262).

In this thesis, we mainly addressed the topic of UX from a methodological perspective. Even though the present work might seem to be more focused on methods from an evaluation perspective than from a design perspective, we did not clearly separate design and evaluation. As highlighted by Scapin

and colleagues (2012), this distinction is unpractical in a UX iterative development process where continuous evaluation informs UX design and is integrated as part of design and redesign processes.

Although each of the six studies described in this manuscript entails a specific research methodology, they share the common point of mixing several research methods in order to provide a comprehensive view of UX. Our choice to investigate the topic of UX using a mixed-method approach is explained by the complementarity of both qualitative and quantitative perspectives in accounting for the complexity of UX.
Our approach also attempts to conciliate scientific requirements and relevance to practice and to explore the topic of UX using sound scientific methodologies.

5. RESEARCH OUTLINE

This thesis includes the following research papers (Table 1). The papers are preceded by an introductory section, which aims at introducing the general research topic (Chapter 1), the research objectives and the contributions of the thesis to the research area (Chapter 2). Each study contributes to the consolidation of UX evaluation and design methods from a different perspective (Table 1).

Table 1. Papers included in this manuscript

Paper A	User Experience: a concept without consensus? Exploring practitioners' perspectives through an international survey	Replication study
Paper B	Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs	Translation and validation
Paper C	The contextual dimension of user experience: development and application of the user experience context scale (UXCS)	Scale development
Paper D	UX Cards – Experience design through psychological needs	Tool development
Paper E	How UX alters established HCI evaluation methods? Expert versus laboratory evaluation	Analysis of methods transferability
Paper F	Enhancing the Design Process by Embedding HCI Research into Experience Triggers	Bridging research and practice

Paper A presents a direct replication of a UX survey. In this paper, we conducted an international survey on User Experience and collected data from 758 practitioners and researchers from 35 nationalities. This study is a replication of a previous survey, which was conducted in 2008 (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009) on a smaller sample and was published under the form of a conference paper only. As UX has grown into a key concept in the field of Human-Computer Interaction, it was worthwhile exploring practitioners' perspectives and analyzing potential evolutions over time in the understanding and practice regarding this concept. Our results highlight that, while some statements about UX have reached consensus, substantial differences in the understanding of UX exist between practitioners from different backgrounds, mainly opposing Industry and Academia. We also provide the readers with a topic-based analysis on UX and a comparative analysis between the original and the

replicated survey, allowing them to understand the trends and changes in UX practice.

Paper B describes the translation and validation of a French version of the AttrakDiff scale. While UX evaluation is a core concern within the field of Human-Computer Interaction (HCI), there is currently no valid self-administered UX evaluation tool in French. Following the cross-cultural methodology developed by Vallerand (1989), the questionnaire was translated by trilingual researchers before being back-translated and validated by a panel of experts. A pre-test was conducted on 26 participants. The characteristics of the French version of the AttrakDiff scale were then evaluated through a quantitative online study involving a sample of 381 users. The results confirm the expected 3-factors structure and a good internal consistency of each subscale. The links between factors are consistent with Hassenzahl's theoretical model (2003) where pragmatic and hedonic perceived attributes combine to form a judgment of attractiveness. The current French version of the AttrakDiff scale is majorly reliable with regards to the initial German version and presents satisfactory levels of validity and reliability. Some problematic items are discussed with regards to the characteristics of the use case chosen for the validation study.

Paper C reports on the development and first validation of the User Experience Context Scale (UXCS), a new tool for the evaluation of the contextual dimension impacting user experience. While the context of use has been highlighted for many years as a key factor impacting UX, so far there is no tool based on self-reported quantitative data that could support an explicit context investigation, relevant to both design practice and research purposes. With the ever-growing trend for mobile and innovative technologies, knowing how the interaction context impacts end-users' experience is of prime importance. The UXCS is based on a review of relevant literature and a fine-grained categorization of contextual dimensions. The UXCS supports both a measure of objective contextual factors and a measure of perceived context and has been built based on a thorough process following best practices for summated rating scale construction. The scale has been validated through an online study involving 137 participants. Beyond all aspects related to the construction of the UXCS, our results also provide better insight on the links between UX and context, therefore concretely showing the potential benefits of the UXCS for UX research and practice.

Paper D presents the iterative development of the UX cards as a pragmatic tool able to support psychological needs-driven UX evaluation and design. While UX has become a key concern of product development, designing for UX or evaluating UX still remains a challenge. To close the gap between UX research and industry, practical methods need to be developed and current UX research methods need to be adapted to the requirements of industrial UX development. The psychological needs-driven UX approach is a well-explored area in UX research and appears to be a powerful framework for the design of more experiential interactive systems. However, the transfer from UX research to practice is difficult and slow and this specific approach is not yet widely used by UX practitioners. As card-based methods have been shown to support designers in both the generation of ideas and the evaluation of their designs, we created the UX

cards as a pragmatic tool able to support psychological needs-driven UX evaluation and design. In this paper, we present the iterative development of the UX cards, which might be used for design, evaluation or training purposes. We report on an experiment involving 33 UX experts and aimed at validating the use of the UX cards for UX evaluation. We also present two idea generation techniques to be used for UX design using the UX cards. Our findings suggest that the UX cards are a valuable tool able to support psychological needs-driven UX design or evaluation.

Paper E investigates how UX alters established HCI methods through two complementary experiments. We selected two widely used HCI methods - expert evaluation and user testing in a laboratory setting - and applied them to the evaluation of UX using a psychological needs-driven approach. In the first experiment, we asked thirty-three UX experts to perform a UX expert evaluation on four given interactive systems. For each system, we collected data on the predicted fulfilment of seven UX needs as assessed by the experts. We also used observation and interview to detect the issues encountered by practitioners when conducting an expert evaluation of UX. In the second experiment, seventy users were asked to evaluate their experience with two given systems (the same ones than two of the systems used for the expert evaluation), by filling out the AttrakDiff scale and a UX needs fulfilment questionnaire. Qualitative data was also collected for each participant through a think aloud protocol and debriefing interview. We then compared the results of both experiments in order to assess the quality and limitations of each method for the evaluation of UX.
These results are discussed from a methodological perspective by analysing, on the one hand, what evaluation practices remain unchanged yet effective, valid and reliable for the study of UX and, on the other hand, what are the challenges and limitations of established HCI methods for the evaluation of UX. By highlighting why and how UX alters widely used evaluation methods, this paper offers a discussion on the way we evaluate UX and reviews alternative paradigms for UX evaluation.

Paper F is a position paper describing a work-in-progress, which is the result of collaboration between psychology and design. It represents both a reflection on the existing gap between UX research and practice and a creative attempt to build bridges under the form of science-based UX design. A special focus of this paper is to translate theoretical work into experiential objects (or situations) called "Experience Triggers". Through their materiality, these artefacts bring emotions and sensations to the design process and designers can immerge into and understand the theories on experience. As a consequence of this immersion, the final product designed by the team is assumed to be more experiential. Experience Triggers are introduced here as a potential new tool for science-based UX design.

Finally, the third part of this manuscript concludes the research described in this thesis by reflecting on its theoretical, methodological, and practical contributions and sketching directions for future research.

Part II

STUDIES

USER EXPERIENCE: A CONCEPT WITHOUT CONSENSUS? EXPLORING PRACTITIONERS' PERSPECTIVES THROUGH AN INTERNATIONAL SURVEY

Carine Lallemand, Guillaume Gronier, & Vincent Koenig

Computers in Human Behavior (2015), Vol 43, pp. 35-48
doi: 10.1016/j.chb.2014.10.048

Keywords: user experience, survey, definition, concept, replication study

ABSTRACT

For more than a decade, User Experience (UX) has grown into a core concept of Human-Computer Interaction (HCI). Practitioners and researchers from a wide range of disciplines are daily working with this concept. However, despite many attempts to understand, define and scope UX, one may still wonder whether a consensus has been reached on this concept. In a willingness to address the complexity of this research topic and bring the concept of UX to maturity, a replication of an international survey has been conducted.

The main goal of the present study is to get a better understanding of practitioners' viewpoints on the notion of UX and to analyse potential evolutions over time in the understanding and practical use of the concept. As both practical and theoretical implications of UX are of the greatest importance for whoever designs interactive systems, the exploration of practitioners' perspectives is a valuable step toward continual improvement of UX activities. The present survey has been conducted amongst 758 practitioners and researchers from 35 nationalities. It allows to better understand how this concept is understood and used throughout the world. Amongst interesting results, important differences were observed according to the geographical location and background of the respondents.

1. INTRODUCTION

Some concepts in the field of HCI are commonly used by practitioners even if a lack of empirical research has prevented their full understanding and impact. User experience (UX) could be one of those fashion and fuzzy terms that is increasingly used even though no clear consensus has been reached yet regarding its definition or scope. While some authors question the added value of UX compared to established concepts such as usability, ergonomics or user acceptance (Barcenilla & Bastien, 2009), some also agree that UX is a "truly extended and distinct perspective on the quality of interactive products" (Hassenzahl, 2008).

Since the 2000s, the concept of UX is widely used but understood in different ways (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009). Many definitions and models have been proposed (Forlizzi & Ford, 2000, Hassenzahl, 2003; Karapanos et al., 2010) without resulting in a true consensus. According to Law et al. (2009), this lack of conceptual clarity can be explained by the fact that UX is associated with a wide range of fuzzy and dynamic concepts and is used as a generic term combining several HCI notions. Understanding UX thus appears as an important challenge for HCI as it constitutes a first conditional step toward UX measurement and design (Law, Vermeeren, Hassenzahl, & Blythe, 2007). As stated by Fenton and Pfleeger (1997): "you cannot control what you cannot measure and you cannot measure what you cannot define" (p. 14). One of the three pillars of the UX Manifesto, published in 2007 (Law et al., 2007) therefore consisted in answering the question "What is UX?", in particular by studying the basic concepts and assumptions related to UX. Several studies have tried to meet this challenge since. Indeed, research attempts to understand UX have been made following two main approaches: reviewing UX research on the one hand (Bargas-Avila & Hornbæk, 2011; Law et al., 2014), and interviewing or surveying UX professionals on the other hand (Law et al., 2009, Law et al., 2014; Tokkonen & Saariluoma, 2013). This paper follows the approach of understanding UX by surveying UX professionals. By replicating a UX survey conducted in 2008 by Law et al. we explore what UX is according to UX practitioners.

In the first and upcoming section, we will briefly describe the origins of UX and the attempts made by the research community to understand, scope and define UX. In the second section of the paper, we will describe the methodology used to conduct our replication study. In the third and fourth sections, we will present the results of the survey and summarize the main findings in a topic-based analysis. The fifth section will focus on the comparison between the results of the present study and the results of the original study conducted in 2008. Finally, we will discuss the challenges, successes and limitations of this replication study.

2. USER EXPERIENCE: A "TRULY DISTINCT PERSPECTIVE" BEYOND USABILITY?

In the 1990s, Donald Norman was amongst the first authors to use the term "User Experience" in order to describe all aspects of a person's experience with a system (Norman, Miller, & Henderson, 1995). Norman explains he introduced the term UX because he believed "usability" to be too narrow to represent a holistic vision of human–computer interactions. Nearly at the same period, Alben's paper entitled "quality of experience" (Alben, 1996) put the focus on users' sensations, their understanding of how things work, their feelings during usage, the achievement of their goals and also on the overall interaction context.

At the theoretical level, UX relies on several trends (Rogers, 2012): activity theory (Kuutti, 1996), distributed cognition (Hollan, Hutchins, & Kirsh, 2000), but also usability studies (Nielsen, 1993; Shackel, 1991) and emotional design (Jordan, 2002 and Norman, 2004). Activity theory and distributed cognition have outlined a comprehensive view of user experience as a complex and socially situated phenomenon where technology acts as a mediator between the user and the activity. The user is therefore not considered as a distinct entity, but rather dependent from the context of the entire system, including the environment, the user characteristics and also the technological objects and tools. Interactive systems provide rich and complex functionalities and the quality of UX will of course also rely on their usability. Numerous usability studies from the 1980s (Nielsen, 1993; Shackel, 1991) are indeed the fundamental bases on which the field of UX is grounded. Usability refers to "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, 1998). In major UX models (Hassenzahl, 2003; Mahlke, 2008), the usability concerns for effectiveness and efficiency were included as "pragmatic" or "instrumental" qualities of a system, whereas the notion of satisfaction has been extended to the one of "hedonic" system's quality. Finally, emotional design (Jordan, 2002 and Norman, 2004) has helped to formalize the link between UX and emotion generation. Indeed, pleasure and emotions generated by the use of products or technological systems, and even the satisfaction of basic needs through technological interactions (Hassenzahl, 2010), are major research topics within the field of UX.

Across the numerous definitions that have been proposed (Desmet and Hekkert, 2007, Hassenzahl and Tractinsky, 2006 and Law et al., 2009), researchers and practitioners agree that UX is the result of the interaction between three elements: the user, the system and the context (Roto, Law, Vermeeren, & Hoonhout, 2011). Following this common vision in the field of HCI, 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*" (p. 95). Despite sharing many common grounds with the concept of usability, UX expands it by including emotional, subjective and temporal aspects involved in the interaction. As stated by McCarthy and Wright (2004): "*Experience of technology refers to something larger than usability or one of its dimensions such as satisfaction or attitude."* (p. 6) While the concept of usability mainly focuses on an objective approach of

the interaction, UX is also exploring subjective factors characterizing the experience between human and technology. Some authors even described UX as a "counter-movement to usability thinking" (Gegner, Runonen, & Keinonen, 2011). However, usability concerns are generally included as parts of UX, often under the title "pragmatic" (Hassenzahl, 2003) or "instrumental" (Mahlke, 2008) aspects of the interaction. UX might therefore offer a holistic approach to understanding human–computer relationship and its underlying experience.

At a higher level, usability and UX are both considered part of User-Centred Design (also called Human-Centred Design), which is defined as "an approach to interactive systems development that aims to make systems usable and useful" (ISO 9241-210, 2010). This process has been first formalized as an ISO standard in 1999 (ISO 13407, 1999) and was at that time only focused on usability. However this major usability standard has been updated in 2010 for the user experience era (ISO 9241-210, 2010) and now includes the concern for UX as one of the six key principles that will ensure that a design is user-centred. More recently, the ever-growing interest and focus on UX has given birth to a novel design process. While User-Centred Design designates the process of designing usable and useful technologies, the term "Experience-Driven Design" or "Experience Design" (Hassenzahl, 2010) is now used to designate the process of designing for UX. The objective of Experience Design is to "is to bring the resulting experience to the fore – to design the experience before the product." (Hassenzahl, 2013, Section 3.4). Despite the use of a distinct terminology, Experience Design is not mutually exclusive with traditional User-Centred Design. The user's perspective still remains the central reference of the development process in both cases. As UX goes beyond usability by bringing experiential aspects into the process, similarly, Experience Design goes beyond User-Centred Design by putting more emphasis on the quality of the experience as felt by the user.

As academic and business contexts seem to rely on different practices, understanding the field of UX might be done either by reviewing UX Research or by interviewing or surveying UX Practitioners. On the one hand, from an academic perspective—despite the existence of a variety of perspectives and approaches—some commonalities on UX have nevertheless been summarized in a valuable UX White Paper (Roto et al., 2011). At a generic level, the research community agrees that UX as a phenomenon designates the experience of using an interactive system. Based on the literature, we can assume this experience to be unique to an individual and influenced by several factors, encompassing prior experiences and expectations based on those experiences (Hassenzahl, 2008; Roto et al., 2011). The social and cultural context of the interaction also plays an important role by impacting the felt experience: "UX may change when the context changes, even if the system does not change." (Roto et al., 2011, p. 10). Finally, the temporal dimension of UX was also highlighted in prominent publications in the field of HCI (Karapanos et al., 2010 and Roto et al., 2011), which showed that UX starts way before the direct interaction with a product (e.g., through advertising, experience of related technologies or reading of users' online reviews) and does not stop just after usage. Time spans of UX have therefore been identified to help thinking about the dynamics of UX (Karapanos et al., 2010; Roto et al., 2011). On the other

hand, from a business perspective, studies have shown the heterogeneity of views on the nature and scope of UX (Law et al., 2009, Law et al., 2014; Tokkonen & Saariluoma, 2013). Practitioners nevertheless tend to agree on the fact that UX is a subjective and dynamic concept, influenced by several contextual factors (Law et al., 2009). Respondents' background variables were rarely able to significantly explain variations in views on UX (Law et al., 2009). In the next section of this paper, we will provide rationale and further details for our survey on UX Practitioners around the world, done by replicating a previous survey conducted in 2008 (Law et al., 2009).

3. A REPLICATION OF THE SURVEY "UNDERSTANDING, SCOPING AND DEFINING USER EXPERIENCE"

We decided to replicate a previous survey entitled "understanding, scoping and defining UX: a survey approach" (Law et al., 2009), for gaining a better insight into how the UX concept might have evolved over time and how it shapes practice. The original study has been first spread during the main conference CHI'08 before being further broadcast through additional communication channels. Results have been published the following year in the proceedings of CHI'09, as a 10-pages long paper. 275 answers had been collected at that time from 25 countries.

In order to facilitate multicultural participation and to reach a wider audience within the French-speaking community of UX practitioners and researchers, we translated all questionnaire items from the English master version to both French and German. A back translation process has been applied to ensure the quality and validity of the process. More than 758 valid answers have been collected from all over the world. Preliminary results have been published in 2013 (Lallemand, Koenig, & Gronier, 2013).

3.1. RATIONALE FOR A REPLICATION

Several reasons explain our choice to replicate this survey. First of all, as UX is still a concept in maturation, it is worth taking stock of the situation four years after the initial study in order to observe a potential evolution in representations, points of view and practices associated to UX. Replication acts here as a way to check whether the results still apply in a different context to the original study, especially in a different temporality.

Moreover, the translation into two other languages allowed us to reach a wider and more diverse audience, especially in the multicultural context in which the present work took place. As this study constituted an exploratory step within a wider Luxemburgish project focused on UX Design, gathering additional knowledge about the French- and German-speaking practitioners' community (not well represented in the initial study) seemed crucial to us. By trying to draw an accurate picture of the current situation of UX and building on that basis, we ultimately aim at better methodologies, frameworks and metrics to design for UX.

3.2. FORM OF REPLICATION

This study may be considered as a direct replication, since differences between both studies are limited to:

- A minor extension through the translation into French and German languages. The original English version was kept as default language and still represented 58.4% of the completed surveys.
- Additional items on sociodemographics aimed at better categorizing participants and acting as control variables to analyze the data.

3.3. STRUCTURE OF THE SURVEY

The UX questionnaire encompasses 3 sections:

(1) Background: in this first section, respondents were asked to answer 13 questions about their job and educational background, their level of familiarity with UX or the importance of UX in their current job. Finally, sociodemographic information (age, gender, country of residence) was also collected. As mentioned before, one of the main differences between the initial study and its replication consists of additional sociodemographic questions allowing better categorizing respondents. The following questions have therefore been added to the initial survey: current job position, level of familiarity with the concept of UX and collaboration with people working in the field of UX (in case UX was not central at all to their own job).

(2) UX Statements: respondents were asked to assess their agreement level with 23 UX Statements on a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Possibility was also given to select the answer "I don't understand" when they felt that a statement was not clear enough for them to take a stand. Statements were formulated similarly to the following example: "*User experience should be assessed while interacting with an artifact*". See complete list of statements in Section 4.2.

(3) UX Definitions: Five UX definitions were presented (Table 1). For each of them, respondents were asked to answer the following open-ended question "What do you think of this definition?" Finally, respondents were asked to choose which definition suits them best ("If you had to pick one of these UX definitions, which one would it be?") and to freely comment on the reasoning for their choice. These open-ended questions aimed at providing participants with a free space to express their view on the topic and to discuss the definitions. It was expected to provide us with valuable information on how participants define and scope UX. It allows for both identifying a range of issues not previously conceptualized through the UX Statements and knowing what the requirements are for a useful UX definition.

Table 1. UX Definitions used, as drawn from original survey.

D1	All aspects of the end-user's interaction with the company. Its services and its products. The first requirement for an exemplary user experience is to meet the exact needs of the customer without fuss or bother. Next comes simplicity and elegance that produce products that are a joy to own, a joy to use. True user experience goes far beyond giving customers what they say they want, or providing checklist features. [Nielsen & Norman Group, nngroup.com]
D2	A consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.) the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.) [Hassenzahl & Tractinsky, 2006]
D3	The entire set of affects that is elicited by the interaction between a user and a product including the degree to which all our senses are gratified (aesthetic experience) the meanings we attach to the product (experience of meaning) and the feelings and emotions that are elicited (emotional experience). [Desmet & Hekkert, 2007]
D4	The value derived from the interaction(s) [or anticipated interaction(s)] with a product or service and the supporting cast in the context of use (e.g. time, location, and user disposition). [Sward & MacArthur, 2007]
D5	The quality of experience a person has when interacting with a specific design. This can range from a specific artefact such as a cup toy or website up to larger integrated experiences such as a museum or an airport. [UXnet.org]

3.4. SAMPLING AND DIFFUSION OF THE SURVEY

The survey was broadcast online from February to April 2012, on multiple advertisement channels. As for the original study, practitioners' forums, social networks and mailing lists were the main vector of dissemination.
With regard to sampling requirements, this exploratory survey did not involve a strictly random and representative sample. As the whole population of practitioners working in a field related to UX is not clearly defined, we decided to simply broadcast the survey on the web. This allowed reaching a wide audience in line with the primary exploratory goal of the study. It also provided us with information on which kind of practitioners declare working in the field of UX. However, we are aware of several biases potentially impacting our results, especially the fact that only self-motivated and careful respondents would answer the questionnaire. Moreover, this method does not allow controlling either the absolute coverage of our advertisement or the relative coverage with regard to the target population.

3.5. PARTICIPANTS

A total of 898 questionnaires have been collected and 758 valid questionnaires have been used for computing the data. Invalid questionnaires mainly resulted from bugs in the online survey system (automatic log off after a period of inactivity) or incomplete answers. A questionnaire had to be filled out at least to the 4th question (demographics on current job position) to be considered valid, 569 participants (75% of valid questionnaires) provided their agreement level with at least one of the UX statements. 428 participants (56.5% of valid questionnaires) have picked a definition. Due to these differences in survey completion amongst our

respondents, we draw the reader's attention on the fact that sample sizes in the results analyses will vary according to the number of participants having completed each survey section. We used SPSS software 22 to perform statistical analyses.

The mean age of the sample was 35.8 years (*Min* = 21, *Max* = 70, *SD* = 9.4) encompassing 44.3% of females (*n* = 320) and 55.7% of males (*n* = 403). 58.4% of the respondents have answered the English version of the questionnaire, against 39.2% for the French version and only 2.4% for the German one. Thirty-five distinct nationalities are represented, with a majority of Europeans (61.9%), especially coming from France (34.7%). North Americans represent 25.4% of the respondents, Asians 5.7% and less than 3% of participants come from other geographical areas.

4. RESULTS

We will first provide the reader with a linear description of the main results of the survey, organized following the three questionnaire sections: Background, UX Statements and UX Definitions. In this section the reader learns more about the respondents' profile, their level of familiarity with the concept of UX or the reasons why they are interested in understanding the nature of UX. We will also describe the levels of agreement of our respondents with the 23 UX Statements. Finally, we will report on which of the five UX definitions the participants preferred and analyze their open answers to understand what they expect from a useful UX definition.

In a second section, we will explore the study's main outcomes through a topic-based analysis in order to better understand how UX professionals perceive several facets of UX. This topic-based analysis will use the survey results to answer the following questions:

- Is UX a new approach?
- What are the links between UX and previous/related fields like usability or user-centered design?
- Is there a need for a standardized definition of UX?
- How to approach UX: quantitatively or qualitatively?
- Is UX individual or social?
- What shapes UX? user-related factors? contextual factors? a temporal dynamic?

Finally, in a third section, we will compare our results to those obtained in 2008 by Law et al. (2009) and investigate a potential evolution in the understanding of the UX concept. Our respondents' agreement levels on UX statements and choice of a UX definition will be compared to the results collected in 2008.

4.1. BACKGROUND, DOMAIN AND ROLE

Regarding their educational background (*n valid* = 755), two-thirds of the respondents were primarily educated in Psychology/Social Sciences (24.1%), Technology/Software (22.3%) or HCI (16%). The remaining third encompasses Design (14.3%), and other training backgrounds (Arts,

Marketing, Business, Info/Comm., and Miscellaneous). The general profile of the respondents is presented in Table 2 using the same format as the original study in order to facilitate the comparison between both studies (presented in Section 6).

Table 2. General profile of the respondents

Variable	Frequency	Valid Percent
I work in (n valid=758):		
Industry	503	66.4%
Academia	100	13.2%
Both or between	155	20.4%
My primary role is (n valid=758):		
Researcher	128	16.9%
Consultant	199	26.3%
Manager	80	10.6%
Practitioner	278	36.7%
Student	73	9.6%
I was originally educated in the field of (n valid=758):		
Arts	37	4.9%
Design	105	13.9%
Marketing	12	1.6%
Business	30	4%
Quality / Processes	5	0.7%
Psychology / Social Sciences	164	21.6%
Technology / Software	149	19.7%
Human-Computer Interaction	116	15.3%
Other	140	18.5%
Which applies the best to your primary interest in UX? (n valid=755, 3 missing)		
I'm interested in understanding the nature of UX:		
Per se	98	13%
To design better products	392	51.9%
To better sell products	28	3.7%
To make people happier	161	21.3%
Other	76	10.1%
How central is UX to your professional work? (n valid=752, 6 missing)		
Very central	438	58.2%
Central	193	25.7%
Less central	89	11.8%
Not central at all	32	4.3%

The distribution of respondents according to their business domain or role is presented in Table 3. A majority of them are working in industry (66.4%). The distribution of the sample according to roles is quite balanced, although managers and students are less represented.

Table 3. Distribution of respondents according to their business domain and role

Role	Domain			Total	
	Industry	Academia	Both or between		
Researcher	27	60	41	128	16.9%
Consultant	159	0	40	199	26.3%
Manager	64	4	12	80	10.6%
Practitioner	243	5	30	278	36.7%
Student	10	31	32	73	9.6%
Total	**503** 66.4%	**100** 13.2%	**155** 20.4%	*N*= 758	

Without having accurate data on the distribution of people working in the field of UX across the population, the ratios related to all of these roles might reflect the reality on the field. In any case, the relatively large sample size allows us to draw conclusions related to each category of respondents.

4.1.1. LEVEL OF FAMILIARITY WITH THE CONCEPT OF UX

On average, respondents have 6.53 years of experience in the field of UX (SD = 6.45) and 6.44 years (SD = 6.24) in the overall field of User-Centered Design (both measures being obviously highly correlated with r = .83, p < .01).

The level of familiarity with the concept of UX (*n valid* = 743), self-assessed by the respondents on a 10-points scale, scored 7.91 on average (SD = 2.29). 1.3% of the respondents declared having never heard about this concept and were therefore thanked for their participation and immediately filtered out of the survey, before reaching the UX Statements section. Note that this high average level of familiarity might be explained by the open advertisement of the survey, where only the most motivated participants decided to answer the survey. It is thus safe to assume we captured personal opinions about what is important in UX for people who really are and feel involved in this topic.

We conducted one-way between subjects ANOVAs to compare the effects of domain and role on the level of familiarity with UX. The extent of familiarity was assessed as higher in Industry (M = 8.13, SD = 2.19) than in Academia (M = 6.80, SD = 2.47) or Both or Between (M = 7.90, SD = 2.29), $F(2, 740) = 14.09$, $p < .001$, $\eta^2 = .04$ (inter-group differences tested by post hoc analyses significant at $p < .001$ level). This might be explained either by the fact that scientists are more cautious when dealing with "familiarity" regarding a concept still in evolution; or maybe also because practitioners tend to use it more often in their daily practice, resulting in actually higher familiarity.

The same tendency holds for the roles: consultants are most familiar with the concept (M = 8.38, SD = 2.08), followed by practitioners (M = 8.08, SD = 2.11), managers (M = 7.69, SD = 2.52) and researchers (M = 7.58, SD = 2.47). Not surprisingly, least familiar with UX are the students (M = 6.75, SD = 2.46), $F(4, 738) = 7.94$, $p < .001$, $\eta^2 = .04$ (inter-group differences tested by post hoc analyses significant at $p < .001$ level). This observation might confirm some authors' belief (Law et al., 2014) that there is a need for enhanced education and training in UX.

Regarding the language, an independent-samples t-test was conducted to compare the effects of language on the level of familiarity with UX. The concept of UX is more familiar to English-speakers (M = 8.36, SD = 1.96) than French-speakers (M = 7.25, SD = 2.53), $t(743) = 6.69$, $p < .001$, $\eta^2 = .06$. It is worth noting that French-speaking respondents assessed, on average, their familiarity level under the global mean of the sample. Familiarity level also correlates with years of work experience (r = .35, $p < .001$), and only weakly with age (r = .12, p = .001).

4.1.2. PRIMARY INTEREST FOR UX AND IMPORTANCE OF UX

UX is considered "central" or "very central" to their professional work by 83.9% of respondents. Only 4.3% of them declare that UX is not central at all to their professional work, however 40.6% collaborate with someone working in the field of UX. We conducted one-way between subjects ANOVAs to compare the effects of role and domain on the importance of UX. Once again, UX appears less central for researchers (M = 2.16, SD = 0.86) and students (M = 2.08, SD = 1) than for managers (M = 2.54, SD = 0.76), consultants (M = 2.48, SD = 0.81) or practitioners (M = 2.44, SD = 0.83), $F(4, 747) = 6.18$, $p < .001$. Unsurprisingly, UX is considered much more central in Industry (M = 2.48, SD = 0.81) or Both or between Industry and Academia (M = 2.35, SD = 0.81) than in Academia (M = 1.93, SD = 0.89), $F(2, 749) = 17.96$, $p < .001$, $\eta^2 = .046$. All inter-group differences were tested by post hoc analyses and significant at $p < .001$ level.

Looking at the reasons why our respondents are interested in UX (Fig. 1), a majority of them are interested in UX to design better products (51.9%), and, unsurprisingly, this holds particularly true for practitioners in Industry. On the other hand, in Academia, respondents are more interested in UX per se, as an object of study ($\chi^2 = 22.56$, $p < .01$).

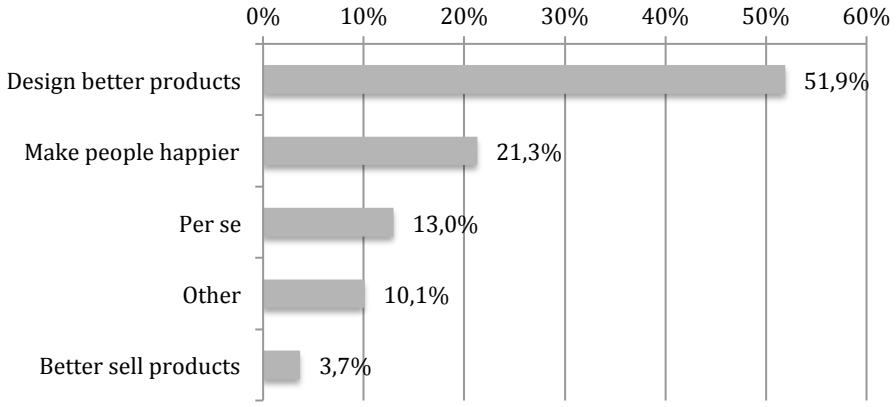


Fig. 1. Interest in understanding the nature of UX.

4.2. UX STATEMENTS

To analyze participants' answers to UX Statements (ST), we first computed descriptive statistics (minimum, maximum, mean and standard deviation) and ranked the 23 statements according to their mean level of agreement (a score of 1 being the minimum level of agreement and a score 5 the highest). Table 4 presents the UX statements sorted by mean agreement. This is what we will refer to as "agreed statements" in the following sections.

Table 4. Statements about UX sorted by mean agreement.

	N	Min	Max	Mean	SD
[ST03] Fleeting and more stable aspects of a person's internal state (e.g., needs, motivations) affect a person's experience of something	565	1	5	4.54	.63
[ST05] UX occurs in. and is dependent on the context in which the artefact is experienced	561	1	5	4.34	.91
[ST18] Designing (for) UX must be grounded in User-Centred Design	549	1	5	4.29	.83
[ST8] Prior exposure to an artefact shapes subsequent UX	552	1	5	4.21	.78
[ST12] Usability is a necessary precondition for good UX	560	1	5	4.15	.99
[ST14] Measuring UX implies determination of merits, values, and significance of an artefact in relation to a person's goals and needs	539	1	5	4.12	.76
[ST17] UX should be assessed while interacting with an artifact	556	1	5	4.02	.86
[ST13] We cannot design UX, but we can design for UX	531	1	5	3.96	1.08
[ST23] UX can change even after a person has stopped interacting with the artefact	545	1	5	3.96	.87
[ST1] UX is highly dynamic - it changes constantly while interacting with a product	576	1	5	3.93	1.01
[ST15] UX refers to affective states, i.e., any combination of valence (good - bad. pleasant – unpleasant) and physiological arousal (calm – excited)	534	1	5	3.89	.89
[ST02] Imagined use of a product can result in real experiences	524	1	5	3.80	1.06
[ST07] There is a definite need for a standardized definition of the term UX	568	1	5	3.71	1.07
[ST22] UX must be approached qualitatively	561	1	5	3.66	.99
[ST16] UX can be quantified and thus compared across similar (or competitive) artefacts	553	1	5	3.62	.96
[ST11] UX is based on how a person perceives the characteristics of an artefact but not on the characteristics per se	539	1	5	3.56	1.12
[ST06] UX is not about people's performance (ability to understand and use) in their relation with an artefact, but about the person's perception of that performance	564	1	5	3.47	1.22
[ST10] UX should be assessed after interacting with an artifact	550	1	5	3.44	1.19
[ST21] UX is not new, it is already covered by existing engineering approaches	552	1	5	3.14	1.11
[ST19] Only an individual person can have an experience. An experience is something personal. Something 'within' a person.	560	1	5	3.14	1.23
[ST09] People will never have comparable UX - each and every interaction with a product results in a unique experience	566	1	5	3.02	1.16
[ST20] UX is equal to emotional attachment	547	1	5	2.71	1.14
[ST04] UX is best viewed in terms of marketing	561	1	5	2.38	1.13

Out of twenty-three Statements about UX, seven collected a mean higher than 4 out of 5 (which corresponds to a level of agreement ranging from "agree" to "strongly agree"). Considering the statements this applies to, it seems that the respondents agree on the importance of a user's internal state, his past experiences, goals and needs, and also on the importance of the context in which the artefact is experienced. We noticed also that usability and User-Centered Design remain closely related to UX and are described as its basis. Conversely, while the majority of statements lead to an average score corresponding to a level of agreement ranging from "neutral" to "agree", the respondents disagreed on average with two of the statements. Thus, no confusion is made between UX and emotional attachment (M = 2.71, SD = 1.14). Similarly, UX is not considered as a marketing concept (M = 2.38, SD = 1.13).

The average rate of non-understandability (i.e., percentage of respondents choosing to tick "I don't understand" for a statement) was 2.84%. The level of comprehensibility can therefore be considered sufficient. As for the initial study, Statement n° 02 (ST02) collected the highest level of non-understandability with an average rate of 8.3%. No differences were found for the non-understandability rate between the three versions of the survey (one-way between subjects ANOVA), which reflects a sufficient quality of the French and German translations.

When looking at the frequency distribution of answers among the five agreement levels, it is possible to distinguish several groups of statements

according to the shape of their distribution graph. This allows a better and more accurate understanding of participants' opinions on the statements.

Fig. 2 summarizes the statements' distribution graphs according to four tendencies:

a. almost uniform and symmetrical graphs showing no clear consensus
b. skewed right graphs showing shared disagreement
c. skewed left graphs showing shared agreement
d. skewed left graphs with almost no disagreement at all showing clear agreement.

In some cases, background variables seem to impact the evaluation of these UX Statements. Independent-samples *t*-test were conducted to compare the effects of language, gender and domain on the level of agreement with one or several UX statements.

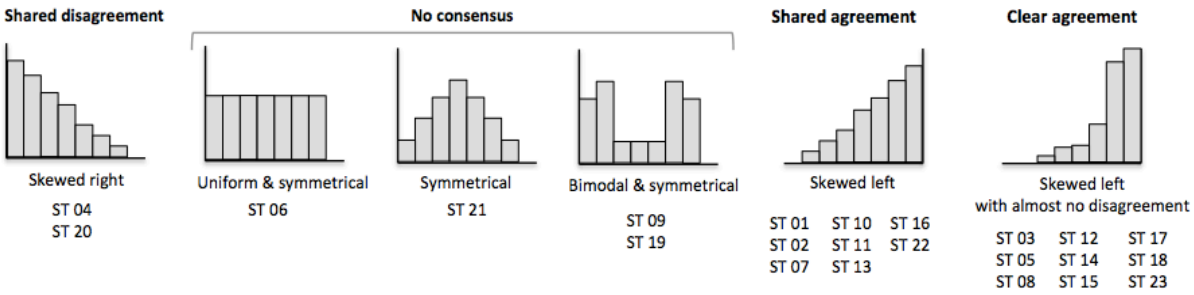


Fig. 2. UX statements (ST) sorted by frequency distribution of answers. The statements are classified according to four global types of frequency distribution: shared disagreement, no consensus, shared agreement or clear agreement.

Regarding language, 9 significant differences (out of 23 potential differences) were found (ST 04, 07, 09, 15, 17, 18, 19, 22, 23). French-speaking respondents ($M = 3.95$) felt a higher need for a standardized definition of UX than their English-speaking counterparts ($M = 3.54$), $t(545) = -4.51$, $p < .01$, $\eta^2 = .036$. Similarly, UX is more likely to be considered a marketing concept in French-speaking countries ($M = 2.69$) than in English-speaking ones ($M = 2.17$), $t(546) = -5.41$, $p < .01$, $\eta^2 = .051$. Such differences could be explained by the greater familiarity with the concept of UX that was first developed in Anglo-Saxon countries.

Significant gender differences could be observed in only 3 statements out of 23 (ST03, ST06, ST07), which confirms previous work (Law et al., 2009) by showing that opinions do not contrast sharply according to gender. Educational background significantly impacted UX Statements in 5 cases out of 23 (ST06, ST08, ST19, ST20, ST22).

Regarding the domain, some differences contrast academic and corporate environments (Industry) (5 significant differences out of 23). These relate in particular to the view of UX as a marketing concept, more pronounced in Academia (ST04; $M = 2.8$ vs. $M = 2.26$ in Industry), $t(439) = -4.03$, $p < .01$, $\eta^2 = .028$. On the other hand, Industry ($M = 2.96$) less agrees on the uniqueness of lived experience than Academia ($M = 3.21$), ST09;

$t(443) = -1.78$, $p < .05$, $\eta^2 = .006$. Surprisingly, it is rather in Industry that one considers UX should be addressed in a qualitative manner (ST22; $M = 3.70$ vs. $M = 3.51$ in Academia), $t(439) = 1.90$, $p < .05$, $\eta^2 = .005$.

Finally, bivariate correlations analyses were performed in order to study the degree of relationship between agreement with UX statements and demographic variables such as age and work experience. The number of years of work experience in the field of UX is negatively correlated with many statements (ST 04, 06, 07, 08, 11, 14, 15, 20, 22). It is worth noting that the more experienced we are, the less we consider UX from a qualitative standpoint (ST22; $r = -.18$, $p < .01$) and dependent of users' perceptions (ST11; $r = -.16$, $p < .01$). The more experienced respondents also feel less need for a standardized definition of UX (ST07;$r = -.16$, $p < .01$), probably because they are using the concept for a long time and have gone through an appropriation process. Unsurprisingly (because Age and Work Experience are highly correlated), similar findings hold true regarding the differences between respondents according to their age (7 significant correlations out of 23). We can notice negative correlations between Age and ST 04, 11, 15, 22, 23. On the other hand, Age is positively correlated with ST 08 and 19.

Further results related to UX Statements will be examined following a topic-based analysis in Section 5.

4.3. UX DEFINITIONS

4.3.1. CHOICE OF A UX DEFINITION

The authors of the initial study had selected five UX definitions to be presented to the participants (Table 1), different in terms of perspectives on UX and highlighting distinct aspects of UX (see Law et al., 2009 for more details).

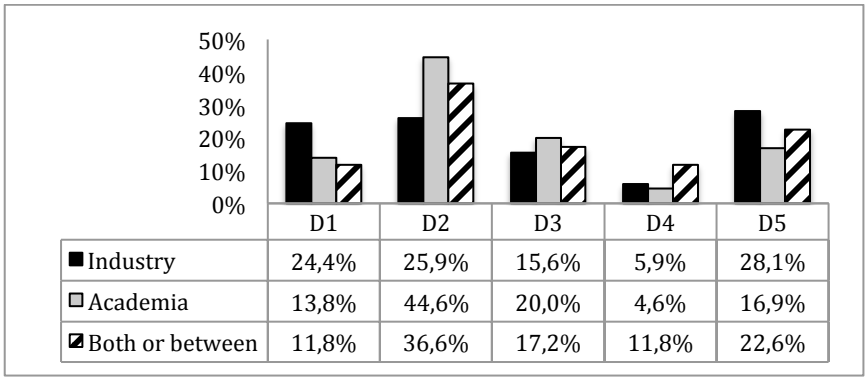
In the present study, 428 respondents indicated which of the five definitions they preferred (Table 5). Definition 2 (D2), focused on what shapes UX, was selected in about a third of cases (31.1%), while D5 and D1 still collected a lot of votes, with respectively 25.2% and 20.1% of the opinions. The less preferred definition is undoubtedly D4 with only 7% of the votes. This latter definition puts the focus on value as an interaction outcome.

Table 5. Distributions of the preferred definitions in our study

	D1	**D2**	**D3**	**D4**	**D5**
Total	86	133	71	30	108
% out of 428	20.1%	31.1%	16.6%	7%	25.2%

Among our respondents, the choice of a definition was correlated to socio-demographic factors. Thus, there are differences in the choice of a UX definition depending on the domain (Fig. 3) ($\chi^2 = 21.67$,$p = .006$), explaining 5% of the total variance in the data ($Phi^2 = 0.05$). Respondents from industry ($n = 270$) mostly choose D1 and D5, while their academic

counterparts (n = 65) mostly supported D2 and D3. Respondents qualifying themselves as working for both or between industry and academia tend to favor D2 and D4.



	D1	D2	D3	D4	D5
Industry	24,4%	25,9%	15,6%	5,9%	28,1%
Academia	13,8%	44,6%	20,0%	4,6%	16,9%
Both or between	11,8%	36,6%	17,2%	11,8%	22,6%

Fig. 3. Choice of a UX definition according to the domain.

Those results seem in line with the origins of the UX definitions presented to the participants. D2 and D3 are academic definitions (Desmet and Hekkert, 2007 and Hassenzahl and Tractinsky, 2006), whereas D1 and D5 originate in Industry.

Differences of opinion also exist depending on the role (χ^2 = 30.27, p = .017), explaining 7% of the variance (Phi^2 = 0.07). Results from Chi-square tests confirm that researchers and students tend to choose D2 and D3. Practitioners preferred D1, whereas consultants preferred D4 and D5. Finally, managers preferred D1 and D5 (Table 6).

Table 6. Choice of a UX definition according to the Role. Numbers represent frequencies.

	D1	D2	D3	D4	D5	Total
Researcher	12	27	22	4	12	77
Consultant	16	30	12	11	32	101
Manager	15	10	5	3	14	47
Practitioner	38	47	27	8	39	159
Student	5	19	5	4	11	44
Total	86	133	71	30	108	428

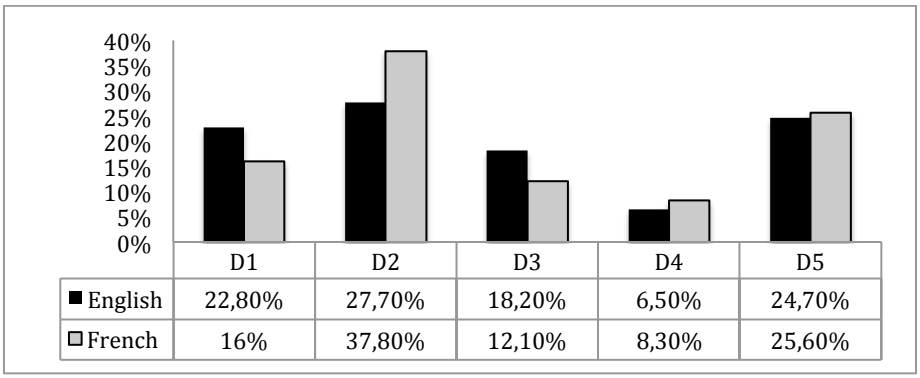
Similarly, educational background also explains 7% of the variance in the choice of a UX definition (χ^2 = 26.32, p = .05, Phi^2 = 0.07) (Table 7).

Table 7. Choice of a UX definition according to the educational background. Numbers represent frequencies.

	D1	D2	D3	D4	D5	Total
Design/Arts	19	18	11	8	20	76
Business/Marketing	13	6	6	0	8	33
Psychology/Soc.	15	41	14	7	27	104
Technology/SW	16	30	23	5	24	98
HCI	11	24	6	4	18	63
Total	74	119	60	24	97	374

Significant differences in the choice of a UX definition seem also to depend on the language in which the survey was answered (χ^2 = 8.14, p = .43), which explains 1.9% of the total variance (Phi^2 = 0.01). English-Speaking participants preferred D1 and D3, whereas French-Speaking participants

tend to choose D2, D4 or D5 (Fig. 4). No statistics were calculated regarding respondents who answered the survey in German, as their sub-sample (n = 9) was too small to allow for statistical comparison.



	D1	D2	D3	D4	D5
English	22,80%	27,70%	18,20%	6,50%	24,70%
French	16%	37,80%	12,10%	8,30%	25,60%

Figure 4. Choice of a UX definition according to the language in which the survey was answered

Gender, interest in UX or the fact that UX is central to the activity of the respondents do not appear significant in the choice of a UX definition, as opposed to age, $F(4, 422) = 7.38$, $p < .001$, $\eta^2 = .065$, and years of experience in the domain of UX, $F(4, 405) = 3.87$, $p = .004$, $\eta^2 = .037$. Respondents preferring D2 are on average younger (M = 32.7, SD = 8.1) than respondents preferring D4 (M = 39.2, SD = 12.5) or D1 (M = 38.4, SD = 9.65). Similarly, respondents preferring D2 have on average less experience in the domain of UX (M = 5.32, SD = 5.70) than respondents preferring D4 (M = 8.80, SD = 7.35) or D1 (M = 8.65, SD = 7.63).

The qualitative analysis of answers to the open-ended questions, "*What do you think of this definition?*" specifically shed light on the elements an acceptable definition of UX should include and how it should be formalized. A synthesis of most frequent comments, as expressed by the participants, has been made and is presented in the following.

4.3.2. WHAT SHOULD A GOOD DEFINITION OF UX HIGHLIGHT?

4.3.2.1. FOCUS ON THE USER

According to the respondents, a good UX definition should definitely be focused on the user. D4 particularly drew criticism for not mentioning the user: "*not a very good definition; it lacks a clear focus on the user.*" Conversely, respondents liked D3 for being "*much more user-centric, which is the way to go when talking about UX*". User's goals and user's affective and emotional states are highlighted as crucial aspects to be included in a UX definition.

Conversely, many respondents stated that a UX definition should not relate to companies or marketing and therefore should not use the word 'customer' to designate the user ("*UX is not necessarily about marketing a product. It's much larger.*"). D1 drew criticism for being too focused on business while participants also disliked the use of the word 'product' in

D3. Similarly, the notion of 'value' used in D4 was more associated with Service Design than with UX.

4.3.2.2. INTERPLAY OF FACTORS

Respondents mostly see UX as a multidimensional concept. Thus, UX professionals are looking for a definition able to encompass several aspects of the interaction, accounting for the complexity of UX. Respondents especially liked the interplay of factors involved in D2, while they found D3 "*way too focused on emotions*". System-related aspects, contextual aspects, social-aspects as well as temporality were often cited as important to mention in any good UX definition.

4.3.2.3. COMPONENTS AND RESULTS OF UX

Participants further commented that a good UX definition should both mention the components of UX and the results (outcomes) of an experience. D2 was typically understood as a description of the main elements composing UX, while D3 was seen as a "*post-UX definition, only focusing on the results*". Some participants suggested combining D2 and D3 in order to have a broader and more comprehensive definition of UX.

Regarding the outcomes of UX, respondents often deplore that the five definitions included in the survey tend to describe only what a positive UX would be. Participants underlined the fact that UX might be positive or negative, and that the latter should not be avoided.

4.3.3. HOW SHOULD A GOOD DEFINITION OF UX BE FORMULATED?

4.3.3.1. DEFINITION'S LENGTH

UX professionals agree on the fact that a good UX definition should be short enough to be easily understandable and memorisable, while at the same time being detailed enough to encompass every important UX-related aspect. D1 for example was often described as "*too long*" or "*way too wordy*" while D4 was described as "*too short*", "*only encompassing the bare minimum*" or even "*too brief to be comprehensive*". Similarly, D2 was criticized for entailing too many brackets, making it hard to read and understand. One of the participants even stated, "*if you have that many "(e.g.,)" in a single sentence, your sentence is probably not clear enough.*"

4.3.3.2. DEFINITION'S WORDING

Many participants would enjoy having a user-friendly UX definition. The words used within the definition should be accurate and clear. A UX definition should not entail terms or concepts that are vague and make things even less understandable. As an illustration, the notions of 'value' or 'supporting cast' evoked in D4 are criticized for being "*vague*" or "*obscure*". Defining a complex concept by using even more complex concepts definitely leads to an unusable definition. Similarly, respondents deplore the fact that D5 defines UX by the too generic wording 'quality of experience', which is actually a statement of the obvious.

4.3.3.3. DEFINITION'S SCOPE

Respondents underlined that the scope of a UX definition should neither be too restrictive (because UX is commonly defined as holistic) nor should it be too large in order to be distinct from other related concepts (such as usability for example). D4 is for example described as "*far too narrow*", "*too simplistic*" and "*incomplete*", therefore "*leaving out some important points*". On the contrary, D2 drew criticism for lacking precision. While a majority of respondents agree that D2 mentions the three main elements of an interaction (i.e., user, system, and context), many respondents pinpointed the fact that D2 could as well be a definition of usability or user-centred design in general. Regarding D3, respondents are unhappy with the fact that "*it reduces UX only to the affective part of the interaction*".

Moreover, respondents feel also concerned by the universality of a UX definition, which should not be focused on a specific type of product or service. D2 was appreciated because it "*works for many different applications*" and "*feels more universal*". Similarly, D5 was described as "*applicable across a broad range of experiences*". D3 was disliked for only using the word 'product' and omitting the 'service' aspect of UX.

4.3.3.4. DEFINITION'S AIM

Many respondents feel the need for a definition that would allow translating the concept into practice. D2 is thus described as "*too academic*" and "*too far away to support a business case or development of a product/service*". It is seen as "*descriptive, but providing little direction as to what exactly should be done"*. Similarly, D3 was described as "*too conceptual and intellectual*". Several respondents believed D3 to describe "*an experience, but it does not indicate how the designer fits into the experience. It is not designed, just experienced*". Participants also often mentioned the issue of measurability of UX as being a main concern for UX professionals.
Finally, regarding the question why we should consider having pragmatic definitions, one of the respondents states, "*a good definition – if we have to have one at all, needs to be easy to 'sell' to the community that will use it!*".

5. TOPIC-BASED ANALYSIS

5.1. IS UX A NEW APPROACH? (ST 21)

Conversely to Hassenzahl's (2008) famous assumption "UX is not just old wine in new bottles", our respondents do not, on average, consider UX as a new approach. Their rating of the statement "UX is not new, it is already covered by existing engineering approaches" is globally neutral (M = 3.14, SD = 1.11). Country of residence is one of the only background-related variables a one-way between subjects ANOVA shows a slight significant difference on that statement, $F(542, 6) = 3.80$, $p < .001$, $\eta^2 = .04$. Continents that less agree with the statement are Europe (M = 3.03, SD = 1.12) and North America (M = 3.12, SD = 1.08). They are opposed to Oceania (M = 3.61, SD = 1.09), Africa (M = 3.67, SD = 0.58), Asia (M = 3.71,SD = 1.08), South America (M = 3.72, SD = 0.83) and Middle East (M = 4, SD = 0.82). The occidental origin of the UX concept and its progressive geographical spread may explain those differences.

Regarding possible links with other statements, results show several positive (however quite low) correlations with statements focused on related fields or concepts, as ST04 "UX is best viewed in terms of marketing" ($r = .14$, $p < .001$), ST12 "Usability is a necessary precondition for good UX" ($r = .11$, $p = .013$) or ST20 "UX is equal to emotional attachment" ($r = .24$, $p < .001$). This is not surprising considering the fact that, if one thinks UX is not a new approach, then one might more likely think of it as related to other concepts.

5.2. WHAT ARE THE LINKS BETWEEN UX AND PREVIOUS/RELATED FIELDS? (ST 04, 12, 18, 20, 21)

According to our respondents, UX is undoubtedly rooted in User-Centered Design and Usability. The first assumption (i.e., UX is rooted in UCD) is in the top three statements regarding the level of agreement: "Designing (for) UX must be grounded in UCD" ($M = 4.29$, $SD = 0.83$). Moreover, usability is seen as "a necessary precondition for good UX" by 81.4% of participants ("agree" or "strongly agree") ($M = 4.15$, $SD = 0.99$). Definition D2, which is the most agreed upon definition, also emphasizes this link with usability and UCD by stating that UX would be "*the consequence of a user's internal state, the characteristics of the designed system and the context within which the interaction occurs*". The three classical pillars of usability and UCD (user, system and context of use) are thus used here, highlighting a close link between usability and UX.

On the one hand, considering UX as rooted in UCD seems compliant with the academic literature. As mentioned in the first section of this paper, UCD and UX Design are not mutually exclusive and share both the primary concern to incorporate the user's perspective into the development process and several methods to achieve this goal (Maguire, 2001). On the other hand, considering usability as a precondition for good UX would basically mean that a positive experience might not occur unless the system is easy to use. It is noteworthy that studies have shown that perceived usability might be influenced by more subjective factors, as aesthetics (Tractinsky, Katz, & Ikar, 2000) or hedonic qualities of the system (Thüring & Mahlke, 2007). Conversely to what respondents stated, usability would therefore not be a precondition for good UX (at least in the conditions explored by those cited studies) and UX-related factors like aesthetics or hedonic qualities might even significantly influence perceived usability.

5.3. IS THERE A NEED FOR A STANDARDIZED DEFINITION OF UX? (ST 07)

The starting point of this study was the apparent lack of consensus on what UX is. Considering the diversity of UX definitions, from both business and academic worlds, it is interesting to wonder whether a single and standardized definition of UX should really be a reason for concern.
It seems that the need for a standardized definition of UX is felt differently amongst different cultures and levels of expertise. Thus, if Statement 07 collects a rather neutral agreement score ($M = 3.71$, $SD = 1.7$), it is interesting to notice that French-speaking participants ($M = 3.95$) feel a higher need for a standardized definition compared to their English-

speaking counterparts (M = 3.54), $t(545) = -4.51$, $p < .01$, $\eta^2 = .036$. The delayed emergence of the UX concept in French-speaking countries might explain a difference in the appropriation of UX, and therefore, a higher need for a definition that could act as a guidance.

Similarly, the most experienced practitioners feel less need for a standardized definition of UX (ST07; $r = -.16$, $p < .01$). It seems that with progressing experience and increasing integration of UX into business processes, experts have developed their own understanding of this concept and do not need a shared view on this topic any more. As outlined in the Manifesto for UX (2007), a standardized UX definition would in fact mostly contribute to communicate on UX, to teach this concept and to progress in this field of research. An agenda for UX research and practice has been published in 2010 (Law & Van Schaik, 2010) and focuses primarily on the need to model UX. It seems however that this concern is more relevant to academia as practitioners are more likely to think that UX should be approached in a qualitative way (M = 3.70 vs. M = 3.51 for researchers), $t(439) = 1.90$, $p < .05$, $\eta^2 = .005$.

5.4. HOW TO APPROACH UX: QUANTITATIVELY OR QUALITATIVELY? (ST 16, 22)

On the question whether UX should be approached quantitatively or qualitatively, no clear answer emerged out of the survey. Respondents agree on both the statements that "UX must be approached qualitatively" (ST22; M = 3.66; SD = 0.99) and that "UX can be quantified and thus compared across similar (or competitive) artifacts" (ST16; M = 3.62, SD = 0.96). As mentioned above, Industry (M = 2.96) less agrees on the uniqueness of lived experienced than Academia (M = 3.21), $t(443) = -.78$, $p < .05$, $\eta^2 = .006$. This means that respondents in Industry tend to believe more in the comparability of peoples' experiences than respondents in Academia. Surprisingly, it is rather in Industry that one considers that UX should be addressed in a qualitative manner (ST22; M = 3.70 vs. M = 3.51 in Academia), $t(439) = 1.90$, $p < .05$, $\eta^2 = .005$. As underlined by Law et al. (2014), UX practitioners might lack training to master quantitative measurement methods. Moreover, we could also assume that they might favor qualitative approaches because of end-user recruitment issues or the common use of "guerilla" usability methods, which involve similar techniques to traditional usability methods with however a more quick-and dirty, informal approach.

Following these results, it is worth comparing these self-reported views on the qualitative versus quantitative aspects with other research on actual UX practice. Interesting research papers give us some clues to understand the reality of UX research and practice. In their review on empirical research on UX, Bargas-Avila and Hornbæk (2011) showed that among 66 empirical studies, half were qualitative, while researchers used quantitative methods in 33% of the cases and combined approaches in 17% of the cases. By comparing their results to Barkhuus and Rode's review (2007), they also show an apparent shift in Academia from quantitative methods to qualitative methods. More recently, Law et al. (2014) conducted both interviews and a survey to collect attitudes toward UX measurement. The

outcomes of their study allowed them to build a synthesis of arguments for and against UX measurement. Here again, no clear consensus emerged and respondents mentioned distinct pros and cons for both qualitative and quantitative UX assessment methods. Some participants described quantitative measures as useful to convince decision makers to modify a problematic design, whereas others highlighted the easiness to derive alternative design ideas from qualitative UX feedbacks.

5.5. IS UX INDIVIDUAL OR SOCIAL? (ST 09, 19)

On average, respondents agree on the fact that "only an individual person can have an experience" (ST19; $M = 3.14$, $SD = 1.23$). However, professionals' opinions are more balanced regarding the assumption that "people will never have comparable UX" (ST09; $M = 3.02$, $SD = 1.16$). The frequency distributions of answers among the five agreement levels on these two statements are almost bimodal and symmetrical, which means that participants have an opinion about this question (neutral answers are not frequent), but there is no consensus. For example, 43.1% disagree or strongly disagree with ST09, while 37% agree or strongly agree with the same statement (20% of the respondents being neutral).

In the literature, UX might be sometimes studied as a social experience, when for example a group of people is experiencing together. By introducing the notion of "co-experience", Battarbee (2003) and Battarbee and Koskinen (2005) underlined how the meanings of individual experiences emerge and change, as they become part of social interaction. However, we believe this is not contradicting the statement that "only an individual person can have an experience" as it is the meaning attached to the experience, and not the fact of having lived the experience, that changes in a social context. The presence of people sharing the experience at the same time is therefore one of the contextual factors shaping UX.

5.6. WHAT SHAPES UX?

5.6.1. DOES UX DEPEND ON USER-RELATED FACTORS? (ST 03, 14, 15, 11)

Several statements included in the survey aimed at collecting UX professionals' opinions on which factors they believe are impacting UX. Among the suggested factors, those related to the users seem to reach consensus. Thus, ST03 is the statement showing the highest level of agreement (ST03; $M = 4.54$, $SD = 0.63$): 94.7% of respondents either agree or strongly agree on the fact that "Fleeting and more stable aspects of a person's internal state affect a person's experience of something". Similarly, we observe a shared agreement on ST14 ($M = 4.12$, $SD = 0.76$), ST15 ($M = 3.89$, $SD = 0.89$) and ST11 ($M = 3.56$, $SD = 1.12$).

In the original paper, Law et al. (2009) explained this tendency by referring to Preece, Rogers, and Sharp (2002) who stated that "user experience goals differ from the more objective usability goals in that they are concerned with how users experience an interactive product from their perspective rather than assessing how useful or productive a system is from its own

perspective." (Preece et al., 2002, p. 19). This might therefore explain why our respondents believe that "We cannot design UX, but we can design for UX" (ST13; *M* = 3.96, *SD* = 1.08). UX professionals are actually designing systems or products meant to trigger a specific type of experience but they are not designing the experience in itself.

5.6.2. DOES UX DEPEND ON CONTEXTUAL FACTORS? (ST 05)

In addition to user-related factors, UX experts also agree that contextual factors strongly contribute to shape UX. The second most agreed upon statement "UX occurs in, and is dependent on the context in which the artifact is experienced" exemplifies this point of view (ST05; *M* = 4.34, *SD* = 0.91). These outcomes, along with the choice of D2 as the preferred UX definition, seem to confirm that the classical usability triad user-system-context (ISO 9241–210, 2010) still remains relevant in the context of UX.

5.6.3. ARE THERE TEMPORAL DYNAMICS OF UX? (ST 08, 17, 23, 01, 02, 10)

As stated in the User Experience White Paper (Roto et al., 2011), there seem to be several time spans of user experience, depending on the moment of usage. The respondents globally agree on the fact that "UX is highly dynamic and changes constantly when interacting with a product" (*M* = 3.93, *SD* = 1.01). The temporal dynamics of UX therefore seem to reach consensus amongst the respondents (Fig. 5).

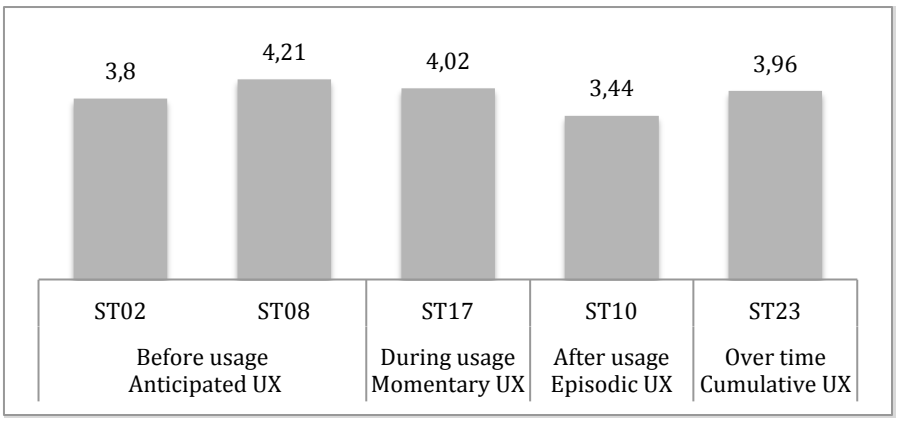


Fig. 5. Time spans of user experience: mean agreement of time-related statements

A majority of them acknowledge the existence of an anticipated UX where "imagined use of a product can result in real experiences" (ST02; *M* = 3.80, *SD* = 1.06) and "prior exposure to an artifact shapes subsequent UX" (ST08; *M* = 4.21, *SD* = 0.78). Momentary UX (the one that is experienced during usage) seems to be favored against Episodic UX (evaluated after usage). Respondents therefore agree on the fact that UX should be assessed "while interacting with an artifact" (ST17; *M* = 4.02, *SD* = 0.86) more than "after interacting with an artifact" (ST10; *M* = 3.44, *SD* = 1.19). Finally, the notion of Cumulative UX

computed by the recollection of multiple periods of use also succeeds to reach a consensus among the respondents. Hence, they are likely to agree on the fact that "UX can change even after a person has stopped interacting with the artifact" (ST23; $M = 3.96$, $SD = 0.87$).

6. WHAT ARE THE TRENDS AND CHANGES OF THE UX CONCEPT? COMPARISON BETWEEN ORIGINAL AND REPLICATED SURVEY

6.1. SAMPLING COMPARISON

As sampling differences might have an impact on the comparability of the results between this study and the original survey broadcast in 2008, it is important to study the structure of both samples.

First of all, the sample size in the current study is almost three times bigger than the one used in 2008. At that time, 275 answers from 25 countries had been collected whereas we collected 758 answers from 35 countries. Vectors of dissemination for both surveys were quite similar (mailing list, conference attendees, professional networks).

In 2008, the geographical distribution was mainly focused on Finland (17.4%), USA (15.6%), UK (13%) and the Netherlands (11.6%). In our study, the main nationalities represented are France (33%), USA (20.8%) and UK (7%). The Netherlands only represents 2% of our respondents. It is important to note that the better coverage of French-speaking respondents was among the main goals of this replication study.

Table 8 presents a comparison between the distributions of respondents according to other background variables. The first main difference between the two samples regards the domain and role. Thus, while Industry represents more than half of the sample in both cases, Academia was much more represented in 2008 (26.2%) than in 2012 (13.2%). Consequently, this is also reflected by the percentage of researchers, which was higher in 2008 (37.8%) than in 2012.

Regarding the respondents' educational background, we can observe two main differences. There are more respondents trained in the field of Arts/Design in 2012 (18.8%) than in 2008 (12.2%) and fewer respondents educated in the field of HCI in 2012 (15.3%) than in 2008 (24.8%).

In both surveys, more than half of the respondents declared their interest in understanding the nature of UX to design better products. This therefore remains the main concern of UX professionals. In the initial survey, a bigger ratio of respondents were interested in understanding the nature of UX per se than in the present study (18.5% vs. 13% respectively). As this concern is typically related to UX research, it might be explained by the higher ratio of respondents from Academia in the initial study. Finally, the willingness to make people happier through the understanding of UX has increased from 14% of the respondents in 2008 to 21.3% of the respondents in 2012.

Table 8. General profile of the respondents: comparison between the initial and the present surveys

Variable	Initial study	Current study
Domain		
Industry	51.1%	66.4%
Academia	26.2%	13.2%
Both or between	22.6%	20.4%
Role		
Researcher	37.8%	16.9%
Consultant/Manager	26.1%	36.9%
Practitioner	19.4%	36.7%
Student / Other	16.7%	9.6%
Educational background		
Arts, Design	12.2%	18.8%
Psychology/Social Sciences	22%	21.6%
Technology / Software	18%	19.7%
Human-Computer Interaction	24.8%	15.3%
Other	23%	24.8%
Interest in understanding UX*:*		
Per se	18.5%	13%
To design better products	55.4%	51.9%
To make people happier	14%	21.3%
To better sell products/Other	12.2%	13.8%
How central is UX to your work?		
Very central	56.8%	58.2%
Central	36%	25.7%
Less central/Not central	7.2%	16.2%

Regarding how central UX is to the respondents' work, we observe in both studies a very high ratio of respondents declaring that UX is very central (56.8% and 58.2%). However, we can notice that a considerable percentage of respondents in 2012 declare UX to be less central or not central at all to their work (16.2%). This ratio was much lower in 2008 (7.2%). We suggest two main explanations: first, the dissemination of the survey in 2012 might have reached a broader population than the initial survey, thus explaining the presence of respondents for whom UX is less central to their work. Another sound explanation might also be that UX has become more popular so that even people who will not define themselves as UX professionals have become familiar with this concept and have been willing to provide their point of view on this topic.

To summarize, in comparison to the sample used in 2008, our sample is bigger in size and involves more French respondents. The proportion of practitioners (vs. academics) is higher than in the initial study.

6.2. COMPARISON REGARDING UX STATEMENTS

The comparison between the rankings of the UX statements sorted by mean agreement shows a very similar ranking pattern between the present study and the initial one. Statements ST03, ST05, ST18 and ST08 rank highest on mean agreement, whereas ST04 and ST20 ranked lowest on mean agreement in both surveys.

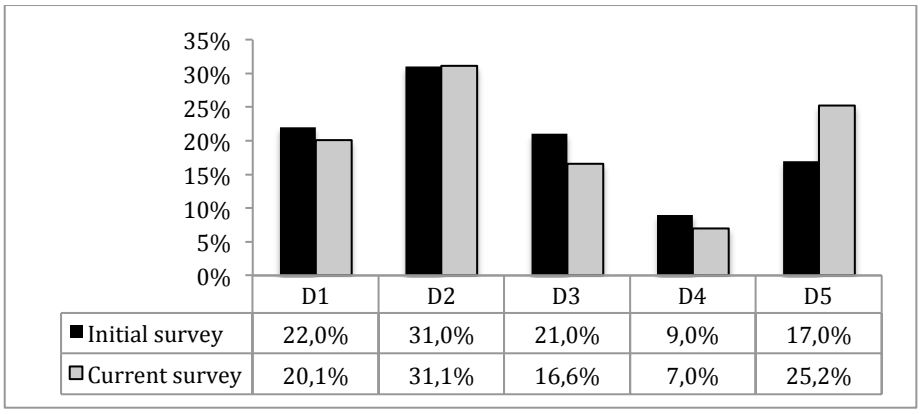
Despite the overall closeness of statements' mean scores, we can still observe some changes regarding the ranking of some statements. First of all, the need for a standardized definition of UX (ST07) seems more

prominent in the present study (ranked 13 out of 23; $M = 3.71$, $SD = 1.07$) than in the original one (ranked 17 out of 23; $M = 3.49$, $SD = 0.7$). This might be explained by a difference in sampling (see previous section for sampling comparison) or by the fact that some professionals feel that they need such a definition to better communicate on UX. Similarly, the respondents in the present study agree more on the assumption that "usability is a necessary precondition for good UX" (ST12; ranked 5 out of 23 in 2012 vs. 11 out of 23 in 2008). This observation is quite unexpected considering the fact that UX research (Thüring & Mahlke, 2007; Tractinsky et al., 2000) has shown that the links between usability and UX seem not to comply with this assumption. We could have expected a better dissemination of UX research into Industry over time. It seems however this goal remains difficult to achieve.

On the other hand, we observe a decrease in ST11 "UX is based on how a person perceives the characteristics of an artefact, but not on the characteristics per se". This statement was the 6th most agreed upon statement in 2008 and is only ranking 13th in 2012. We could assume that the concept of UX was first mostly associated to user-related aspects (e.g., values, affects, emotions). The evolution of UX design methods and the better understanding of UX in general might lead the professionals to think that "the characteristics of the system per se" are able to support UX in a consistent way.

6.3. COMPARISON REGARDING UX DEFINITIONS

The comparison between the initial and the current study regarding the choice of a preferred UX definition is presented in Fig. 6.



	D1	D2	D3	D4	D5
Initial survey	22,0%	31,0%	21,0%	9,0%	17,0%
Current survey	20,1%	31,1%	16,6%	7,0%	25,2%

Fig. 6. Evolutions in the choice of a preferred UX definition

First, it is worth noting that D2 remains the first chosen definition with approximately a third of overall participants' votes. The main evolution regarding UX definitions regards D5, with a percentage of votes increasing from 17% in 2008 to 25.2% in 2012 (+8.2%). This increase in D5 comes mostly at the expense of D3, which drops from 21% in 2008 to 16.6% in 2012 (−4.4%). Finally, D4 and D1 also collected less votes in 2012 than in 2008 with a decrease of approximately 2% for each.

We suggest two hypotheses to explain this evolution. First, our sampling has a higher ratio of practitioners from Industry and we saw in the previous section that they tend to favor D5. Second, the evolution might also be explained by the fact that UX has become more popular and is therefore applied in a wide variety of contexts. D5 is an easy-to-understand definition that confers an extremely broad scope to UX and might therefore unite numerous respondents.

Now that we have presented and discussed all results, it is worthwhile reflecting on the challenges, successes and limitations of this replication study.

7. CHALLENGES AND LIMITATIONS OF THE REPLICATION

7.1. VOLATILITY OF CONCEPTS IN THE FIELD OF HCI

Repeating a conceptual survey presents inherent challenges due to the relative volatility of some concepts and notions developed in HCI, but also due to the volatility of both humans and technology. Driven toward novelty and innovation, some terms used in this research field tend to emerge as popular trends and fade away quickly without having been really analyzed through the lens of empirical research. Some authors in HCI suspect that it could have been the case for UX, which is often used as an umbrella term to designate a wide range of fuzzy and dynamic concepts such as affects, hedonism or aesthetics (Roto et al., 2011). Moreover, after 4 years of intensive use by both practitioners and researchers, it was challenging to repeat a survey aimed at the basics and the definition of UX. We encountered for example the case of a group leader on LinkedIn who refused to broadcast the study claiming that it was now useless because every good practitioner knows what UX is, even though he was unable to provide an accurate definition of UX, himself. Fortunately, beyond this single case, the replicated survey has been received warmly by the community, which demonstrates the need and openness to reflect and examine the concept of UX once again, in a new temporal context. Understanding and validating previous findings seemed nevertheless highly valuable and our approach succeeded in analyzing the maturational process of the UX concept.

7.2. LANGUAGE AND TRANSLATION OF MATERIAL

When working in a non-English speaking country, replication (or even partial use of existing tools only) generally involves the translation of those tools into the native language of the users composing the target population and sample. The administration of a questionnaire in the native language of respondents allows to provide them with a better understanding of the items and to decrease the rate of people being excluded or who abort due to language difficulties. However, translating a survey may become very complex when dealing with conceptual topics (as it is the case here), which already involve several ambiguous items (whether intended or not by their authors) in their original version. The present study was translated into

German and French. Even if a back translation process has been used to verify the reliability of the translation, it is not yet sure whether concepts were understood precisely the same way across different languages (and maybe even across different respondents for the same language). To overcome this difficulty when computing the data, we also compared the level of non- understandability of the items (respondents had the option to check "I don't understand"). Being almost similar for each language and similar to the level found in the original study, the translation was considered fairly reliable.

7.3. COMPARABILITY OF THE RESULTS: SAMPLING AND ADVERTISEMENT OF THE SURVEY

Replicating a research work dealing with the definition of a concept implies reaching a comparable sample both in terms of sample size and demographics. This however, is a difficult task to achieve, given this exploratory study does not involve a random and representative sample. As the whole population of practitioners working in a field related to UX is not clearly defined, it was decided to simply broadcast the survey on the web. We were aware that several biases might have impacted previous results (and may also impact ours), especially the fact that only self- motivated and careful respondents would answer the questionnaire. Moreover, it was impossible to accurately know neither the number of people touched by the advertisement of our survey (probably several thousands), nor the coverage of the target population. However, every research design choice has strengths and weaknesses. The diffusion method chosen for the original study has clear advantages in terms of reaching a wide audience, which fulfilled the primary exploratory goal of the study and provided us with information on what kind of practitioners declare working directly or indirectly on topics related to UX. We succeeded in reaching an international sample larger than the original one ($N = 758$ in 2012 vs. $N = 275$ in 2008) but still almost equivalent in characteristics. The larger sample size had two main advantages: first it allowed detecting more subtle differences in the understanding and perceptions of the notion of UX according to background variables; second it allowed detecting societal evolution related to the field of HCI (e.g., an increase in the number of UX practitioners from Asia, Middle-East or Africa).

7.4. LIMITATIONS HIGHLIGHTED IN THE ORIGINAL SURVEY DESIGN

Replicating research implies repeating a study exactly the way it has been conducted the first time. It is however close to impossible to design studies without any limitation and thus most studies present some limitations, highlighted by the authors or not, that need to be copied for the sake of replication. We emphasize this aspect for it is a peculiar activity in research to design a study with full awareness of limitations one could address if the study was not a replication study. This indeed seems counter-intuitive to the constant strive for progress, but ultimately, due to the need for solid replication in science, leads to more solid progress. In the case of the UX Survey, we noticed some possibilities for improvement regarding the survey design (e.g., reduction of the number of items, rephrasing of ambiguous UX

statements, rotation/counterbalancing of items or the modification of open-ended questions). These improvements could have been done quite easily with a new pre-testing phase involving a limited set of participants. For the sake of replication however, no major change has been implemented with regard to the original study and the only revision regarded a slight extension of the study (cf. Section 3.2).

8. CONCLUSION

By replicating and extending a previous UX survey, we intended to gain further insight into the maturational process of the UX concept. We also aimed at validating previous findings almost taken for granted by the HCI community (e.g., uniqueness of an experience, influence of the context, or temporal dynamics of UX). Despite some challenges and difficulties to overcome, replication of such a survey appeared valuable and highly interesting for the community.

Regarding the levels of agreement with the 23 UX Statements included in the survey, our results largely confirm previous findings. Respondents agree on the importance of both user-related factors and contextual factors as important variables shaping UX. The temporal dynamic of UX also reached consensus amongst the respondents. It is also worth noting that the assessment of Momentary UX (while interacting with an artifact) was favored against Episodic UX (evaluated after usage). Conversely to what respondents declared in 2008 (Law et al., 2009), our respondents believe that UX is not a new concept and that "it is already covered by existing engineering approaches", undoubtedly rooted in User-Centred Design and usability. On the question whether UX should be approached quantitatively or qualitatively, no clear answer emerged out of the survey. Similarly, respondents do not share a clear view on UX being individual or social. In some cases, background variables seem to impact the evaluation of these UX Statements. Amongst them, domain, role, language and years of experience had the greatest impact while gender differences were rarely significant, which confirms previous work by showing that opinions do not contrast sharply according to gender (Law et al., 2009).

The choice of a UX definition was influenced by several factors, including the language in which the survey was answered, respondents' age, educational background, domain, role and years of experience in the UX domain. It also seems that the need for a standardized UX definition (ST07) is felt differently amongst different cultures and levels of expertise. French-speaking participants feel a higher need for a standardized definition compared to their English-speaking counterparts. Similarly, the most experienced practitioners feel less need for a standardized definition of UX. We can therefore wonder why and whether we need a standardized definition? On the one hand, previous authors have highlighted the need for such a definition in order to help teaching and disseminating the UX concept (Law et al., 2014). On the other hand, practitioners have pinpointed their desire to keep a diversity and freedom in their practices. According to our respondents, a good UX definition should definitely be focused on the user. It should not relate to companies or marketing and therefore should not use the word *customer*. UX practitioners are also looking for a definition

able to encompass several aspects of the interaction, accounting for the complexity and multidimensionality of UX. A good UX Definition should both mention the components of UX and the outcomes of an experience. Participants also underlined the fact that UX might be positive or negative and definitions should therefore not be focused on positive UX only. Regarding the formulation of a good UX definition, it should be short enough to be easily understandable and memorisable while at the same time detailed enough to encompass every UX-related aspect. The words used within the definition should be accurate and clear. The scope should neither be too restrictive nor too large. Finally, respondents feel the need for a definition that would allow translating the UX concept into practice. As one can see, practitioners expressed high expectations regarding a UX definition and one can wonder whether it is possible to specify a unique definition fulfilling all these requirements. It might indeed not be possible to provide a standardized definition for both practitioners and researchers from all fields and providing them with different definitions, able to fulfill their specific needs, might be a promising solution?

Another main difference between Academia and Industry was pointed out by our results and regards the level of familiarity with UX, which appears to be significantly higher in Industry than in Academia. Similarly, UX is considered to be much more central in Industry than in Academia. This could suggest that UX originated as a business concept before finding its way into research and theoretical bodies. While UX practitioners are interested in understanding the nature of UX to better design products, researchers are more interested in UX per se, as an object of study. As mentioned earlier, the choice of a UX definition was also influenced by the domain and role of the respondents.

Finally, the concept of UX definitely opens numerous valuable perspectives for the HCI domain. Nevertheless, many questions also remain unanswered, both at a theoretical or a methodological level. Several initiatives have been launched to structure and support the development of UX at the academic level (Law & Van Schaik, 2010; Law et al., 2007). The results of the present study contribute to this objective. By trying to draw an accurate picture of the current situation of UX and building on that basis, we ultimately aim at better methodologies, frameworks and metrics to design for UX. The better understanding gained of practitioners' perspectives is a necessary step toward continual improvement of UX activities. The numerous differences observed between Industry and Academia clearly indicate that there is however still a gap between both perspectives. A better integration of theories and practice should thus be a primary goal, undoubtedly leading to a win–win situation for both Academia and Industry. On the one hand, UX research should be better taught to students around the world as they represent the next generation of UX practitioners. On the other hand, UX research should thrive on practice to better answer the needs and expectations of the UX industry and also provide practitioners with solid and valid tools needed when assessing or designing for UX.

REFERENCES FOR PAPER A

Alben, L. (1996). Quality of Experience: Defining the Criteria for Effective Interaction Design. *Interactions, 3*, 11.

Barcenilla, J., & Bastien, J.M. (2009). L'acceptabilité des nouvelles technologies : quelles relations avec l'ergonomie, l'utilisabilité et l'expérience utilisateur ? *Le Travail Humain, 72*(4), 311-331.

Bargas-Avila, J.A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. Proceedings of the *ACM Conference on Human Factors in Computing Systems*, 2689-2698. doi:10.1145/1978942.1979336

Barkhuus, L., & Rode, J. (2007). From Mice to Men-24 years of Evaluation in CHI. Proceedings of the *ACM Conference on Human Factors in Computing Systems (CHI 2007)*. doi:10.1145/1240624.2180963

Battarbee, K. (2003). Defining co-experience. *Proceedings of the 2003 international Conference on Designing Pleasurable Products and Interfaces (PPPI'03)* (pp. 109-113), Pittsburgh, PA, USA. New York: ACM. doi:10.1145/782896.782923

Battarbee, K., & Koskinen, I. (2005). Co-experience: user experience as interaction. *CoDesign, 1*(1), 5-18. doi: 10.1080/15710880412331289917

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

Fenton, N. E., & Pfleeger, S. L. (1997). *Software Metrics: A Rigorous & Practical Approach*. Boston: PWS Publishing Company.

Forlizzi. J., & Ford. S. (2000). The building blocks of experience: An early framework for interaction designers. *Proceedings of Designing Interactive Systems (DIS 2000)*, New York, USA. doi:10.1145/347642.347800

Gegner, L., Runonen, M., & Keinonen, T. (2011). Oscillating Between Extremes: A Framework for Mapping Differing Views on User eXperience. *Proceedings of the 2003 international Conference on Designing Pleasurable Products and Interfaces (DPPI'11)*, Milan, Italy. doi:10.1145/2347504.2347566

Hassenzahl, M. (2003). The thing and I: understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A.F. Monk, & P.C. Wright (Eds.), *Funology: From Usability to Enjoyment* (pp. 31-42). Kluwer: Dordrecht.

Hassenzahl, M. (2008) User Experience (UX): Towards an experiential perspective on product quality. *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine (IHM 2008)*, Metz, France. doi:10.1145/1512714.1512717

Hassenzahl, M. (2010). *Experience Design: Technology for All the Right Reasons*. San Rafael, California: Morgan & Claypool.

Hassenzahl, M. (2013). User Experience and Experience Design. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "*The Encyclopedia of Human-Computer Interaction, 2nd Ed.*". Aarhus, Denmark: The Interaction Design Foundation. Available online at https://www.interaction-design.org/encyclopedia/user_experience_and_experience_design.html

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

Hollan, J.D., Hutchins, E.L., & Kirsh, D. (2000). Distributed Cognition: A New Theoretical Foundation for Human-computer Interaction Research. *ACM Transactions on Human-Computer Interaction*, 7, 2, 174-196.

http://www.nngroup.com/about/userexperience.html .The definition of User Experience. (n.d.). Retrieved January 22, 2014, from Nielsen and Norman Group,

http://www.uxnet.org/. User Experience Network. (n.d.). Retrieved January 22, 2014, from UXNet, (last verified 22 January 2014)

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

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

International Organization for Standardization. (1998). ISO 9241-11:1998, *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability*. Geneva, Switzerland: International Organization for Standardization.

Jordan, P. (2002). *Designing pleasurable products: An introduction to the new human factors*. Boca Raton, FL: CRC Press.

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. doi: 10.1016/j.intcom.2010.04.003

Kuutti, K. (1996). Activity Theory as a Potential Framework for Human-computer Interaction Research. In B.A. Nardi (ed.), *Context and Consciousness: Activity Theory and Human-computer Interaction* (pp. 17-44). Cambridge, MA: MIT Press.

Lallemand, C., Koenig, V., & Gronier, G. (2013). Replicating an International Survey on User Experience: Challenges, Successes and

Limitations. Proceedings of the CHI'2013 Workshop on the Replication of HCI Research, CEUR Workshop Proceedings, Vol-976.

Law, E.., van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies*, 72 (6), 526-541. doi: 10.1016/j.ijhcs.2013.09.006i

Law, E., Roto, V., Hassenzahl, M., Vermeeren, A. & Kort, J. (2009). Understanding, scoping and defining UX: a survey approach. Proceedings of the *ACM Conference on Human Factors in Computing Systems (CHI 2009)*, Boston, USA. doi: 10.1145/1518701.1518813

Law, E., & Van Schaik, P. (2010). Modelling User Experience – An agenda for Research and Practice. *Interacting with computers, 22 (5),* 313-323. doi: 10.1016/j.intcom.2010.04.006

Law, E., Vermeeren, A., Hassenzahl, M., & Blythe, M. (2007). Towards a UX Manifesto. COST294-MAUSE affiliated workshop, Lancaster, UK.

Maguire, M. (2001). Methods to support human-centred design. *International Journal of Human-Computer Studies, 55*, 4, 587-634. doi: 10.1006/ijhc.2001.0503

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.

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

Nielsen, J. (1993). *Usability Engineering*. New York: AP Professional.

Norman, D. (2004). *Emotional design: Why we love (or hate) everyday things*. New York: Basic Books.

Norman, D., Miller, J., & Henderson, A. (1995). What You See, Some of What's in the Future, And How We Go About Doing It: HI at Apple Computer. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1995)*, Denver, Colorado, USA.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction*. New York, NY: John Wiley & Sons.

Rogers, Y. (2012). *HCI Theory: Classical, Modern, and Contemporary*. Synthesis Lectures on Human-Centered Informatics, May 2012, Vol. 5, No. 2.

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (2011) User Experience White Paper: Bringing clarity to the concept of user experience. Result from Dagstuhl Seminar on Demarcating User Experience, Finland. Retrieved from *http://www.allaboutux.org/uxwhitepaper*

Shackel, B. (1991). Usability—context, framework, definition, design and evaluation. In B. Shackel, & S. Richardson (Eds.), *Human Factors for Informatics Usability* (pp. 21-38). Cambridge: Cambridge University Press.

Sward D., & MacArthur, G. (2007). Making user experience a business strategy. In E. Law et al. (Eds.), *Proceedings of the Workshop on Towards a UX Manifesto*, Lancaster, UK (pp. 35-40). Retrieved from http://www.cost294.org

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology*, *42*(4), 253–264. doi: 10.1080/00207590701396674

Tokkonen, H., & Saariluoma, P. (2013). How User Experience is Understood. *Proceedings of 2013 Science and Information Conference (SAI 2013)*, 791-795.

Tractinsky, N., Katz, A., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*(2), 127-145. doi: 10.1016/S0953-5438(00)00031X

ENGLISH SUMMARY OF PAPER B

A FRENCH VERSION OF THE ATTRAKDIFF SCALE: TRANSLATION AND VALIDATION STUDY

Carine Lallemand, Vincent Koenig, Guillaume Gronier, & Romain Martin

Introduction. – While user experience (UX) evaluation is a core concern within the field of Human-Computer Interaction (HCI), there is currently no valid self-administered UX evaluation tool in French. The AttrakDiff scale (Hassenzahl, Burmester, & Koller, 2003) is a UX evaluation tool, which relies on a theoretical model distinguishing pragmatic and hedonic qualities of interactive systems.

Objective. – This paper describes the translation and validation of the French version of the AttrakDiff scale in order to ease UX assessment in French-speaking users.

Method. - Following the cross-cultural methodology developed by Vallerand (1989), the questionnaire was translated by trilingual researchers before being back-translated and validated by a panel of experts. A pre-test was conducted on 26 participants. The characteristics of the French version of the AttrakDiff scale were then evaluated through a quantitative online study involving a sample of 381 users.

Results. – The results confirm the expected 3-factors structure and a good internal consistency of each subscale. The links between factors are consistent with Hassenzahl's theoretical model (2003) where pragmatic and hedonic perceived attributes combine to form a judgment of attractiveness.

Conclusion. - The current French version of the AttrakDiff scale is majorly reliable with regards to the initial German version and presents satisfactory levels of validity and reliability. Three problematic items are discussed with regards to the characteristics of the system used for the validation study.

Keywords: Human-Computer Interaction, interactive systems evaluation, user experience, questionnaire, AttrakDiff scale, transcultural validation process

B

CREATION ET VALIDATION D'UNE VERSION FRANCAISE DU QUESTIONNAIRE ATTRAKDIFF POUR L'EVALUATION DE L'EXPERIENCE UTILISATEUR DES SYSTEMES INTERACTIF

Carine Lallemand, Vincent Koenig, Guillaume Gronier, & Romain Martin

Submitted for journal publication

Keywords: Interaction Homme-Machine, évaluation de systèmes interactifs, expérience utilisateur, questionnaire, AttrakDiff, processus de validation transculturelle.

ABSTRACT

Introduction. – Alors que l'évaluation de l'expérience utilisateur (UX) est au cœur des préoccupations dans le domaine des Interactions Homme-Machine, aucun outil d'évaluation auto-administré valide de l'UX n'existe actuellement en langue française. Le questionnaire AttrakDiff 2 (Hassenzahl, Burmester, & Koller, 2003) est un outil d'évaluation de l'UX qui repose sur un modèle théorique distinguant qualités pragmatiques et qualités hédoniques des systèmes interactifs.
Objectif. - Cet article décrit la traduction et la validation de la version française du questionnaire AttrakDiff 2 en vue de son utilisation sur des échantillons de population francophone.
Méthode. - Suivant la méthodologie de validation transculturelle proposée par Vallerand (1989), le questionnaire a été traduit par des chercheurs trilingues puis a fait l'objet d'un processus de traduction renversée et d'une validation par un comité d'experts. Un prétest a été effectué sur 26 participants. Les caractéristiques de la version française de l'AttrakDiff 2 ont ensuite été évaluées par une étude quantitative en ligne sur un échantillon de 381 utilisateurs.
Résultats. - Les résultats des analyses effectuées confirment la structure factorielle attendue en 3 facteurs et une bonne consistance interne des sous-échelles. Les liens entre les facteurs sont consistants avec le modèle théorique d'Hassenzahl (2003) où attributs pragmatiques et hédoniques perçus se combinent pour former un jugement d'attractivité.
Conclusion. - La présente version française de l'AttrakDiff 2 est globalement conforme à la version initiale allemande de l'outil et présente des niveaux de validité et de fidélité satisfaisants

1. INTRODUCTION

L'expérience utilisateur est communément décrite dans la littérature comme la qualité globale de l'interaction entre un utilisateur et un système interactif. Ce concept, envisagé comme une perspective étendue et distincte sur la qualité des produits interactifs (Hassenzahl, 2008), connaît une popularité croissante depuis plus d'une décennie. Dans une société post-matérialiste (Hassenzahl, 2013), où l' « économie de l'expérience » (Pine & Gilmore, 1998) prend une place prédominante, les enjeux liés à la conception et à l'évaluation de l'expérience utilisateur des systèmes interactifs deviennent essentiels.

Le questionnaire AttrakDiff 2 (Hassenzahl, Burmester, & Koller, 2003) est l'un des outils d'évaluation de l'expérience utilisateur les plus utilisés au niveau académique. Basé sur un modèle théorique illustrant comment l'évaluation des qualités hédoniques et pragmatiques d'un système influence la perception de l'attractivité globale d'un système (Hassenzahl, Beu, & Burmester, 2001), il est constitué de 28 paires d'items opposés à évaluer sur une échelle de Likert. Initialement développé en allemand, l'AttrakDiff 2 a ensuite été traduit et utilisé en anglais pour l'évaluation de systèmes interactifs de tous types.
Cependant, plusieurs auteurs rapportent des difficultés d'utilisation du questionnaire sur des populations non germanophones ou non anglophones (Holm & Lehtiniemi, 2011 ; Raita & Oulasvirta, 2011), principalement en raison de l'absence de traductions valides dans la langue natale des participants. Ainsi, Holm et Lehtiniemi (2011) rapportent que seuls 9 des 41 participants finlandais à leur étude ont complété l'AttrakDiff 2 et expliquent cela en partie par le fait qu'ils « *ont trouvé cela fastidieux et peu pratique de traduire les termes utilisés dans l'outil aux participants ne parlant pas couramment anglais* » (p. 331).
En France, la problématique est similaire. Bien que largement utilisé à travers le monde académique pour l'évaluation de tous types de systèmes, il n'existe pas à l'heure actuelle de traduction valide de l'AttrakDiff 2 pour des populations francophones. Or, le concept d'UX s'est considérablement développé aussi bien au niveau académique que professionnel et le besoin d'un outil de mesure francophone valide de la qualité de l'expérience avec un système interactif se fait donc pressant. L'objectif de cette étude est de développer une version française de l'AttrakDiff 2 qui pourra être utilisée pour l'évaluation de l'expérience utilisateur sur des échantillons francophones, des comparaisons interculturelles impliquant des utilisateurs francophones, ou encore en tant que mesure de la validité convergente ou divergente de divers outils qui sont ou seront à l'avenir développés en français dans le domaine de l'expérience utilisateur.

La traduction méthodique d'un instrument de mesure constitue un travail fastidieux pour les chercheurs. Pourtant, cette procédure s'impose comme la seule scientifiquement valide. En effet, les traductions non rigoureuses d'échelles de mesure, courantes dans les pays non anglophones, peuvent mener à des problèmes d'interprétation et de comparaison des résultats (Gudmundsson, 2009). Une traduction menée par un seul chercheur pour une étude spécifique risque ainsi de comporter des biais liés à l'interprétation et à la subjectivité du chercheur (Vallerand, 1989). C'est

pourquoi des principes d'adaptation et de traduction de tests et d'instruments, notamment psychologiques, ont été proposés (Gudmundsson, 2009 ; International Test Commission, 2010; Vallerand, 1989 ; van de Vijver & Hambleton, 1996).

Les instruments de mesure verbaux tels que l'AttrakDiff présentent souvent l'inconvénient d'être dépendants du langage, et parfois également de la culture dans laquelle ils sont administrés (Hartson & Pyla, 2012). Ainsi, le vocabulaire utilisé pour définir les paires de mots dans l'AttrakDiff est difficile à traduire précisément. De plus, une traduction mot à mot s'avère parfois inadéquate pour assurer la validité conceptuelle de l'outil traduit (Hilton & Skrutkowski, 2002). Or, l'intérêt principal d'utiliser un instrument de mesure existant réside dans la validité du construit théorique sous-jacent et la possibilité de comparaisons interculturelles sur les dimensions mesurées. Dans le cas de l'AttrakDiff, l'adoption d'un processus de validation de la traduction est nécessaire pour assurer l'équivalence du questionnaire français avec sa version originale, et ainsi mesurer les mêmes sous-dimensions du concept d'expérience utilisateur.

Cet article détaille la méthodologie de traduction et de validation de la version française du questionnaire AttrakDiff, utilisé pour l'évaluation de l'expérience utilisateur d'un système interactif. Nous évoquerons tout d'abord l'émergence du concept d'expérience utilisateur et les enjeux liés à son évaluation. Puis, nous expliciterons le modèle théorique sur lequel repose l'AttrakDiff (Hassenzahl et al., 2003) ainsi que la méthodologie utilisée pour la création du questionnaire. Nous nous focaliserons enfin sur la traduction en langue française de l'AttrakDiff et les étapes suivies pour assurer et tester sa validité.

1.1. L'EVALUATION DE L'EXPERIENCE UTILISATEUR AU CŒUR DES PREOCCUPATIONS

Le terme « expérience utilisateur » (UX) a été utilisé pour la première fois par Norman dans les années 1990 afin d'étendre le champ trop étroit de l'utilisabilité et couvrir tous les aspects de l'expérience d'une personne avec un système (Norman, Miller, & Henderson, 1995). Parmi les nombreuses définitions proposées (Desmet & Hekkert, 2007 ; Hassenzahl & Tractinsky, 2006 ; Lallemand, Gronier, & Koenig, in press ; Law et al., 2009), chercheurs et praticiens s'accordent à dire que l'UX est le résultat de l'interaction entre trois éléments : l'utilisateur, le système et le contexte (Roto, Law, Vermeeren, & Hoonhout, 2011). Suivant cette vision assez commune dans le domaine des Interactions Homme-Machine (IHM), Hassenzahl et Tractinsky (2006) définissent l'UX comme « *une conséquence de l'état interne d'un utilisateur (prédispositions, attentes, besoins, motivation, humeur, etc), des caractéristiques du système conçu (complexité, but, utilisabilité, fonctionnalité, etc.) et du contexte (ou environnement) dans lequel l'interaction prend place (cadre organisationnel/social, sens de l'activité, volonté d'usage, etc.).* » (p. 95, traduction des auteurs). Partageant de multiples bases communes avec le concept d'utilisabilité (Nielsen, 1993), l'UX vient élargir ce dernier par des aspects émotionnels, subjectifs et temporels à la fois plus nombreux et plus précis. Là où le concept d'utilisabilité se concentrait majoritairement sur une approche objective de l'interaction, l'UX va spécifiquement creuser les

aspects subjectifs caractérisant le vécu d'un être humain au contact avec la technologie, en valorisant notamment l'importance du vécu rapporté.

Les enjeux de l'UX sont nombreux puisque, au-delà des avantages associé à une bonne utilisabilité (Maguire, 2001), créer des produits ayant une bonne expérience utilisateur constitue un facteur concurrentiel déterminant sur les marchés de consommation matures (Pine & Gilmore, 1998). Étape clé du processus de conception itératif (ISO 9241-210, 2010), la question de l'évaluation de l'UX est au cœur des préoccupations des chercheurs et professionnels (Hassenzahl & Tractinsky, 2006). Ainsi, l'opérationnalisation des connaissances théoriques produites est vue comme une étape cruciale vers la maturité du concept, nécessaire à son utilisation optimale pour la conception des systèmes interactifs. De nombreux outils et méthodologies d'évaluation de l'UX ont été proposés. Près d'une centaine d'entre eux ont été recensés et catégorisés (Vermeeren et al., 2010). Parmi les outils auto-administrés, l'AttrakDiff (Hassenzahl, Beu & Burmester, 2001) évalue par un questionnaire les perceptions de l'utilisateur envers le système. L'AttrakDiff présente plusieurs qualités en tant qu'outil de mesure de l'UX (Hartson & Pyla, 2012). D'une part, il est relativement court, facile à administrer, et l'échelle verbale utilisée est simple à comprendre (Hassenzahl et al., 2001). D'autre part il est ancré dans un modèle théorique dont plusieurs études (Hassenzahl et al., 2001; Hassenzahl et al., 2003 ; Hassenzahl, 2004 ; Hassenzahl & Monk, 2010 ; van Schaik & Ling, 2008) ont souligné la cohérence. L'apparente exhaustivité des items qui le composent confère à l'AttrakDiff une validité de contenu satisfaisante pour la mesure de l'UX. Enfin, il est disponible gratuitement en ligne et a été utilisé avec succès dans de nombreuses études.

1.2. ATTRAKDIFF : UNE EVALUATION DES QUALITES HEDONIQUES ET PRAGMATIQUES DES SYSTEMES INTERACTIFS

Le questionnaire AttrakDiff a été développé par Hassenzahl et al. (2003) pour évaluer les qualités pragmatiques, les qualités hédoniques ainsi que l'attractivité des produits interactifs. La création de cet outil par une méthode de génération participative d'items impliquant des experts permet à l'AttrakDiff de couvrir de manière étendue les différents aspects liés à l'interaction d'un utilisateur avec une technologie. Ainsi, les aspects pragmatiques et hédoniques revendiqués par les auteurs englobent aussi les émotions, les aspects esthétiques de l'interaction ainsi que la mesure de l'attractivité globale d'un produit ou d'un système. Quant au contexte d'usage, autre élément clé de la qualité d'une interaction, il est considéré comme incorporé au ressenti de l'utilisateur et se voit donc indirectement représenté.

Rapidement devenu populaire, l'AttrakDiff est utilisé depuis une dizaine d'années dans de nombreuses études scientifiques en tant que mesure de la qualité de l'expérience utilisateur avec un système interactif. Ainsi, il a été utilisé dans presque tous les domaines d'application. On peut notamment citer l'évaluation de jeux vidéo (Christou, 2013 ; Lankes, Bernhaupt, & Tscheligi, 2010), d'applications professionnelles (Grün et al., 2005 ; Schrepp, Held, & Laugwitz, 2006), de systèmes liés à la santé (Klaassen, op

den Akker, Lavrysen, & van Wissen, 2013), de systèmes multimodaux (Wechsung et al., 2009), de mondes virtuels (Holm & Lehtiniemi, 2011) ou encore de plateformes d'e-learning (Eimler et al., 2010). L'Attrakdiff a également été utilisé dans la construction d'autres outils de mesure, notamment en tant que mesure de la validité convergente ou divergente de ces derniers (Harbich & Auer, 2005 – ISONORM 9241/10 ; Laugwitz, Held, & Shrepp, 2008 – The User Experience Questionnaire ; Leuteritz, Widlroither, & Klüh, 2009 – ISOMetrics Questionnaire ; Moshagen & Thielsch, 2010 – VisAWI visual aesthetic scale ; Wechsung & Naumann, 2005). Enfin, le questionnaire AttrakDiff est classiquement inclus dans les études conceptuelles sur l'UX (Hassenzahl, 2004 ; Hassenzahl, 2010 ; Tuch, Trusell, & Hornbæk, 2013 ; van Schaik & Ling, 2008) ou encore pour évaluer comparativement des méthodes de conception (Zwinderman et al., 2013).

Le succès de l'Attrakdiff s'explique probablement par le fait que peu d'outils d'évaluation auto-administrés proposent de mesurer la perception des attributs du produit contribuant à la qualité globale de l'UX. D'une part, de nombreux questionnaires d'utilisabilité existent pour mesurer les aspects pragmatiques de l'expérience, comme par exemple le « System Usability Scale » (SUS ; Brooke, 1996) ou le « Software Usability Measurement Inventory » (SUMI ; Kirakowski & Corbett, 1993). La majorité de ces questionnaires centrés sur le versant pragmatique de l'interaction sont basés sur les modèles de l'utilisabilité tels que décrits dans la norme ISO-9241 (1998). Certains, tels que le « Questionnaire for User Interface Satisfaction » (QUIS ; Chin, Diehl, & Norman, 1988) ou le « Ease of Use Questionnaire » (USE ; Lund, 2001), comprennent également une mesure de la satisfaction liée à l'interaction (qui pourrait donc être assimilée à une mesure d'un aspect hédonique de l'interaction), mais cela constitue une minorité des items. D'autre part, de nombreux outils se centrent uniquement sur des aspects hédoniques de l'interaction, tels que les affects (Russell, Weiss, & Mendelsohn, 1989) ou les émotions (Scherer, 2005). Mais peu d'outils rassemblent à la fois les aspects pragmatiques et hédoniques de l'interaction tout en proposant un modèle théorique sous-jacent comme le fait l'AttrakDiff. En plus de l'AttrakDiff, c'est également le cas du User Experience Questionnaire (UEQ ; Laugwitz et al., 2008), qui comprend 6 sous-échelles : *Attractivité*, *Clarté*, *Efficacité*, *Fiabilité*, *Stimulation* et *Nouveauté*. Le UEQ repose sur le même modèle théorique que l'AttrakDiff et a été créé après ce dernier dans une volonté de neutralité de l'outil par rapport aux différents aspects constitutifs d'une interaction. La principale critique adressée à l'AttrakDiff par les auteurs du UEQ (Laugwitz et al., 2008) est qu'il met plus l'accent sur les aspects hédoniques que sur les aspects pragmatiques de l'interaction. En effet, avec 7 items constituant la sous-échelle des aspects pragmatiques contre 14 items constituant les deux sous-échelles des aspects hédoniques (versant stimulation et versant identification), Laugwitz et al. (2008) reprochent à l'AttrakDiff d'être déséquilibré et de favoriser certains aspects de l'interaction par rapport à d'autres. Selon eux, l'AttrakDiff ne serait pas à même de fournir une évaluation appropriée pour certains types de systèmes, notamment des logiciels professionnels. Cependant, le modèle conceptuel sous-jacent à l'AttrakDiff (Hassenzahl, 2003) permet d'expliquer ce déséquilibre apparent du nombre d'items liés à la dimension pragmatique par rapport au nombre d'items liés à la dimension hédonique. En effet, dans le modèle

d'Hassenzahl, les qualités hédoniques priment sur les qualités pragmatiques, ces dernières n'ayant selon l'auteur de valeur que « parce qu'elles facilitent l'atteinte de « be-goals » [les buts soutenus par les qualités hédoniques] (Hassenzahl, 2008, p. 12).

Au niveau de son format, notons que l'AttrakDiff a également pour avantage de comporter moins d'items que d'autres instruments populaires, tels que le SUMI (Kirakowski & Corbett, 1993) qui comprend 50 affirmations sur l'utilisabilité, ou bien le ISOMetrics Questionnaire qui ne comporte pas moins de 75 items (Leuteritz, Widlroither, & Klüh, 2009). Bien que cette différence de nombre d'items s'explique par des finalités différentes (le SUMI et l'ISOMetrics Questionnaire ont pour but de caractériser avec précision des problèmes d'utilisabilité), la facilité d'utilisation et de passation est tout de même en faveur de l'AttrakDiff.

1.2.1. DÉVELOPPEMENT DE L'ATTRAKDIFF : MODÈLE THÉORIQUE ET PREMIÈRE VERSION DE L'OUTIL

Une première version de l'AttrakDiff (AttrakDiff 1) a été développée en 2003 (Hassenzahl et al., 2003) sur base d'un modèle théorique (Hassenzahl 2002, 2003 ; Hassenzahl, Platz, Burmester, & Lehner, 2000) (Figure 1) distinguant qualité pragmatique, qualité hédonique et attractivité d'un produit.

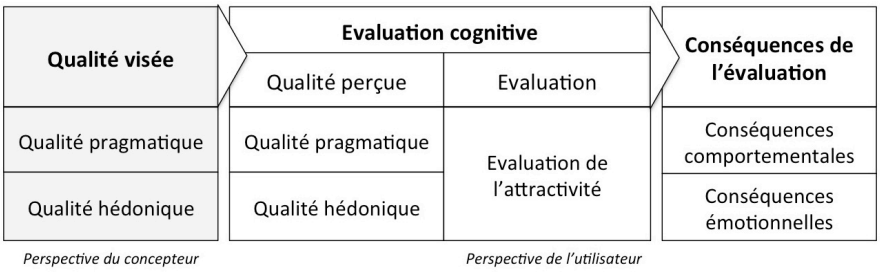


Figure 1. Modèle théorique illustrant comment les qualités pragmatiques et hédoniques influencent la perception subjective de l'attractivité donnant naissance à des comportements et des émotions (adapté de Hassenzahl, 2003)

Selon Hassenzahl et al. (2003), la qualité perçue d'un système dépend de deux attributs principaux : sa qualité pragmatique et sa qualité hédonique. La qualité pragmatique désigne principalement les aspects instrumentaux du système ou produit, c'est-à-dire son utilité et son utilisabilité. Ces qualités pragmatiques vont soutenir la réalisation d'objectifs – tâches (appelés « do-goals »). La clarté du système, sa structure ou sa prévisibilité sont autant d'attributs liés à la qualité pragmatique. La qualité hédonique quant à elle est non instrumentale et se réfère au soi. Elle est donc liée à l'utilisateur, et se base sur un jugement du potentiel du produit à procurer du plaisir et à satisfaire l'épanouissement de besoins humains plus profonds appelés « be-goals ». La capacité du système interactif à stimuler l'utilisateur, à le connecter aux autres, à lui donner un sentiment de contrôle ou encore à lui conférer une certaine popularité sont des attributs liés à la qualité hédonique. Ces deux attributs pragmatique et hédonique se combinent pour

générer une évaluation globale de l'attractivité du produit, qui désigne l'appréciation globale découlant de la qualité perçue. Cette dernière va finalement être à l'origine de conséquences comportementales (ex : accroissement de l'usage) et émotionnelles (ex : joie).

L'AttrakDiff 1 était composé de 23 items sous forme de différentiateurs sémantiques. Le score de l'échelle pour chaque dimension était obtenu à partir des valeurs moyennes recueillies pour chaque item. Les études préliminaires de validation de l'échelle (Hassenzahl, 2003) ont montré l'indépendance entre qualité hédonique perçue et qualité pragmatique perçue d'un système, toutes deux étant à l'origine de l'évaluation de l'attractivité. Bien que prometteur, ce premier questionnaire avait toutefois pour limite de ne pas faire de distinction entre la stimulation et l'identification, deux sous-composantes de la qualité hédonique d'un système interactif. La qualité hédonique-stimulation trouve son origine dans la quête constante des individus de développer leurs connaissances et compétences. Pour soutenir le développement personnel des utilisateurs, les produits et systèmes doivent être stimulants, c'est-à-dire donner aux utilisateurs de nouvelles impressions, opportunités ou idées (Hassenzahl, 2003). La qualité hédonique-identification provient quant à elle du fait que les individus expriment qui ils sont, leur « self », à travers les objets matériels qu'ils possèdent et utilisent. Ainsi, les systèmes interactifs doivent soutenir une fonction sociale d'auto-expression et communiquer une certaine identité de l'utilisateur auprès des autres.

1.2.2. ETUDES DE VALIDATION INITIALES DE L'ATTRAKDIFF

Une deuxième version de l'AttrakDiff (AttrakDiff 2) a été créée pour affiner la représentation conceptuelle et prendre en compte la distinction qualité hédonique-stimulation vs. qualité hédonique-identification. Les deux études de validation subséquentes (Hassenzahl, 2003 ; Hassenzahl, 2004) ont souligné la validité de cette distinction dans l'évaluation de la qualité hédonique d'un produit. Les items de l'AttrakDiff 2 ont été générés lors d'un atelier réunissant six experts ergonomes impliqués dans une séance de brainstorming, d'évaluation et de sélection (Hassenzahl et al., 2003). Sur 133 items générés via une phase initiale de créativité, 50 items ont été retenus, auxquels ont été ajoutés les 7 items de l'AttrakDiff 1 mesurant l'attractivité. Cette version expérimentale du nouveau questionnaire a ensuite été testée à travers une étude pilote sur 22 sujets. Les participants devaient évaluer trois sites web, différents sur le plan de l'interaction et de l'aspect, mais semblables sur celui des fonctionnalités proposées. Les sujets étaient divisés en deux groupes, un premier groupe ayant des tâches précises à accomplir sur chacun des sites, et un autre groupe étant placé en situation d'exploration libre. Une fois la navigation achevée, chaque site était évalué par chaque participant à l'aide du questionnaire. Tous les items (sauf ceux de l'AttrakDiff 1) ont été traités à l'aide d'une Analyse en Composantes Principales (ACP), afin de les regrouper et de supprimer ceux qui ne se différenciaient pas suffisamment (21 items retenus à ce stade). Une seconde ACP a ensuite été menée et trois composantes ont été extraites (sur base du critère de Kaiser >1), expliquant 72% de la variance. La validité de construit a été évaluée à l'aide d'un test d'hypothèses concernant le contenu des différents sites web, et leur évaluation en matière de qualité hédonique

(stimulation uniquement) et pragmatique. Une deuxième étude comportant 33 sujets a été menée, avec pour objet un lecteur MP3 présentant quatre habillages différents. Les participants devaient évaluer le lecteur sur la base de l'image de chaque habillage à l'aide du questionnaire. Les auteurs évoquent une validité de construit satisfaisante (cohérence interne et indépendance entre les dimensions) avec des alphas de Cronbach allant de .73 à .90.

Il est toutefois important de souligner les limites de la validation initiale de l'AttrakDiff, qui a été réalisée sur deux études pilotes impliquant de petits échantillons. Le ratio sujets-variables utilisé dans l'étude de validation initiale de l'échelle allemande est insuffisant (22 participants pour un pool initial de 50 items puis un set d'items de 21 items) et ne permet théoriquement pas de réaliser une analyse factorielle valide. Le ratio sujets-variables minimal recommandé pour la validation d'une échelle par analyse factorielle est de cinq fois le nombre d'items de l'échelle et de 100 participants au minimum (Grimm & Yarnold, 1995), soit un échantillon recommandé pour une validation de l'AttrakDiff d'au minimum 250 participants dans la phase de réduction du pool de 50 items, puis d'au minimum 105 participants pour l'analyse de la structure factorielle des 21 items restants sélectionnés. La solution factorielle initiale telle que décrite par Hassenzahl et al. (2003), bien que conforme au modèle théorique, présente potentiellement des problèmes de stabilité et de fidélité.

1.2.3. VERSION ACTUELLE : ATTRAKDIFF 2

Dans sa version actuelle, l'AttrakDiff 2 se présente sous forme de 28 paires de mots contrastés, aussi appelés différenciateurs sémantiques (Osgood, Suci, & Tannenbaum, 1957). L'évaluation, réalisée de manière autonome par les participants, est faite via des échelles bipolaires en 7 points présentant respectivement à chaque extrémité un mot et son antonyme. Ainsi, des paires de mots telles que « *bon* – *mauvais* » ou « *ennuyeux* – *captivant* » sont présentées aux participants. Au niveau de la structure conceptuelle, quatre échelles comprenant chacune 7 paires de mots évaluent respectivement : les aspects pragmatiques (QP), les aspects hédoniques - divisés en deux sous-échelles : identification (QH-I) et stimulation (QH-S) - et l'attractivité globale du système interactif (ATT).

L'outil est disponible gratuitement en allemand et en anglais (http://attrakdiff.de). La passation peut se faire en ligne et les résultats de l'étude sont produits sous forme d'un rapport automatisé. Notons qu'il existe également une version abrégée de l'AttrakDiff (« abridged AttrakDiff »), qui ne comprend que 10 items (Hassenzahl, 2010 ; Hassenzahl & Monk, 2010), choisis parmi les 28 items de l'échelle, pour leur représentativité des construits. Ainsi, 4 items ont été sélectionnés pour mesurer les aspects pragmatiques (QP2, QP3, QP5, QP6), 4 items pour les aspects hédoniques (QHS2, QHS5, QHI3, QHI4) et 2 items pour l'évaluation globale de l'attractivité (ATT2, ATT5). Cette échelle abrégée a été testée et validée lors d'une étude sur quatre échantillons hétérogènes (Hassenzahl & Monk, 2010).

1.3. LA VALIDATION TRANSCULTURELLE DE L'ATTRAKDIFF

L'AttrakDiff a fait l'objet de plusieurs traductions dans différentes langues, principalement européennes. Développé à l'origine en langue allemande, le questionnaire AttrakDiff a ensuite fait l'objet d'une traduction simple (sans procédure de traduction rigoureuse) en langue anglaise. Des éléments sur les caractéristiques de cette version anglaise ont été publiés par van Schaik et Ling (2008) dans le cadre d'une étude sur 111 participants utilisant plusieurs variantes du site intranet d'un département universitaire. Une analyse factorielle réalisée sur les 21 items des échelles pragmatiques, hédonique-identification et hédonique-stimulation confirme une structure en 3 facteurs avec une consistance interne élevée (alpha de Cronbach de .83 pour QHI à .90 pour QP). Néanmoins, van Schaik et Ling précisent avoir supprimé 6 items pour parvenir à une solution factorielle satisfaisante : QP_1, HQI_1, HQI_4, HQI_5, HQI_6 et HQS_6, sans donner plus de détails sur les raisons de leur suppression. Avec 4 items supprimés sur 7, la sous-échelle QHI est incontestablement la plus problématique dans cette étude. Trois de ces 4 items réfèrent directement à l'intégration sociale : QHI_1 (*m'isole/me sociabilise*), QHI_5 (*m'exclut/m'intègre*) et QHI_6 (*me rapproche des autres/me sépare des autres*). Le cas d'usage choisi n'impliquant pas de caractère social ou relationnel, la formulation de ces items a pu paraître inadaptée.

L'AttrakDiff a également été traduit en islandais (Isleifsdottir & Larusdottir, 2008), en estonien (Peedu, 2011) ou encore en finlandais (Holm & Lehtiniemi, 2011 ; Raita & Oulasvirta, 2011). Toutefois, tout comme pour la version anglaise, le point commun de ces traductions semble être leur caractère insuffisamment méthodique, avec des processus de validation inexistants ou incomplets des versions traduites. Dans l'étude de Isleifsdottir et Larusdottir (2008), la traduction de l'AttrakDiff de l'anglais vers l'islandais a été réalisée par une seule personne et il n'est pas précisé si elle a fait l'objet de validation avant la passation. Dans la version islandaise, l'item QHI_5 (*alienating/integrating*) a été exclu de l'échelle car la traductrice n'a pas su trouver une traduction distincte entre cette paire d'items et une autre paire très similaire. L'échelle islandaise a été administrée à un petit échantillon de 10 personnes réalisant un test utilisateur. Les scores de consistance interne des différentes échelles de l'outil sont indiqués dans l'article mais ne sont, selon les auteurs, « *malheureusement pas aussi élevés que ceux mesurés précédemment par Hassenzahl* » (p. 100), les alphas de Cronbach allant de .42 (QH-S) à .58 (QP) avant usage et de .46 (QH-I) à .86 (QP). Les limites de cette étude sont nombreuses : au-delà d'une traduction réalisée par une seule personne et testée sur un très petit échantillon, il faut souligner également que la traduction n'a pas été faite sur base de l'instrument source en allemand mais sur base de la version anglaise de l'AttrakDiff, qui elle-même n'a jamais été rigoureusement validée. Les auteurs proposent des modifications dans la traduction de certains items, et font également l'hypothèse que l'échelle hédonique-stimulation pourrait ne pas se conformer au modèle théorique dans le cas de logiciels professionnels.

Pour l'étude finlandaise de Raita et Oulasvirta (2011), le processus de traduction semble avoir été plus élaboré, avec plusieurs itérations, mais l'article ne renseigne pas sur la structure factorielle de l'échelle sur l'échantillon utilisé ni sur la consistance interne des différentes sous-

échelles. Enfin, l'étude estonienne de Peedu (2011) ne donne absolument aucune information sur la traduction et la validation du questionnaire.

Comme le rappellent Chang, Chau, et Holroyd (1999), la validité des études utilisant des instruments traduits peut être remise en question dès lors que peu d'attention a été portée aux procédures déterminant l'équivalence entre l'instrument original et la version traduite. L'application d'un instrument à une population nouvelle implique bien plus que la simple production du texte dans une autre langue, l'administration de la version traduite et la comparaison des résultats (van de Vijver & Hambleton, 1996). De nombreuses autres problématiques doivent être considérées consciencieusement. Van de Vijver et Poortinga (1997) distinguent ainsi trois grands types de biais pouvant menacer l'adéquation d'un instrument traduit : les biais de construit (liés à la non équivalence des construits entre les groupes culturels), les biais de méthode (résultant de problèmes d'administration de l'instrument) et les biais d'items (souvent liés à des problèmes de traduction. Dans notre étude, une attention particulière a été portée à l'application de pratiques standardisées et valides pour la traduction d'instruments de mesure (International Test Commission, 2010).

2. METHODE

2.1. PARTICIPANTS

Quatre-cent six participants ont participé à une étude en ligne. Parmi eux, 25 réponses ont été considérées comme non valides car les participants n'avaient pas le français pour langue maternelle. 381 réponses uniques et valides ont ainsi été traitées.

Notre échantillon est constitué de 240 femmes et 141 hommes, ayant tous le français pour langue maternelle. L'âge moyen des participants est de 32 ans (σ = 12,15 ; *Min* = 16, *Max* = 78). Sur une échelle allant de 1 (pas du tout) à 7 (tout à fait), les participants se déclarent majoritairement à l'aise avec l'utilisation des technologies (M = 5,86, σ = 1,21). 90,8% d'entre eux sont inscrits sur Facebook depuis plusieurs années et seulement 3,4% déclarent visiter le site pour la première fois. En termes de fréquence de visite, 74% des participants déclarent visiter Facebook chaque jour ou plusieurs fois par jour.

2.2. PROCEDURES

2.2.1. TRADUCTION DE L'ATTRAKDIFF EN FRANÇAIS : PREPARATION D'UNE VERSION PRELIMINAIRE

Une version préliminaire du questionnaire en français a été réalisée par la traduction de l'échelle originale allemande. L'échelle anglaise a également été utilisée pour calibrer ou affiner le sens des items les plus complexes à traduire. Cette traduction initiale a été réalisée par deux experts chercheurs en Interaction Homme-Machine (IHM) : le premier auteur accompagné d'un expert germanophone de langue maternelle. Il est important de préciser que l'utilisation de l'échelle anglaise pour affiner la traduction

française semble pertinente au vu de sa popularité et de sa grande diffusion. Bien que des différences notoires existent entre version originale allemande et version anglaise, nous avons recherché dans cette traduction de l'outil en français une certaine homogénéité à la fois avec les versions allemandes et anglaises, et ce, afin d'être le plus fidèle possible à l'esprit de l'outil et aux construits mesurés par chacun des items.

La justesse de cette première version française a ensuite été évaluée et affinée par un processus de traduction renversée. Ce processus de traduction nécessite, une fois une version préliminaire créée, de demander à d'autres experts bilingues de traduire l'échelle depuis la version traduite dans sa langue d'origine. La justesse de la traduction correspond ainsi au degré selon lequel elle permet de reproduire fidèlement la version originale (Vallerand, 1989). Selon Vallerand, cette technique permet d'éviter les biais liés à l'interprétation et à la subjectivité du chercheur, fréquemment observables dans le cas d'une traduction traditionnelle. C'est la méthode idéale pour évaluer et corriger une version préliminaire traduite (Brislin 1986; Brislin, Lonner, & Thorndike, 1973). Suivant les recommandations de Vallerand, deux experts bilingues (experts également au niveau de la recherche) ont été sollicités pour réaliser chacun en parallèle une traduction renversée. Sur 56 mots au total (28 paires) ayant fait l'objet d'une traduction renversée, 21 ont été traduits conformément à la version originale allemande. Les autres termes ont été examinés et validés par une approche de type comité.

2.2.2. EVALUATION ET MODIFICATION DE LA VERSION PRELIMINAIRE PAR APPROCHE DE TYPE COMITE

Afin d'évaluer et de valider le choix de la version préliminaire de l'outil, une approche de type comité a été adoptée. Cette approche permet de réduire les biais linguistiques, théoriques et psychologiques susceptibles d'apparaître dans le cas où le chercheur initiant la traduction de l'outil serait amené à valider lui-même la version préliminaire sans avis extérieur objectif (Vallerand, 1989).

Le comité d'évaluation était composé de cinq personnes, comprenant deux chercheurs trilingues spécialisés en expérience utilisateur ayant réalisé la traduction initiale, deux chercheurs trilingues en charge des traductions renversées, et un chercheur externe spécialisé en UX et ayant utilisé régulièrement l'AttrakDiff dans sa version anglaise. La double expertise linguistique et scientifique des membres du comité d'évaluation permet d'assurer une meilleure adéquation de l'instrument traduit avec le construit théorique dans lequel il s'inscrit (Van de Vivjer & Hambleton, 1996). De plus, la traduction d'une échelle de mesure requiert une connaissance approfondie du langage source comme du langage cible (Van de Vivjer & Hambleton, 1996). C'est pourquoi nous avons favorisé la mixité culturelle et le trilinguisme des membres du comité.

Les termes ayant récolté un consensus durant le processus de traduction renversée ont été rapidement validés par le comité, qui s'est ensuite concentré sur les paires de mots plus problématiques. Les choix du comité ont été dictés à la fois par le sens du terme dans la langue originale mais

également par la cohérence de chaque paire de mot à l'intérieur de sa sous-échelle et par rapport aux autres paires de mots de l'outil. En effet, la traduction mot à mot d'une langue à une autre n'est pas appropriée pour rendre compte des différences linguistiques et culturelles pouvant exister entre les deux pays (Hilton & Skrutkowski, 2002). Les traductions retenues par le comité pour chaque paire de mot sont présentées dans le Tableau 1.

Tableau 1. Présentation des items des versions originales allemande et anglaise de l'AttrakDiff et de leur traduction française validée par comité

	Version originale (DE)	Version Originale (EN)	Validation comité (FR)
Qualité Pragmatique (QP)	Technisch - Menschlich	Technical - Human	Technique - Humain
	Kompliziert - Einfach	Complicated - Simple	Compliqué - Simple
	Unpraktisch - Praktisch	Impractical - Practical	Pas pratique - Pratique
	Umständlich - Direkt	Cumbersome - Straightforward	Fastidieux - Efficace
	Unberechenbar - Voraussagbar	Unpredictable - Predictable	Imprévisible - Prévisible
	Verwirrend - Übersichtlich	Confusing - Clearly structured	Confus - Clair
	Widerspenstig - Handhabbar	Unruly - Manageable	Incontrôlable - Maîtrisable
Qualité Hédonique Identification (QH-I)	Isolierend - Verbindend	Isolating - Connective	M'isole - Me sociabilise
	Laienhaft - Fachmännisch	Unprofessional - Professional	Amateur - Professionnel
	Stillos - Stilvoll	Tacky - Stylish	De mauvais goût - De bon goût
	Minderwertig - Wertvoll	Cheap - Premium	Bas de gamme - Haut de gamme
	Ausgrenzend - Einbeziehend	Alienating - Integrating	M'exclut - M'intègre
	Trennt mich - Bringt mich näher	Separates me - Brings me closer	Me sépare des autres - Me rapproche
	Nicht vorzeigbar - Vorzeigbar	Unpresentable - Presentable	Non présentable - Présentable
Qualité Hédonique Stimulation (QH-S)	Konventionell - Originell	Conventional - Inventive	Conventionnel - Original
	Phantasielos - Kreativ	Unimaginative - Creative	Sans imagination - Créatif
	Vorsichtig - Mutig	Cautious - Bold	Prudent - Audacieux
	Konservativ - Innovativ	Conservative - Innovative	Conservateur - Novateur
	Lahm - Fesselnd	Dull - Captivating	Ennuyeux - Captivant
	Harmlos - Herausfordernd	Undemanding - Challenging	Peu exigeant - Challenging
	Herkömmlich - Neuartig	Ordinary - Novel	Commun - Nouveau
Attractivité (ATT)	Unangenehm - Angenehm	Unpleasant - Pleasant	Déplaisant - Plaisant
	Hässlich - Schön	Ugly - Attractive	Laid - Beau
	Unsympathisch - Sympathisch	Disagreeable - Likeable	Désagréable - Agréable
	Zurückweisend - Einladend	Rejecting - Inviting	Rebutant - Attirant
	Schlecht - Gut	Bad - Good	Mauvais - Bon
	Abstossend - Anziehend	Repelling - Appealing	Repoussant - Attrayant
	Entmutigend - Motivierend	Discouraging - Motivating	Décourageant - Motivant

Une difficulté rencontrée par le comité de validation a été, pour certaines paires de mots, la non cohérence entre les versions originales allemande et anglaise. Ainsi, les deux paires de mots « unangenehm – angenehm » et « unsympathisch – sympathisch », respectivement traduites par « unpleasant – pleasant » et « disagreeable – likeable », ont des traductions qui semblent inversées. « unangenehm » en allemand signifie en effet plutôt « disagreeable » tandis que « unsympathisch » signifierait plutôt « unpleasant ». Ces paires de mots appartenant à la même échelle, et ayant dans tous les cas un sens très proche, l'impact de cette distinction ne devrait pas se ressentir sur la qualité globale de l'échelle traduite en français. Pour éviter les biais de méthode liés à l'administration du questionnaire, les consignes générales ainsi que le format des échelles de mesure ont également été passés en revue.

2.2.3. PRETEST DE LA VERSION EXPERIMENTALE

Une fois la version expérimentale de la traduction validée par le comité, un prétest de l'outil a été réalisé sur 26 utilisateurs. Le but de ce prétest était de valider en situation réelle la compréhensibilité de chaque paire d'items par le public ciblé. Pour ce faire, les items de l'échelle ont été présentés aux utilisateurs accompagnés, pour chaque paire de mots, d'une question portant sur l'évaluation de la compréhensibilité des paires de mots. Afin de simuler avec réalisme l'utilisation réelle de l'outil, consigne était donnée aux participants de ce prétest d'évaluer leur expérience du dernier système interactif utilisé.

Les passations ont été réalisées en face à face afin de recueillir les impressions et explications des sujets. Ce type d'administration informelle permet de vérifier notamment l'interprétation des instructions et des items par le groupe cible ainsi que la bonne compréhension de l'échelle de mesure. Selon les recommandations de van de Vijver et Hambleton (1996), ceci contribue à minimiser les biais de méthodes liés à l'administration du questionnaire. Les consignes ont été comprises par la totalité des répondants. De même, aucun problème n'a été identifié au niveau de la compréhension du format de réponse, les échelles de Likert étant communément utilisées dans les questionnaires francophones.

L'échelle de compréhensibilité des items se présentait sous forme d'une échelle de Likert en 7 points : « Cette paire de mots est-elle compréhensible ? » (de 1 « pas du tout » à 7 « tout à fait »). Afin d'être considéré compréhensible, chaque item devait atteindre une moyenne de compréhensibilité minimale de 4 sur 7. Le tableau 2 présente les scores de compréhensibilité pour chaque paire d'items. Tous les items dépassent le seuil attendu de 4 sur 7 et seuls 4 items sur 28 présentent des moyennes inférieures à 5. La moyenne de compréhensibilité globale de l'échelle est de 5,64 (σ = 1,49).

Tableau 2. Scores de compréhensibilité des items de la version française de l'AttrakDff recueillis lors d'un prétest sur 26 utilisateurs

Items		Score de compréhensibilité	
		Moyenne	Ecart-type
QP_1	Technique - Humain	4,81	1,92
QP_2	Compliqué - Simple	6,72	0,54
QP_3	Pas pratique - Pratique	6,42	0,95
QP_4	Fastidieux - Efficace	5,16	1,77
QP_5	Imprévisible - Prévisible	5,36	1,75
QP_6	Confus - Clair	6,42	0,99
QP_7	Incontrôlable - Maîtrisable	6,00	1,20
QHI_1	M'isole - Me sociabilise	5,38	1,96
QHI_2	Amateur - Professionnel	5,92	1,41
QHI_3	De mauvais goût - De bon goût	4,80	2,24
QHI_4	Bas de gamme - Haut de gamme	5,96	1,51
QHI_5	M'exclut - M'intègre	5,04	1,73
QHI_6	Me sépare des autres - Me rapproche des autres	5,42	1,68
QHI_7	Non présentable - Présentable	5,04	2,03

QHS_1	Conventionnel - Original	6,00	1,17
QHS_2	Sans imagination - Créatif	5,60	1,53
QHS_3	Prudent - Audacieux	4,50	2,04
QHS_4	Conservateur - Novateur	5,00	1,76
QHS_5	Ennuyeux - Captivant	5,67	1,35
QHS_6	Peu exigeant - Challenging	4,42	2,14
QHS_7	Commun - Nouveau	5,38	1,65
ATT_1	Déplaisant - Plaisant	6,54	0,95
ATT_2	Laid - Beau	5,88	1,63
ATT_3	Désagréable - Agréable	6,54	0,95
ATT_4	Rebutant - Attirant	6,08	1,08
ATT_5	Mauvais - Bon	6,38	0,97
ATT_6	Repoussant - Attrayant	6,12	1,30
ATT_7	Décourageant - Motivant	5,42	1,63
TOTAL		5,64	1,49

Bien que tous les items dépassent le seuil de compréhension attendu, les écarts-types élevés sur certains items (σ = 2,24 pour QHI_3 ou σ = 2,14 pour QHS_6) montrent une certaine variabilité dans la compréhension de certaines paires de mots. L'analyse de la fidélité par la consistance interne nous permettra de mieux détecter l'éventuelle inadéquation de certains items.

Ces étapes préliminaires de traduction rigoureuse de l'instrument et de test de la compréhensibilité des items sont nécessaires pour conférer à notre version traduite une validité de contenu et une validité concomitante équivalentes à celle de l'instrument d'origine. La validité de contenu sert à vérifier que les items composant un test représentent de manière adéquate l'ensemble du domaine que le test prétend mesurer (Anastasi, 1976). Dire de l'AttrakDiff qu'il est valide de contenu signifie ainsi que les 28 items qui le composent représentent bien la définition du construit. La validité concomitante quant à elle représente la corrélation entre les scores obtenus sur l'échelle et les scores obtenus sur un critère mesurant le ou les mêmes concepts (Allen & Yen, 1979).
Dans le cas d'une traduction d'instruments, il n'est pas nécessaire d'analyser ces validités à nouveau sur la version traduite. Le fait de prouver qu'il existe une équivalence transculturelle entre la version originale et la version traduite, suffit pour conférer à la version traduite ces validités concomitante et de contenu (Spielberger & Sharma, 1976 ; Vallerand, 1989). Ces caractéristiques de validité ont été établies sur le test original (Hassenzahl, 2003). La qualité de la traduction de notre version française, et la vérification par un comité mixte d'experts de son équivalence transculturelle avec la version originale, devrait ainsi être suffisante pour conférer à notre échelle traduite en français ces validités concomitante et de contenu.

2.3. PASSATION

L'étude portait sur l'évaluation de l'expérience des utilisateurs du réseau social Facebook. Pour nous assurer de toucher des utilisateurs finaux de ce système, le questionnaire a été diffusé directement sur le réseau social. Afin d'éviter tout biais lié à la compréhension des items, le seul prérequis à la participation était d'être francophone de langue maternelle. Le choix de ce mode de passation en ligne, n'impliquant donc pas la présence d'un

expérimentateur, a été fait dans un souci de validité écologique des résultats. En effet, l'expérience utilisateur est dynamique et fortement dépendante de facteurs contextuels (Roto et al., 2011). Ainsi, une expérimentation en laboratoire, telle qu'elle a été réalisée dans l'étude initiale d'Hassenzahl et al. (2003), prend d'une part le risque de dénaturer l'expérience réelle vécue et d'autre part, augmente les risques de biais liés à la désirabilité sociale.

La consigne de l'étude était assez générique : « *Dans le cadre d'un projet de recherche sur l'expérience utilisateur des systèmes interactifs, nous souhaiterions évaluer votre expérience du site web Facebook* ». Il n'était pas précisé aux participants que l'échelle française faisait l'objet d'une validation. Les consignes originales de l'AttrakDiff, elles aussi traduites en français, étaient ensuite présentées aux participants. Ces derniers complétaient ensuite les 28 items de l'AttrakDiff en français suivi d'une mesure par item unique (single-item) sur la qualité globale de leur expérience. Enfin, des informations sociodémographiques telles que le sexe, l'âge, la langue maternelle, le niveau d'expertise avec les technologies étaient recueillies. La durée d'inscription sur le site de Facebook ainsi que la fréquence de visite ont également été recueillies pour renseigner sur l'usage du système.

2.4. MESURES

2.4.1. QUESTIONNAIRE ATTRAKDIFF 2

Le questionnaire AttrakDiff 2 (Hassenzahl et al., 2003) est un auto-questionnaire composé de 28 items sous forme de différenciateurs sémantiques cotés sur une échelle de Likert à 7 points. Ces 28 items composent quatre sous-échelles, chaque sous-échelle étant formée de 7 items. Les sous-échelles de l'AttrakDiff 2 sont les suivantes : Qualité Pragmatique (QP), Qualité Hédonique-Stimulation (QH-S), Qualité Hédonique-Identification (QH-I) et enfin Attractivité (ATT). Notons que les items ne sont pas regroupés par sous-échelles mais présentés dans le même ordre que sur le site http://attrakdiff.de, la valence de certains items étant par ailleurs alternée pour éviter la tendance à l'acquiescement. Ainsi, le score des items QP1, QP2, QP3, QP5, QHS1, QHS3, QHS4, QHS7, QHI2, QHI3, QHI6, ATT1, ATT3, ATT5, ATT7 est inversé. Lors des analyses statistiques, ces scores ont été renversés pour que la valence de chaque item aille toujours du négatif au positif.

2.4.2. MESURE DE LA QUALITE GLOBALE DE L'EXPERIENCE PAR ITEM UNIQUE

Les participants étaient invités à évaluer la qualité globale de leur expérience avec Facebook sur une échelle de 1 à 100 (1 désignant le moins bon score et 100 désignant le meilleur score). Notre hypothèse est que le score d'expérience recueilli par une traduction valide de l'AttrakDiff devrait corréler avec la variable générique de qualité globale de l'expérience (variable que nous nommerons QUAL_UX) mesurée par un item unique. L'utilisation de l'item unique QUAL_UX a pour but ici de proposer une mesure complémentaire de validité concomitante de la version française de l'AttrakDiff, en l'absence d'outil comparable en langue française reconnu comme valide et fiable.

2.5. ANALYSE DES DONNEES

Toutes les analyses statistiques ont été réalisées à l'aide du logiciel SPSS version 22. Les données utilisées ne comportent aucune donnée manquante. La structure factorielle de la version française de l'AttrakDiff 2 a été évaluée par une analyse en composantes principales avec rotation Varimax. La consistance interne du questionnaire AttrakDiff 2 en français a été évaluée en calculant l'alpha de Cronbach pour chaque sous-échelle.

3. RESULTATS

Le tableau 3 présente les moyennes et écarts-types recueillis pour chaque item de la version française de l'AttrakDiff. Pour une meilleure lisibilité, ceux-ci sont présentés par sous-échelles.

Tableau 3. Statistiques descriptives des sous-échelles de la version française de l'AttrakDiff

	N	Min	Max	Moyenne	Ecart-type
QP_1 [Humain | Technique]	381	1	7	3,94	1,53
QP_2 [Simple | Compliqué]	381	1	7	4,85	1,63
QP_3 [Pratique | Pas pratique]	381	1	7	4,98	1,54
QP_4 [Fastidieux | Efficace]	381	1	7	4,63	1,39
QP_5 [Prévisible | Imprévisible]	381	1	7	4,64	1,39
QP_6 [Confus | Clair]	381	1	7	4,31	1,57
QP_7 [Incontrôlable | Maîtrisable]	381	1	7	3,48	1,78
Qualité Pragmatique – total (α = .75)	381	1	6,86	4,41	0,98
ATT_1 [Plaisant | Déplaisant]	381	1	7	4,74	1,44
ATT_2 [Laid | Beau]	381	1	7	4,06	1,21
ATT_3 [Agréable | Désagréable]	381	1	7	4,86	1,35
ATT_4 [Rebutant | Attirant]	381	1	7	4,71	1,31
ATT_5 [Bon | Mauvais]	381	1	7	4,29	1,38
ATT_6 [Repoussant | Attrayant]	381	1	7	4,74	1,23
ATT_7 [Motivant | Décourageant]	381	1	7	4,09	1,16
Attractivité – total (α = .88)	381	1	6,71	4,5	1
QHI_1 [M'isole | Me sociabilise]	381	1	7	4,35	1,44
QHI_2 [Professionnel | Amateur]	381	1	7	3,62	1,57
QHI_3 [De bon goût | De mauvais goût]	381	1	7	4,20	1,27
QHI_4 [Bas de gamme | Haut de gamme]	381	1	7	3,86	1,12
QHI_5 [M'exclut | M'intègre]	381	1	7	4,41	1,21
QHI_6 [Me rapproche des autres | Me sépare]	381	1	7	4,53	1,35
QHI_7 [Non présentable | Présentable]	381	1	7	4,69	1,21
Qualité Hédonique Identité – total (α = .77)	381	1	6,29	4,24	0,86
QHS_1 [Original | Conventionnel]	381	1	7	4,24	1,49
QHS_2 [Sans imagination | Créatif]	381	1	7	4,30	1,38
QHS_3 [Audacieux | Prudent]	381	1	7	4,41	1,19
QHS_4 [Novateur | Conservateur]	381	1	7	4,72	1,24
QHS_5 [Ennuyeux | Captivant]	381	1	7	4,38	1,36
QHS_6 [Peu exigeant | Challenging]	381	1	7	3,38	1,37
QHS_7 [Nouveau | Commun]	381	1	7	3,91	1,44
Qualité Hédonique Stimulation – total (α = .78)	381	1	6,43	4,19	0,89
ATTRAKDIFF total	381	1,43	6,04	4,33	0,75
QUAL_UX (Score sur 100 – qualité expérience)	381	0	99	61,78	21,3

L'évaluation de la validité de construit de notre version française a pour objectif de vérifier si la version traduite permet bien de mesurer les qualités pragmatiques et hédoniques d'un système conformément au modèle théorique initial (Hassenzahl et al., 2001). Il s'agit également de montrer si le modèle théorique sous-jacent à l'AttrakDiff s'avère valide dans la culture française. Nous évaluerons ici la validité de construit par l'étude de la structure du construit dans un premier temps, et par l'étude des relations entre les différentes sous-échelles dans un second temps.

3.1. STRUCTURE FACTORIELLE

Une analyse factorielle en composantes principales (ACP) a été réalisée pour tester la validité de construit de notre version française de l'AttrakDiff (rotation Varimax, critère de Kaiser > 1). En adéquation avec le modèle théorique d'Hassenzahl et al. (2003), seuls les 21 items des sous-échelles pragmatique (QP), hédonique-stimulation (QH-S) et hédonique-identification (QH-I) ont été inclus dans l'analyse, l'échelle d'attractivité (ATT) étant traitée séparément. En effet, le modèle précise que l'attractivité (sous-échelle ATT) est un jugement qui découle de la perception des attributs pragmatiques et hédoniques. La validation du construit consistera à vérifier que la dimension d'attractivité est prédite par la perception des attributs pragmatiques et hédoniques.

Un examen de la matrice de corrélation indique que la majorité des items sont corrélés positivement et présentent des corrélations > .3. Seul l'item QP_5 (*Prévisible/Imprévisible*) apparaît comme problématique à ce stade puisqu'il présente des corrélations très faibles avec les autres items de l'échelle, dont aucune n'est supérieure à .3.

L'analyse de la qualité de représentation des items montre que tous les items ont une qualité supérieure à .20, à l'exception de QHI_2 (*Professionnel/Amateur*) qui présente une qualité de représentation très faible de .09. L'analyse sémantique de cet item révèle toutefois un potentiel biais de méthode lié à la plateforme utilisée pour l'étude. En effet, le réseau social en ligne Facebook est développé et géré par une équipe de professionnels, mais son contenu est produit par les utilisateurs, donc des amateurs. Bien que la faible qualité de représentation de QHI_2 ne suffise théoriquement pas à justifier le maintien de cette variable dans l'analyse, nous décidons de conserver cet item car nous ne sommes pas en mesure de savoir s'il est non valide, mal traduit ou s'il est affecté par un biais de méthode. De plus, l'item ne semble pas problématique dans l'analyse factorielle de la version anglaise de l'AttrakDiff réalisée par van Schaik et Ling (2008). Dans la section Discussion, nous suggérons de répliquer cette étude en utilisant le questionnaire français sur d'autres cas d'usage qui n'auraient pas de connotation sémantique en lien avec cet item.

Le test de sphéricité de Bartlett significatif, $\chi 2\ (210, N = 381) = 2781.49, p < .001$, et la mesure d'adéquation de l'échantillon de Kaiser-Meyer-Olkin (KMO) de .87 indiquent que la matrice de corrélation peut être soumise à l'analyse factorielle.

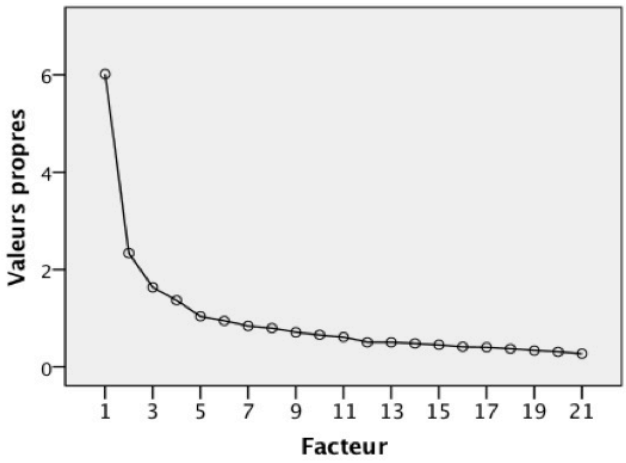


Figure 2. Graphe des valeurs propres pour les scores à l'AttrakDiff des 381 participants de l'échantillon

Bien que le critère de Kaiser extrait 5 facteurs ayant des valeurs propres supérieures à 1, le test des éboulis de Cattell (*scree test* ; Cattell, 1966) suggère une structure à 3 facteurs, conforme au modèle théorique (Figure 2). Une nouvelle ACP avec extraction en 3 facteurs est alors réalisée. La structure factorielle obtenue sur notre échantillon de 381 participants est présentée dans le Tableau 4.

Tableau 4. Analyse factorielle en composantes principales avec Rotation Varimax

	Composante		
	QH-I	QH-S	QP
QP_1 [Humain | Technique]	.44		
QP_2 [Simple | Compliqué]			.84
QP_3 [Pratique | Pas pratique]			.65
QP_4 [Fastidieux | Efficace]	.37		.64
QP_5 [Prévisible | Imprévisible]			.44
QP_6 [Confus | Clair]			.73
QP_7 [Incontrôlable | Maîtrisable]	.48		.31
QHI_1 [M'isole | Me sociabilise]	.70		
QHI_2 [Professionnel | Amateur]	(.26)		
QHI_3 [De bon goût | De mauvais goût]	.60		
QHI_4 [Bas de gamme | Haut de gamme]	.61	.30	
QHI_5 [M'exclut | M'intègre]	.80		
QHI_6 [Me rapproche des autres | Me sépare des autres]	.70		
QHI_7 [Non présentable | Présentable]	.59		
QHS_1 [Original | Conventionnel]		.65	
QHS_2 [Sans imagination | Créatif]	.42	.65	
QHS_3 [Audacieux | Prudent]		.67	
QHS_4 [Novateur | Conservateur]		.77	
QHS_5 [Ennuyeux | Captivant]	.55	.44	
QHS_6 [Peu exigeant | Challenging]	.32	.42	-.39
QHS_7 [Nouveau | Commun]		.63	
Valeur propre	6,01	2,34	1,64
% de variance expliquée	28,64	11,12	7,79

N.B : Les saturations inférieures à < .30 ne sont pas affichés dans le tableau

Cette solution factorielle en 3 facteurs explique 47,5 % de la variance totale. La première composante, qui représente la qualité hédonique-identification (QH-I) explique 28,6% de la variance totale. La saturation des items est globalement élevée (de .59 à .80), mis à part l'item QHI_2 (*Professionnel/Amateur*) qui sature faiblement sur sa composante (.26). Cet item a déjà été identifié comme problématique lors de l'analyse de la qualité de représentation. L'item QHI_4 sature à la fois sur sa composante hédonique-identification (.61) et sur la composante hédonique-stimulation (.30). On remarque également que plusieurs items des composantes pragmatiques et hédoniques-stimulation saturent sur cette dimension hédonique-identification.

La seconde composante extraite, la qualité hédonique-stimulation (QH-S) explique 11,1% de la variance totale. La saturation des items est un peu moins élevée que pour l'échelle QH-I (de .42 à .77). De plus, trois items de cette échelle saturent également sur l'échelle QH-I : QHS_2 (*Sans imagination/Créatif*) (.42), QHS_5 (*Ennuyeux/Captivant*) (.55) et QHS_6 (*Peu exigeant/Challenging*) (.32). Ce dernier item sature par ailleurs négativement sur la composante pragmatique QP (-.39).

Enfin, la troisième et dernière composante extraite est la qualité pragmatique, qui explique 7,8% de la variance totale. La saturation des items est là encore quelque peu hétérogène (de .31 à .84). Les items QP_4 et QP_7 saturent également sur la dimension QH-I (.37 et .48 respectivement). L'item QP_1 (Humain/Technique) est problématique puisqu'il sature uniquement sur la composante QH-I et non sur la composante QP dont il fait initialement partie. C'est le seul item parmi les 28 items de l'échelle qui présente cette particularité. Ces résultats, et notamment l'identification des items problématiques, seront discutés plus loin dans l'article.

3.2. ANALYSE DE LA FIDELITE PAR LA CONSISTANCE INTERNE

La fidélité d'un instrument de mesure concerne sa précision, un outil fidèle mesurant toujours le construit de la même façon (Nunnally, 1970 ; Vallerand, 1989). La méthode d'évaluation la plus commune pour déterminer la fidélité d'un test est la méthode test-retest, qui consiste à administrer le même test à deux reprises sur les mêmes individus. Si le test s'avère fidèle, alors la corrélation entre les deux séries de score devra être élevée (idéalement égale à 1, mais des erreurs de mesure provoquent des variations dans les scores). Cette méthode suppose cependant une stabilité dans l'évaluation et semble donc inappropriée dans le cas de l'AttrakDiff. L'évaluation de l'expérience utilisateur fluctuant sous l'influence de nombreux facteurs (Law, van Schaik, & Roto, 2013), une méthode d'évaluation de la fidélité n'impliquant pas une double passation est nécessaire dans le cas de notre étude. Afin d'estimer la fidélité de notre traduction française de l'AttrakDiff en une seule passation, nous nous sommes par conséquent basés sur l'évaluation de la consistance interne de l'outil, qui sert à vérifier que les items d'une même dimension sont bien homogènes. Dans un instrument de mesure consistant, les items d'une même dimension (mesurant le même attribut) sont sensés produire des

scores fortement corrélés. La consistance interne est notamment estimée par le coefficient alpha de Cronbach. Elle est jugée bonne lorsque la valeur de l'alpha de Cronbach est supérieure à .70 (Nunnally, 1978).

La consistance interne des sous-échelles est satisfaisante avec des scores respectifs de α = .75 (QP), α = .77 (QH-I), α = .78 (QH-S) et des corrélations items-total toujours positives et supérieures à .25. Les corrélations inter-échelles faibles à moyennes (de .27 à .54 avec une moyenne de .42), suggèrent une bonne distinction entre les sous-échelles. Ces résultats sont conformes aux pointages optimaux escomptés dans le cas d'une traduction, étant égaux ou très légèrement inférieurs seulement aux scores décrits dans l'étude de validation initiale de l'AttrakDiff (Hassenzahl et al., 2003) (Tableau 5).

Tableau 5. Consistance interne des échelles de la traduction française et de la version originale allemande de l'échelle AttrakDiff

Echelle	Consistance interne (Alpha de Cronbach)	Consistance interne échelle d'origine (Hassenzahl, 2003)
Qualité Hédonique-Identification (QH-I)	.77	.73 - .83
Qualité Hédonique-Stimulation (QH-S)	.78	.76 - .90
Qualité Pragmatique (QP)	.75	.83 - .85
Attractivité (ATT)	.88	non renseigné

3.3. ETUDE DES RELATIONS ENTRE LES ECHELLES

Les corrélations inter-échelles sont présentées dans le Tableau 6. Tout comme dans l'étude de validation initiale d'Hassenzahl et al. (2003), les deux échelles hédoniques QH-I et QH-S sont corrélées (r = .47, p < .01), témoignant du lien sémantique fort des deux échelles mais permettant tout de même de les distinguer. D'autre part, l'échelle de qualité pragmatique QP est également corrélée avec les échelles hédoniques QH-I (r = .54, p < .001) et QH-S (r = .27, p < .001).

Tableau 6. Matrice de corrélation de Pearson inter-échelles (N = 381)

	QP	QH-I	QH-S	ATT
QP	1			
QH-I	.54**	1		
QH-S	.27**	.47**	1	
ATT	.61**	.73**	.58**	1

** La corrélation est significative au niveau .01 (test bilatéral).

- La corrélation est significative au niveau .05 (test bilatéral).

Le modèle théorique sous jacent à l'AttrakDiff, selon lequel les qualités pragmatiques (QP) et hédoniques (QH-I et QH-S) perçues d'un système contribuent à l'évaluation de l'attractivité globale perçue (ATT), a été testé à l'aide d'un modèle de régression linéaire. Le modèle est significatif, B = -.35, $t(377)$ = -1.97, p = .05 et explique 67 % de la variance dans l'évaluation de l'attractivité, R^2 = .67, $F(3, 377)$ = 254,56, p < .001. Les relations entre les composantes de l'AttrakDiff français sont donc similaires aux relations décrites dans le modèle théorique initial. Ces données soutiennent la validité de construit de notre échelle traduite.

Parmi les variables sociodémographiques, la fréquence à laquelle les utilisateurs visitent le site de Facebook est positivement corrélée à l'évaluation globale du système par l'AttrakDiff (r = .43, p < .001) et explique 18% de la variance du score global à l'AttrakDiff, B = 3.04, R^2 = .18, p < .01. De plus, le nombre d'années depuis lequel les utilisateurs sont inscrits sur Facebook est corrélé positivement avec l'évaluation du système (r = .25, p < .01). On constate une faible corrélation négative entre âge des utilisateurs et qualité de l'expérience (r = -.18, p < .01 pour AttrakDiff global et r = -.26, p < .01 pour QUAL_UX). En revanche, on ne constate pas de différence significative en fonction du sexe des utilisateurs.

3.4. CORRELATION ENTRE ATTRAKDIFF ET ITEM UNIQUE DE MESURE DE LA QUALITE DE L'EXPERIENCE

Conformément à nos hypothèses, le score total à l'AttrakDiff, obtenu par le score moyen des 28 items (α = .91), est fortement corrélé (r = .73, p < .01) à la question demandant aux utilisateurs d'évaluer la qualité globale de leur expérience (QUAL_UX) sur une échelle de 1 à 100. Bien que cet item unique de mesure de la qualité de l'expérience n'ait pas encore été validé, il a été montré dans plusieurs études que des items uniques peuvent constituer des mesures fiables et valides (Christophersen & Konradt, 2011 ; Nichols & Webster, 2013 ; Sauro & Dumas, 2009) sous réserve que l'item utilisé soit focalisé et non ambigu (Tractinsky, Katz, & Ikar, 2000 ; Wanous & Hudy, 2001). L'avantage d'une mesure par item unique est sa validité de contenu, puisque cette dernière se résume à la formulation de l'item. Dans notre cas, la validité de contenu de l'item QUAL_UX semble assurée puisque cet item demande aux utilisateurs d'évaluer la « qualité globale de leur expérience ». Une absence de corrélation entre QUAL_UX et le score à l'AttrakDiff français aurait fortement questionné nos résultats ou aurait jeté le doute sur la validité de l'item unique. La forte corrélation de r = .73 entre QUAL_UX et AttrakDiff global a peu de chances d'être fortuite. Selon le modèle sous-jacent à l'AttrakDiff, on peut par ailleurs s'attendre à ce que QUAL_UX corrèle plus fortement avec l'échelle ATT qu'avec les autres facteurs (QP, QH-I et QH-S) puisque ces derniers se combinent pour constituer le jugement d'attractivité. C'est le cas avec une corrélation entre QUAL_UX et ATT (r = .75) supérieure aux corrélations de l'item unique avec QP (r = .52), QH-S (r = .44) ou QH-I (r = .64). Ces résultats tendent donc à soutenir d'une part la validité et la pertinence de QUAL_UX en tant que critère d'évaluation de la validité concomitante et d'autre part la qualité de notre version française de l'AttrakDiff en soutenant l'hypothèse d'une validité concomitante satisfaisante.

4. DISCUSSION

Les résultats de l'étude menée sur la qualité de l'expérience utilisateur de Facebook auprès de 381 participants tendent à montrer de bonnes propriétés psychométriques de notre traduction française de l'AttrakDiff. Le questionnaire semble fidèle et présente une bonne consistance interne, comme en témoignent les alphas de Cronbach supérieurs à .75 (de .75 à .78) obtenus pour chacune des sous-échelles. Au niveau de la validité de construit, la structure factorielle en 3 facteurs est globalement conforme à la structure factorielle de la version d'origine de l'outil, et surtout au modèle

théorique sous-jacent. Les résultats mettent toutefois en lumière certains problèmes qu'il est d'autant plus important de discuter que la validité de l'étude de validation initiale de l'outil réalisée par Hassenzahl et al. (2003) est discutable.

Nous observons tout d'abord des recouvrements entre les échelles QH-S et QH-I, avec plusieurs items saturant sur les deux échelles. En effet, trois items de l'échelle QH-S saturent également sur l'échelle QH-I : QHS_2 (*Sans imagination/Créatif*) (.42), QHS_5 (*Ennuyeux/Captivant*) (.55) et QHS_6 (*Peu exigeant/Challenging*) (.32). De même, l'item QHI_4 (*Bas de gamme/Haut de gamme*) sature également sur QH-S (.30). Ces liens entre les deux composantes se retrouvent également dans la version originale avec une corrélation inter-échelles de $r = .55$. Hassenzahl (2003) note à ce sujet que ces deux composantes QH-S et QH-I sont naturellement plus liées entre elles (de par leur contenu) qu'elles ne le sont avec la composante QP. Nous pouvons donc considérer que ces recouvrements sont dus à la nature des items de l'échelle initiale et non à une mauvaise traduction du questionnaire. Par ailleurs, certains items de l'échelle QP saturent également sur QH-I : c'est le cas de QP_1 (*Humain/Technique*), QP_4 (*Fastidieux/Efficace*) et QP_7 (*Incontrôlable/Maîtrisable*). Notons à ce sujet que les saturations transversales d'items sur plusieurs sous-échelles ne sont pas surprenantes au vu de la conceptualisation des dimensions de l'AttrakDiff. En effet, si les facettes pragmatiques et hédoniques stimulation / identification sont suffisamment distinctes pour être traités comme des facteurs différenciés, elles n'en partagent pas moins, selon les contextes d'interaction, certaines caractéristiques, comme en témoignent les corrélations inter-échelles et les études réalisées sur la nature de l'expérience utilisateur. Il serait intéressant de voir dans des études ultérieures si les patterns de saturation d'items se retrouvent de manière identique sur différents cas d'usage. Par exemple, l'item QHS_5 (*Ennuyeux/Captivant*) sature dans la présente étude à la fois sur sa composante QH-S (.44) et sur QH-I (.55). Or, on retrouve ce même pattern de saturation dans l'étude de van Schaik et Ling (2008) sur la version anglaise de l'AttrakDiff où QHS_5 sature également de manière transversale sur QH-S (.50) et sur QH-I (.43).

D'autre part, des items constituant l'outil ont été repérés comme problématiques. Ainsi, l'item QP_5 (*Prévisible/Imprévisible*) ne présente aucune corrélation supérieure à .30 et semble ainsi produire une mesure isolée des autres items de l'échelle. L'item QHI_2 (*Professionnel/Amateur*) a une qualité de représentation très faible et ne sature que faiblement (.26) sur sa composante Hédonique-Identification, tout en ne saturant fortement sur aucune autre dimension. Enfin, contrairement à la version initiale de l'outil, l'item QP_1 (*Humain/Technique*), ne sature pas dans la présente étude sur la composante Pragmatique (QP) mais plutôt sur la composante Hédonique-Identification (QH-I) (.44). Pour ces trois items, nous faisons l'hypothèse que les caractéristiques du système utilisé (le réseau social Facebook) pour cette étude de validation ont pu avoir un impact sur les évaluations des utilisateurs. Pour l'item QP_1 (*Humain/Technique*), la saturation sur la composante QH-I pourrait s'expliquer par le côté social du système évalué. Dans le cas d'un réseau social comme Facebook, demander à l'utilisateur si le réseau est « humain » ou « technique » a pu stimuler une réponse logique qui veut que l'on évalue un réseau social comme « humain » puisqu'il implique par nature des relations humaines. Appliqué à un autre cas d'étude

n'impliquant pas de relations sociales, la question prend un autre sens puisqu'alors on évalue plutôt la prise en compte de la complexité et des aspects pragmatiques de l'interaction. La saturation de cet item sur la composante Hédonique-Identification n'est pas anodine, et témoigne potentiellement du fait que le versant humain de Facebook peut soutenir l'expression de l'identité d'un utilisateur auprès de ses contacts sur le réseau social. Pour l'item QHI_2 (*Professionnel/Amateur*), là encore, le caractère social du système a pu interférer dans l'évaluation. Les utilisateurs ont pu interpréter l'item comme le fait que le contenu de Facebook soit plutôt professionnel ou amateur plutôt que de se demander si la conception globale du système (en dehors du contenu posté par ses membres) renvoie à un sentiment de professionnalisme ou d'amateurisme. Enfin, pour l'item QP_5 (*Prévisible/Imprévisible*), le versant « imprévisible » de l'item est sensé renvoyer au niveau théorique à une mauvaise utilisabilité de l'outil, qui ne soutiendrait pas un sentiment de contrôle de l'utilisateur. Or, dans le cas de Facebook, les contenus sont par nature imprévisibles puisqu'ils sont produits par les membres du réseau de l'utilisateur. Ce caractère imprévisible paraît d'ailleurs désirable puisqu'il crée une impression de nouveauté des contenus, que l'utilisateur découvre sous forme de fil d'information continu. Dans tous ces cas, la problématique semble provenir de la potentielle interprétation des items par les utilisateurs en fonction des caractéristiques du système évalué. On peut alors se demander si l'AttrakDiff ne présente pas, y compris dans sa version originale, une faiblesse concernant l'ambiguïté de certains items. En effet, la distinction entre ce qui relève de l'évaluation du système dans sa conception (sa forme) et ce qui relève de l'évaluation du système dans son contenu (le « fond », les interactions avec ses membres) n'est pas toujours claire pour les répondants. On peut d'ailleurs noter que les auteurs de l'AttrakDiff n'ont pas inclus dans leur étude de validation de phase préliminaire vérifiant la compréhensibilité par les utilisateurs finaux des items, des consignes et du format de l'échelle de mesure. Mais peut-être les auteurs originaux ont-ils considéré que ces deux versants distincts de l'évaluation du système (le fond vs. la forme) ne sont pas antinomiques, dans la mesure où seule compterait l'évaluation de la qualité telle que perçue par l'utilisateur, peu importe qu'elle provienne de l'évaluation de la forme ou du fond du système interactif concerné.

L'étude approfondie des caractéristiques des trois items problématiques montre qu'ils ne nuisent pas fondamentalement à la consistance interne de leurs sous-échelles respectives. Certains d'entre eux, tels que QHI_2 ou QP_5, semblent par ailleurs valides dans l'étude anglaise de van Schaik et Ling (2008). C'est pourquoi il nous a semblé prématuré de supprimer ces items de la version française de l'AttrakDiff. Cependant, il sera indispensable de vérifier les saturations des items concernés dans des études ultérieures (non basées sur des réseaux sociaux) afin de vérifier s'ils restent problématiques, auquel cas une reformulation ou une suppression du ou des items devra être envisagée.

De même, comme nous l'avons souligné à plusieurs reprises, l'étude de validation initiale de l'outil ainsi que les études subséquentes utilisant l'AttrakDiff en allemand, anglais, ou dans d'autres langues européennes, présentent de fortes limitations (cf. section 1.3). Si le modèle théorique d'Hassenzahl (2003) semble soutenu dans tous les cas, les analyses des

caractéristiques psychométriques de l'outil apparaissent lacunaires et les processus de traduction peu rigoureux. Au-delà de la version française, il nous semble donc indispensable que des études plus approfondies soient menées sur la version initiale allemande de l'outil ainsi que sur la version anglaise, cette dernière constituant généralement la version source pour la traduction de l'AttrakDiff dans d'autres langues.

Les travaux réalisés sur les traductions d'outils de mesure (Brislin, 1986 ; Gudmunsson, 2009 ; Hambleton & Patsula, 1999 ; Vallerand, 1989) soulignent la nécessité d'adapter les outils aux populations étudiées afin de prendre en compte les différences culturelles. Dans le présent cas, la traduction et l'adaptation de l'outil a été réalisée par un comité d'experts mixte, composé de représentants des deux cultures (germanophones et francophones). Ces experts ont ainsi pu s'assurer ensemble de l'équivalence transculturelle de chaque item lors de la traduction, en vérifiant pour chaque item non seulement la qualité de la traduction dans le sens donné à chaque mot mais également en s'assurant que la « force » du terme employé en allemand soit équivalente à la « force » du terme traduit en français. Ainsi certains adjectifs en allemand (*Isolierend / Verbindend*) ont été traduits par des expressions françaises utilisant un verbe pronominal (*M'isole/Me sociabilise*). Par ailleurs, certains termes ont fait l'objet de considérations culturelles sur la puissance émotionnelle véhiculée par le mot. C'est le cas de « *umständlich* » initialement traduit par « *laborieux* » et remplacé ensuite durant la réunion de comité par le terme « *fastidieux* ».

Bien que l'équivalence transculturelle ait été prise en compte précautionneusement lors de la traduction du questionnaire, nous pouvons tout de même nous demander si les dimensions hédoniques et pragmatiques sont bien la base de l'évaluation globale de la qualité d'un système interactif, et si l'importance de l'une ou de l'autre de ces dimensions sur l'évaluation de l'attractivité est la même pour les utilisateurs français que pour les utilisateurs allemands. En résumé, il semble intéressant de questionner le modèle théorique sous-jacent à la construction de l'AttrakDiff. Comme nous l'avons vu grâce aux analyses de régression, ce modèle théorique semble rester valide sur notre échantillon francophone, avec un impact commun des qualités pragmatiques et hédoniques (indépendantes l'une de l'autre) sur l'évaluation de l'attractivité globale. Ce qui semble changer en revanche est le poids de chacune des dimensions dans l'évaluation de l'attractivité. Cependant, nous émettons l'hypothèse que ces différences ne sont pas dues aux différences culturelles opposant utilisateurs allemands et utilisateurs français, mais plutôt aux différences d'évaluation de différents types de systèmes. Ainsi, l'évaluation des qualités pragmatiques et hédoniques d'un système dépend possiblement de l'importance accordée à chacune de ces dimensions. Dans le cas d'un jeu vidéo, on peut supposer que les qualités hédoniques vont primer sur les qualités pragmatiques dans l'évaluation de l'attractivité globale. A l'inverse, dans le cas d'un logiciel professionnel, les qualités pragmatiques prédomineront probablement sur l'hédonisme. L'étude de Isleifsdottir et Larusdottir (2008) émet la même hypothèse, appuyée par les résultats de l'évaluation de l'expérience utilisateur d'un logiciel professionnel. D'autres études sont toutefois nécessaires pour confirmer ces résultats ou, le cas échéant, identifier les expressions culturelles spécifiques de l'expérience utilisateur dans la population francophone.

Enfin, au-delà des problématiques liées à la traduction de l'outil, cette étude nous permet aussi de détecter des phénomènes intéressants concernant l'étude de l'expérience utilisateur. Ainsi, les résultats montrant des liens significatifs entre fréquence de visite et durée d'inscription sur le site Facebook et qualité de l'expérience perçue suggèrent que la familiarité avec le système mène à une évaluation plus positive de l'expérience globale. Plusieurs suggestions peuvent être avancées pour expliquer ce constat. Selon le modèle d'Hassenzahl (2003), l'attractivité d'un produit (telle que mesurée par les 7 items de la sous-échelle Attractivité) est une appréciation globale qui découle de la qualité perçue qui va être à l'origine de conséquences comportementales et émotionnelles. Or, un accroissement de l'usage peut typiquement être considéré comme une conséquence comportementale découlant de la qualité perçue. Selon cette hypothèse, ce serait donc parce que les utilisateurs ont évalué Facebook comme globalement attractif au début qu'ils ont continué à l'utiliser depuis des années et à le visiter fréquemment. Bien que cohérente, cette hypothèse n'explique pas réellement pourquoi l'évaluation des utilisateurs les plus familiers est meilleure que celle des utilisateurs les moins familiers.

Un éclairage complémentaire peut nous être apporté par l'étude de la dynamique temporelle de l'UX, qui a été décrite dans plusieurs travaux et qui fait l'objet d'un consensus au sein de la communauté (Roto et al., 2011). Les chercheurs s'accordent ainsi à dire que l'UX n'est pas un phénomène statique et que l'on peut distinguer différents types d'expérience à différents stades d'utilisation du produit : de l'UX anticipée (avant l'usage d'un système) à l'UX remémorée (après l'usage du système) ou cumulée (en comparaison avec les systèmes antérieurs ou concurrents rencontrés par l'utilisateur). Dans notre cas, la plus grande maîtrise de Facebook acquise avec le temps pourrait donner l'impression à l'utilisateur que le système est facile à utiliser et provoquer ainsi une expérience positive stimulée par un sentiment de compétence (Sheldon et al., 2001). L'utilisateur expérimenté pourrait avoir oublié l'éventuelle difficulté d'apprentissage du système auquel il a pu faire face plusieurs mois ou années auparavant. Les études menées sur les expériences montrent ainsi que celles-ci tendent à s'améliorer au fil du temps (van Boven & Gilovitch, 2003).

Une dernière hypothèse pourrait enfin expliquer ce phénomène par une volonté de maintien de cohérence de la part des utilisateurs : évaluer comme négatif un système qu'ils utilisent chaque jour depuis des années reviendrait à évaluer négativement leur habitude d'utiliser ce site. L'expérience de cette contradiction entre une cognition (i.e., penser que l'expérience utilisateur procurée par Facebook n'est pas bonne) et une action (i.e., utiliser régulièrement Facebook depuis plusieurs années) est appelée dissonance cognitive (Festinger, 1957) et place l'utilisateur dans une situation de relatif inconfort. Pour réduire cette tension induite, l'utilisateur peut alors adopter une attitude plus favorable envers l'objet de la dissonance. Selon la théorie de l'auto-perception de Bem (1967), les personnes développeraient des attitudes en observant leur propre comportement, de la même façon qu'un observateur externe. Ainsi, l'attitude développée face à un système interactif tel que Facebook serait en fait inférée par les utilisateurs sur base d'indices externes également, tels que la fréquence à laquelle ils utilisent le système.

5. CONCLUSION

Depuis une dizaine d'années, le questionnaire AttrakDiff (Hassenzahl et al., 2003) est utilisé dans la recherche en expérience utilisateur. Cependant, à notre connaissance, aucune étude de validation rigoureuse n'a été publiée à ce jour, ni pour sa version originelle allemande, ni pour ses traductions ultérieures. Dans un souci de consolidation de la recherche en IHM, la méthodologie de traduction et de validation du questionnaire AttrakDiff suivie dans cette étude a permis la création d'une version française de l'outil présentant des niveaux de validité et de fidélité satisfaisants. Afin de minimiser les biais liés au construit, à la méthode ou aux items traduits, nous avons appliqué une méthodologie de traduction et de validation rigoureuse suivant les recommandations de la commission internationale de tests (2010) et des experts en études interculturelles (Brislin, 1980 ; Brislin, 1986 ; Vallerand, 1989 ; Van de Vijver & Hambleton, 1996). Le processus de traduction renversée, puis de validation par un comité mixte composés d'experts bilingues du domaine des interactions homme-machine, a permis d'établir une version expérimentale française de l'outil. Cette version a ensuite été pré-testée sur 26 utilisateurs finaux afin de vérifier la bonne compréhensibilité des consignes, des items et du format de réponse. Enfin, une étude statistique menée sur un échantillon de 381 utilisateurs français a été menée afin de collecter des données quantitatives, indispensables à l'analyse des qualités psychométriques de notre version française de l'AttrakDiff.

Les résultats de cette étude sur la qualité de l'expérience utilisateur de Facebook supportent l'hypothèse de conformité de la version française à la version initiale allemande, en montrant que la consistance interne des sous-échelles a été préservée et que la validité concomitante, évaluée par corrélation avec un item unique d'évaluation de la qualité globale de l'expérience, semble bonne. Les résultats de notre étude supportent également le modèle théorique sur lequel repose l'AttrakDiff (Hassenzahl et al., 2000; Hassenzahl, 2002) en montrant la distinction entre qualité hédonique perçue et qualité pragmatique perçue d'un système, qui se combinent pour générer une évaluation globale de l'attractivité du produit.

La principale limitation de la présente étude est liée au cas d'usage unique utilisé, ayant un caractère social et pouvant constituer un biais de méthode impactant certains items (dont trois ont été identifiés comme problématiques). L'évaluation de la validité concomitante par la comparaison entre le score à l'AttrakDiff et le score obtenu à l'item unique QUAL_UX doit également être complétée, dans des études ultérieures, par d'autres mesures des validités concomitantes et divergentes. Ainsi, des études complémentaires sur des systèmes interactifs non sociaux sont nécessaires pour vérifier la structure interne de la version française de l'AttrakDiff et investiguer plus en profondeur les constats décrits dans cette première étude. De plus, une étude transculturelle comparant la qualité de l'expérience de Facebook chez des utilisateurs allemands à celle d'utilisateurs français pourrait confirmer ou infirmer les spécificités « sociales » de l'UX constatées dans la présente étude. De même, la comparaison des résultats de la version française de l'AttrakDiff avec d'autres échelles de mesure valides permettrait d'inspecter de manière plus

rigoureuse la validité concomitante et la validité divergente de la version française de cet outil. Enfin, nous préconisons des efforts de consolidation plus soutenus de la recherche en expérience utilisateur, notamment via la traduction et la validation transculturelle d'autres outils d'évaluation ou de conception de l'UX.

REFERENCES FOR PAPER B

Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Anastasi, A. (1976). *Psychological Testing* (4th edition). NewYork: MacMillan.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200. doi: 10.1037/h0024835

Brislin, R. W. (1980). *Translation and content analysis of oral and written material.* In H. C. Triandis & J. W. Berry (Eds.). *Handbook of cross-cultural psychology: Vol. 1* (pp. 389-444). Boston: Allyn & Bacon.

Brislin, R. W. (1986). *The wording and translation of research instruments*. In W. J. Lonner, & J. W. Berry (Eds.). *Field methods in cross-cultural research* (pp. 137-164). Newbury Park, CA: Sage.Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods*. New York: Wiley.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.). *Usability Evaluation in Industry*. London: Taylor and Francis.

Canino, G., & Bravo, M. (1999). *The translation and adaptation of diagnostic instruments for cross-cultural use.* In D. Shaffer, C. P. Lucas, & J. E. Richters (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 285–298). New York: Guilford Press.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. doi: 10.1207/s15327906mbr0102_10

Chang, A.M., Chau, J.P., & Holroyd, E. (1999). Translation of questionnaires and issues of equivalence. *Journal of Advanced Nursing*, *29* (2), 316-322.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 213-218. New York: ACM/SIGCHI.

Christophersen, T., & Konradt, U. (2011). Reliability, validity, and sensitivity of a single-item measure of online store usability. I*nternational Journal of Human-Computer Studies*, *69*, 269-280. doi: 10.1016/j.ijhcs.2010.10.005

Christou, G. (2013). A comparison between experiences and inexperienced video games players' perceptions. *Human-centric Computing and Information Sciences*, *3*(15). doi:10.1186/2192-1962-3-15

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

Eimler, S., von der Pütten, A., Schächtle, U., Carstens, L., & Krämer, N. (2010). Following the White Rabbit – A Robot Rabbit as Vocabulary Trainer for Beginners of English. In G. Leitner, M. Hitz, & A. Holzinger (Eds.), *Lecture Notes in Computer Science*: *Vol. 6389. HCI in Work and Learning, Life and Leisure* (pp. 322-339). doi: 10.1007/978-3-642-16607-5_22

Erkut, S., Alarcon, O., Garcia Coll, C., Tropp, L. R., & Garcia, H. A. V. (1999). The dual-focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology*, *30*, 206–218.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. California: Stanford University Press.

Grimm, L. G., & Yarnold, P. R. (Eds.). (1995). Reading and understanding multivariate statistics. Washington, DC: American Psychological Association.

Grün, C., Gerken, J., Jetter, H.-C., Koenig, W., & Reiterer, H. (2005). MedioVis - A user-centred library metadata browser. In A. Rauber et al. (Eds.), *Lecture Notes in Computer Science*: *Vol. 3652. Research and Advanced Technology for Digital Libraries* (pp. 174-185). doi: 10.1007/11551362_16

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, *61*(2), 29-45. doi:10.1027/1901-2276.61.2.29

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, *46*, 1417–1432.

Hambleton, R. (2001). The next generation of the ITC Test translation and adaptation guidelines. *European Journal of Psychological Assessment*, *17*, 164–172. doi:10.1027/1015-5759.17.3.164

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, *1*, 1–30.

Harbich, S., & Auer, S. (2005). The influence of hedonic quality on usability questionnaires. In M. F. Costabile, & F. Paternò (Eds.), *Lecture Notes in Computer Science*: *Vol. 3585* (pp. 1129-1133). doi: 10.1007/11555261_121

Hartson, R., & Pyla, P. (2012). *The UX Book : Process and Guidelines for Ensuring a Quality User Experience*. Waltham, MD: Morgan Kaufmann.

Hassenzahl, M. (2003). *The thing and I: Understanding the relationship between user and product*. In M. A. Blyth, A. F. Monk, K. Overbeeke, & P. C. Wright (Eds.), *Human-Computer Interaction Series: Vol. 3. Funology: From usability to enjoyment* (pp. 31-42). doi: 10.1007/1-4020-2967-5_4

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, *19*, 319-349. doi:10.1207/s15327051hci1904_2

Hassenzahl, M. (2007). AttrakDiff. Internet Resource http://www.attrakdiff.de.

Hassenzahl, M. (2008). User Experience (UX): Towards an experiential perspective on product quality. *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine* (pp. 11-15). New-York, NY: ACM. doi: 10.1145/1512714.1512717

Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*, *3*(1), 1-95. doi: 10.2200/S00261ED1V01Y201003HCI008

Hassenzahl, M. (2013). Experiences Before Things: a Primer for the (Yet) Unconvinced. *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2059-2068). New York, NY: ACM. doi:10.1145/2468356.2468724

Hassenzahl, M., Beu A., & Burmester, M. (2001). Engineering joy. *IEEE Software*, *18*(1), 70-76. doi: 10.1109/52.903170

Hassenzahl, M., Burmester, M., & Koller, F. (2003). *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*. In J. Ziegler, & G. Szwillus (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung* (pp. 187–196). Stuttgart, Germany: B.G. Teubner. doi: 10.1007/978-3-322-80058-9_19

Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human-Computer Interaction*, *25*(3), 235-260. doi: 10.1080/07370024.2010.500139

Hassenzahl, M., Platz, A., Burmester, M., Lehner, K. (2000). Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 201-208. New York, NY: ACM. doi: 10.1145/332040.332432

Hassenzahl, M., & Sandweg, N. (2004). From mental effort to perceived usability: transforming experiences into summary assessments. *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1283-1286). New York, NY: ACM Press. doi: 10.1145/985921.986044

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behavior & Information Technology*, *25*, 91-97. doi: 10.1080/01449290500330331

Herdman, M., Fox-Rushby, J., & Badia, X. (1997). 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, *6*, 237–247.

Hilton, A., & Skrutkowski, M. (2002). Translating instruments into other languages: development and testing processes. *Cancer Nursing*, *25*(1), 1-7.

Holm, J., & Lehtiniemi, A. (2011). A Virtual World Prototype for Interacting with a Music Collection. In A. Ozok & P. Zaphiris (Eds.), *Lecture Notes in Computer Science*: *Vol. 6778. Online Communities and Social Computing.* (pp. 326-335). doi: 10.1007/978-3-642-21796-8_35

International Test Commission (2010). *Guidelines for Translating and Adapting Tests.* [http://www.intestcom.org]

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

Isleifsdottir, J., & Larusdottir, M. (2008). Measuring the User Experience of a Task Oriented Software. In E. Law et al. (Eds.), *Proceedings of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement* (pp. 97-101), Reykjavik, Iceland.

Kirakowski, J., & Corbett, M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology, 24*, 210–212.

Klaassen, R., Akker op den, R., Lavrysen, T., & van Wissen, S. (2013). User preferences for multi-device context-aware feedback in a digital coaching system. *Journal of Multimodal User Interfaces*, *7*(3), 247-267. doi:10.1007/s12193-013-0125-0

Lallemand, C., Gronier, G., & Koenig, V. (in press). User Experience: a concept without consensus? Exploring practitioners' perspectives through an international survey, *Computers in Human Behavior*. doi:10.1016/j.chb.2014.10.048

Lankes, M., Bernhaupt, R., & Tscheligi, M. (2010). Evaluating User Experience Factors Using Experiments: Expressive Artificial Faces Embedded in Contexts. In R. Bernhaupt (ed.), *Human-Computer Interaction Series. Evaluating User Experience in Games*. London: Springer. doi:10.1007/978-1-84882-963-3_10.

Law, E. C., Van Schaik, P., & Roto, V. (2013). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies*, *72*, 6, 526-541. doi:10.1016/j.ijhcs.2013.09.006i

Law, E., Roto, V., Hassenzahl, M., Vermeeren, A. & Kort, J. (2009). Understanding, scoping and defining UX: a survey approach. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New-York, NY: ACM. doi: 10.1145/1518701.1518813

Leuteritz, J.-P., Widlroither, H., & Klüh, M. (2009). *Multi-level Validation of the ISOmetrics Questionnaire Based on Qualitative and Quantitative Data Obtained from a Conventional Usability Test*. In J.A. Jacko (Ed.), *Lecture Notes in Computer*

Science: *Vol. 5610. Human-Computer Interaction. New trends.* (pp. 304–313). doi: 10.1007/978-3-642-02574-7_34

Lund, A.M. (2001). Measuring Usability with the USE Questionnaire. *Usability & User Experience* (STC SIG Newsletter), *8*, 2. http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.

Maguire, M. (2001). Methods to support human-centred design. *International Journal of Human-Computer Studies*, *55* (4), 587-634.

Moshagen, M., & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, *68* (10), 689-709. doi:10.1016/j.ijhcs.2010.05.006

Nichols, A. L., & Webster, G. D. (2013). The single-item need to belong scale. *Personality and Individual Differences, 55*(2), 189-192. doi:10.1016/j.paid.2013.02.018

Nielsen, J. (1993). *Usability Engineering*. New York: AP Professional.

Norman, D., Miller, J., & Henderson, A. (1995). What You See, Some of What's in the Future, And How We Go About Doing It: HI at Apple Computer. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New-York, NY: ACM.

Nunnally, J.C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.

Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Osgood, C.E., Suci, C. J., & Tannenbaum, G. H. (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press.

Peedu, G. (2011). *Enhancing Public Service User Experience in Information Society*. (master's thesis). Tallinn University, Institute of Informatics: Estonia. Retrieved from: www.cs.tlu.ee/teemaderegister/get_file.php?id=108.

Pine, B.J., & Gilmore, J.H. (1998). Welcome to the Experience Economy: work is theatre & every business a stage. *Harvard Business Review*, July-August 1998.

Raita, E., & Oulasvirta, A. (2011). Too good to be bad: Favorable product expectations boost subjective usability ratings, *Interacting with Computers*, *23*, 363-371. doi: 10.1016/j.intcom.2011.04.002

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (2011). User Experience White Paper: Bringing clarity to the concept of user experience. *Result from Dagstuhl Seminar on Demarcating User Experience*, Sept. 15-18, 2010, Finland. http://www.allaboutux.org/uxwhitepaper

Russell, J. A., Weiss, A., & Mendelsohn, G. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*(3), 493–502.

Sauro, J., & Dumas J. (2009). Comparison of Three One-Question, Post-Task Usability Questionnaires. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 1599-1608. New-York, NY: ACM. doi :10.1145/1518701.1518946

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 693-727. doi: 10.1177/0539018405058216

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*, *89*, 325-339. doi: 10.1037//O022-3514.80.2.325

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

Tractinsky, N., Katz, A., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*(2), 127-145. doi:10.1016/S0953-5438(00)00031-X

Tuch, A., Trusell, R., & Hornbæk, K. (2013). Analyzing users' narratives to understand experience with interactive products. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2079-2088. Paris, France: ACM. doi:10.1145/2470654.2481285

Vallerand, R.J. (1989). Vers une méthodologie de validation transculturelle de questionnaires psychologiques : implications pour la recherche en langue française. *Psychologie Canadienne*, *30*(4), 662-689.

Van Boven, L., & Gilovich, T. (2003). To do or to have? That is the question. *Journal of Personality and Social Psychology*, *85*, 1193-1202. doi:10.1037/0022-3514.85.6.1193

Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89-99.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, *13*, 29–37.

Van de Vijver, F., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment. *European Review of Applied Psychology*, *47*, 263–279.

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54,* 119-135. doi: 10.1016/j.erap.2003.12.004

Van Schaik, P., & Ling, J. (2008). Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. *Interacting with computers, 20*, 419-432. doi: 10.1016/j.intcom.2008.03.001

Vermeeren, A, Law, E., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User Experience Evaluation Methods: Current State and Development Needs. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, New York, NY: ACM Press. doi :10.1145/1868914.1868973

Wanous, J. P., & Hudy, M. J. (2001). Single-Item Reliability: A Replication and Extension. *Organizational Research Methods*, *4*, 361-375. doi: 10.1177/109442810144003

Wechsung, I., Engelbrecht, K. –P., Schaffer, S., Seebode, J., Metze, F., & Möller, S. (2009). *Usability Evaluation of Multimodal Interfaces: Is the Whole the Sum of Its Parts?* In J.A. Jacko (Ed.), *Lecture Notes in Computer Science*: *Vol. 5611. Human-Computer Interaction. Novel Interaction Methods and Techniques* (pp. 113-119). doi: 10.1007/978-3-642-02577-8_13

Wechsung, I., & Naumann, A.B. (2008) *Evaluation Methods for Multimodal Systems: A Comparison of Standardized Usability Questionnaires*. In E. André et al. (Eds.), *Lecture Notes in Computer Science*: *Vol. 5078. Perception in Multimodal Dialog Systems* (pp. 276-284). doi: 10.1007/978-3-540-69369-7_32

Zwinderman, M., Leenheer, R., Shirzad, A., Chupriyanov, N., Veugen, G., Zhang, B., & Markopoulos, P. (2013). Using Video Prototypes for Evaluating Design Concepts with Users: A Comparison to Usability Testing. In P. Kotzé et al. (Eds.), *Lecture Notes in Computer Science*: *Vol. 8118* (pp. 774-781). doi: 10.1007/978-3-642-40480-1_55

THE CONTEXTUAL DIMENSION OF USER EXPERIENCE: DEVELOPMENT AND APPLICATION OF THE USER EXPERIENCE CONTEXT SCALE (UXCS)

C

Carine Lallemand, Vincent Koenig, & Guillaume Gronier

Submitted for journal publication

Keywords: user experience, context of use, scale development, UX assessment, UX Context Scale

ABSTRACT

The context of use, in a broad understanding, has been highlighted for a long time as being a key factor impacting User Experience (UX). However, current UX assessment tools and especially questionnaires are often focused only on hedonic and pragmatic qualities of the interactive system and rarely encompass an explicit investigation of the context. With the ever-growing trend for mobile products, this issue yet becomes critical.

Based on a review of relevant literature and a fine-grained categorization of contextual factors, the UX Context Scale (UXCS) has been developed to allow for a measure of context properties, as perceived by the user. The UXCS supports both a measure of objective contextual factors and a measure of perceived context properties. It has been built based on a thorough process following best practices for summated rating scale construction.

The UXCS was tested through an online study (N = 137 valid answers), which investigated the effects of contextual factors on perceived UX. A principal component analysis on perceived contextual factors reveals a 6-components structure encompassing *Physical Context*, *Social Context*, *Internal Context*, *Perceived Resources*, *Task Context*, and *Temporal Context*. Reliability of each subscale is high (ranging from .68 to .93) and further analyses confirm the relevance and validity of this tool for UX evaluation. Beyond aspects related to the construction of the scale, this study also provides valuable insights on the links between UX and context.

1. INTRODUCTION

User experience (UX) is commonly described in the literature as the holistic quality of the interaction between a user and an interactive system. Even if no clear consensus has been reached on the definition of UX, professionals agree on the three classical pillars influencing UX: the user, the system and the context (Roto, Law, Vermeeren, & Hoonhout, 2011). Following this common understanding, Hassenzahl and Tractinsky (2006) define UX as: "A consequence of a *user's internal state* (predispositions, expectations, needs, motivation, mood, etc.), the *characteristics of the designed system* (e.g., complexity, purpose, usability, functionality, etc.) and *the context* (or the environment) within which the interaction occurs (e.g., organizational/ social setting, meaningfulness of the activity, voluntariness of use)" (p. 95)

Classically considered as a crucial factor in the fields of Ergonomics and Human-Computer Interaction, the notion of context is almost always mentioned in UX frameworks or models. The "situatedness of action" (McCarthy & Wright, 2004) and the acknowledgment of a close interaction between action and situation during the interaction with technology contribute to highlight the central role played by the context. Numerous studies have evaluated interactive systems in specific contexts of use, thereby showing what challenges are set by these contextual factors for the interaction (Korhonen, Arrasvuori, & Väänänen-Vainio-Mattila, 2010; Oulasvirta, 2009).

Paradoxically, UX assessment tools, and especially questionnaires, tend to focus on the perceived qualities of a product or system without paying much attention to the explicit measure of both objective and perceived context (Hassenzahl, Burmester, & Koller, 2003). One may argue that this dimension is studied upstream of a project or with more qualitative methodologies. However, since online UX studies become a widespread tool, a quantitative measure of perceived context "quality" grows even more important. Understanding and assessing the influence of contextual factors in user experience is crucial for the design of interactive products.

This paper focuses first on the importance of the context as a main factor impacting User Experience. Theories on the notion of context and its underlying dimensions are then described and discussed. The second part of the paper presents the UX Context Scale (UXCS), its subscales and the rationale behind its design. Finally, we describe the validation of the UXCS through an online study involving 137 participants. The paper concludes with the implications of this study for UX evaluation.

1.1. USER EXPERIENCE IS CONTEXTUAL

According to the Oxford online dictionary (http://oxforddictionaries.com, retrieved on 2014-06-25), the context is generally defined as 'the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood and assessed'. An experience, whatever form it takes, is indeed inevitably embedded in a specific setting and should therefore be considered as "colored" by a specific context. User experiences, being specific categories of experiences "derived from

encountering systems" (Roto et al., 2011), are no exceptions to this rule. The results of a UX survey, originally conducted in 2008 (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009) and replicated in 2012 (Lallemand, Gronier, & Koenig, in press), confirm this observation by showing that UX professionals mainly agree on the fact that "*UX occurs in and is dependent on the context in which the artefact is experienced*".

From a theoretical point of view, UX is contextual per nature: the field of UX emerged out of several theories highlighting the essential role of the context. Situated action (Suchman, 1987), which focuses on the understanding of human acts in context, is one of them. Traditional User-Centred Design, another deep root of UX, also relies on the requirement to "*understand and specify the context of use*" (ISO 9241-210, 2010). The majority of UX models, whether theoretical or more pragmatic, include the context of use as one of the main factors impacting UX (Crilly, Moultrie, & Clarkson, 2004; Forlizzi & Ford, 2000; Hassenzahl & Tractinsky, 2006; ISO, 2010; Mahlke, 2007; Spool, 2005). The UX White Paper (2011), co-written by several UX specialists, states that "*UX may change when the context changes, even if the system does not change*" (p. 10). Numerous field studies confirm this statement (Korhonen et al., 2010; Wigelius & Väätäjä, 2009). Spool (2005) claims that UX Design needs to consider user, interface and context. Similarly, Hassenzahl and Tractinsky (2006) defined three categories of factors influencing UX: the user's internal state, the characteristics of the designed system and the context (or the environment) within which the interaction occurs.

The importance of contextual and situational factors for the quality of human-computer interactions is not something novel, it has been recognized for a while and has given birth to the notion of context-awareness. Context-aware computing is a field of research and practice introduced by Schilit, Adams and Want (1994) to designate the fact that devices may sense several parameters of the environment, use the situation and react accordingly to the situation. As context awareness has mainly been used for applications where the user's context is rapidly changing (e.g., handheld) (Dey & Abowd, 1999), user location was one of the first parameters of interest. Many others followed as the notion of context was conceptualized in HCI.

1.2. WHAT IS THE CONTEXT IN HCI?

Following Dourish's paper on the role of context (2004), it is essential to understand "*what we talk about when we talk about context*". The notion of context in HCI has indeed been defined in various ways. It should be noted that major contextual models or taxonomies were first built to support the specification of the use context for usability studies (Alonso-Rios, Vazquez-Garcia, Mosqueira-Rey & Moret-Bonillo, 2010; Bevan & Macleod, 1994; Maguire, 2001; Maissel, Dillon, Maguire, Rengger, & Sweeney, 1991), and subsequently extended beyond the scope of usability to that of user experience.

The renowned ISO standard 9241-11 (ISO, 1998) identifies *users*, *tasks*, *equipment* and *environment* as the first-level attributes, further broken down into subattributes and categories given as a set of examples. *Users* encompass *user types*, *skills*, *knowledge* and *personal attributes*. Equipment is subdivided into *basic description* and *specification*. *Environment* includes three categories, namely *organizational*, *technical* and *physical* environment. Maguire (2001) provides a description of the main contextual aspects to be considered in a context of use analysis. His first-level components include *user goals and characteristics*, *tasks*, *technical environment*, *physical environment* and finally *social or organizational environment*.

More recently, Alonso-Rios et al. (2010) provided a context-of-use taxonomy for usability studies. This taxonomy aimed at lessening the tendency of usability studies to overlook characteristics of the context, as the authors believed this phenomenon to be caused by the lack of a clear and comprehensive taxonomy including precise definitions of context of use components. Table 1 summarizes context-of-use attributes defined by Alonso-Rios et al. (2010) where the classical triad *user*, *task* and *environment* constitutes the hierarchy's first-level.

Table 1. Context-of-Use taxonomy attributes by Alonso-Rios et al. (2010).

User	Role, Experience, Education, Attitude to the system, Physical characteristics, Cognitive characteristics
Task	Choice in system use, Complexity, Temporal characteristic, Demands, Workflow controllability, Safety, Criticality
Environment	Physical environment, Social environment, Technical environment

Classifications of contextual dimensions were also produced in the field of context-awareness. Dey and Abowd (1999) proposed the distinction between four contextual dimensions, namely *location*, *identity*, *activity* and *time*. Later, Kaltz, Ziegler and Lohmann (2005) differentiated three categories of context: *user and role*, *process and task* and finally *time and device*. However, following Dey's (2001) definition of the context as "*any information that can be used to characterize the situation of an entity,*" more thorough and multidisciplinary models of the context of use needed to be explored in order to contribute to the understanding of user experience. Bradley and Dunlop (2005) proposed a multidisciplinary model of context of use integrating HCI and context-aware computing perspectives. These authors define context of use as "*anything that influences the process in which focal user actions are undertaken*" and identified the following context dimensions:

- *Task context*: people and objects surrounding the task, which offer advantages ("resources") and drawbacks ("constraints") for the achievement of the task.
- *Physical context*: includes the environmental location, nearby physical elements and their attributes, but also weather and light conditions.
- *Social context*: influences of surrounding people in terms of relationship, dialogue, presence and behaviour.
- *Temporal context*: influences of past experiences in terms of expectations, and explanations of the present. It also includes higher-level temporal context (e.g., current time, date or season).

- *Application context*: concerns the information flow between the user and the device (e.g., representation of the application's state and functioning).
- *Cognitive context*: the user's cognitive abilities, habits and attitudes.

Furthermore, Bradley and Dunlop (2005) distinguished these dimensions according to two properties: *meaningful* and *incidental. Meaningful context* is related to the achievement of the user's higher-level goals, and is being selected by the user. *Incidental context* describes, "incidental occurrences in the contextual world … e.g., bumping into a friend", which will exert an influence on the user. Hence, this adds a dynamic dimension to the model (e.g., incidental elements of the context can become meaningful). Finally, Bradley and Dunlop considered user activity as a process in which cognitive goals are "*continually shaped by people's perception of the meaningful and incidental contextual worlds*" (p. 428), while at the same time user actions to fulfil these goals will contribute to shaping future context.

In a recent literature review, Jumisko-Pyykkö and Vainio (2010) summarised the most cited elements of context of use, as described in 109 articles from five major journals and one main conference in the field of HCI. They found six most studied components of context, presented in the following according to the number of articles referring to each component:

- *Social context (66,1% - 72/109 papers)*: influence of others, present physically or in a technology-mediated fashion. Social context also includes social norms, values, and attitudes.
- *Physical context (61.5% - 67/109 papers)*: spatial location, functional space, and place. Includes where the user is located, elements (artefacts) of the near environment he could interact with and general functions and attributes of his surroundings.
- *Technical and informational contexts (36.7% - 40/109 papers)*: technical aspects such as available services and applications, interoperability, type of device, and mixed reality.
- *Temporal context (35.8% - 39/109 papers)*: duration of the interaction, time of day and date (in relation to user activity patterns), level of time pressure, and synchronism of the task.
- *Task context (31.2% - 34/109 papers):* task interruptions (attentional shift away from the task), number of different competing tasks (multitasking) as well as task's category (goal-oriented or action-oriented).
- *Transitions (20,2% - 22/109 papers):* Some cases (e.g., a mobile context of use) also involve a changing environment, which results in "*transitions"* between the dimensions of context.

Furthermore, Jumisko-Pyykkö and Vainio (2010) found that these context elements also have properties, which need to be examined in order to render the dynamic nature of context. Four general properties are highlighted:

1) *Level of magnitude*: depending on the scale that is used (macro vs. micro), the focus changes within a dimension of context (e.g., "in the physical

context, artefacts that are near to the user represent the micro scale whereas functional space is an example of the macro scale").

2) *Level of dynamism*: distinction between static context (e.g., a daily task executed at the same time every day, with constant relation to time) versus dynamic context (same task executed under different contexts).
3) *Type of patterns*: is there a regular context occurrence (rhythmic) or on the contrary, is it occurring in an unpredictable manner (random)?
4) *Typical combinations*: which different context elements are usually present simultaneously?

It should be noted that Jumisko-Pyykkö and Vainio (2010) do not detail the internal context of the user as a proper context dimension. This tendency to isolate the user as a distinct entity might result from the aim of reducing the overload upon the term *context* as suggested by the ISO standard 13407 (1999), updated as ISO 9241-210 (2010).

Knowing which contextual dimensions are involved in human-computer interactions is a first step toward the understanding of how context impacts UX. While there are some differences in the number of dimensions or labelling, various contextual models and taxonomies globally agree on the main dimensions to take into account when studying the context in HCI. Social context, physical context, technical context, temporal context, internal context, and finally task context appear to be the most relevant factors impacting the interaction. Transitions might also occur when the environment is changing during the course of the interaction.

1.3. THE CONTEXT IN UX QUESTIONNAIRES

Amongst available UX assessment tools, online studies and self-administered questionnaires appear to be a good vector for contextual studies by allowing users to answer them during or just after use in their ecological setting. However, remote studies or self-declared measures will require specific contextual-aspects measurement tools.

Existing UX assessment self-administered questionnaires are predominantly focused on pragmatic and hedonic qualities of interactive systems without actually including the context as a variable to be explicitly measured.
The AttrakDiff questionnaire (Hassenzahl et al., 2003) or the User Experience Questionnaire (Laugwitz, Held, & Schrepp, 2008) are two examples of this kind of "product-centred" UX measurement tools. By product-centred, we do not claim that the underlying models of these tools do not take context into account since it is assumed that the evaluation after exposure to a product always implies and includes a certain context. However, product and context are not studied separately and the resulting perception is an evaluation of the product in context.
Other questionnaires may be used as a complement to these measures. As an example, some authors (Tuch, Trusell, & Hornbæk, 2013) report the use of three generic questions about the context extracted from the Geneva Appraisal Questionnaire (Scherer, 2001): they concern when, where, and with whom a reported user experience took place. Partala and Kallinen (2012) used a self-made 10 statements questionnaire studying the context of

the user experience. It was specifically developed for their experiment based on the context framework by Jumisko-Pyykkö and Vainio (2010) and included the most cited contextual dimensions (temporal, physical, social, task and technical - contexts).
Regarding more specific aspects of the context, numerous measurement tools, primarily developed by psychologists, are hence focused on studying what we may call the internal context of the user, that is affects (Russell, Weiss, & Mendelsohn, 1989; Watson, Tellegen, & Clark, 1988), emotions (Bradley & Lang, 1994; Izard, 1977; Scherer, 2001), or intrinsic motivation (Ryan, 1982). Amongst the questionnaires commonly used to study UX, only a few address the notion of social context and those are predominantly used in the context of games (Nacke et al., 2009). Finally, to the best extent of our knowledge, no existing UX questionnaire takes explicitly into account a measure of the physical context of the interaction.

2. DEVELOPMENT OF THE UX CONTEXT SCALE

The development of a UX Context Scale (UXCS) aims at providing a holistic measurement of contextual factors for UX assessment. The creation of this tool contributes to filling the gaps in UX evaluation questionnaires, which were lacking a self-declared measure of the context quality. The measurement of both objective and perceived aspects of the context is especially relevant for controlling the influence of the context or simply knowing more about contextual aspects involved in a situation of interaction.

By trying to build a measurement tool to assess contextual factors, we also aimed at easing and promoting the integration of the context as a core dimension of every UX evaluation. In academia, the UXCS might be used for studying contextual factors impacting UX by providing insights into the interrelationships between contextual dimensions and UX. Research could be transferred into practice through recommendations for UX practitioners. In industry, the UXCS scale might be used both for the design of an interactive system or for its evaluation. In the first case, the focus would be on the process (formative evaluation) allowing product designers to better take into account contextual factors when designing a product. In the second case, the focus would be on the product itself (summative evaluation) and the UXCS would collect valuable information on the influence of context on the perceived quality of interaction. Both applications might be especially relevant for highly context-dependent products, for example mobile devices (Korhonen et al., 2010).

Based on the theoretical findings on the nature and definition of context reviewed in the first section of this paper (Jumisko-Pyykkö & Vainio, 2010), we designed the UXCS to support both an objective measure of contextual factors and a subjective measure of perceived context. Following the best practices on summated rating scale construction (Spector, 1992), we first conducted a literature review to define the dimensions and subdimensions of context relevant for the evaluation of UX. Second we designed a first experimental version of the scale under the format of semantic differential items. We pilot tested the UXCS on 10 participants to ensure the understandability of the items and the format. We also asked two UX

experts to review the scale for face validity. After slight refinements of the scale, we administered the scale to a sample of 147 respondents through an online study. The results of this validation study are described in the Method section.

2.1. CONTEXTUAL DIMENSIONS AND SUBDIMENSIONS

Six main contextual dimensions were identified within the literature as being of primary importance and relevant to compose the UXCS. A fine-grained categorization of contextual dimensions was needed to understand which contextual factors affect user experience and how this occurrs. Each subscale is composed of both objective and perceived contextual aspects. The number of items in each subscale is variable and depends both on the importance of each dimension as described in the literature and on the number of subdimensions included in each dimension.

We submitted the initial pool of item to five HCI experts and therefore checked the scale's face validity by asking them to review each item separately and also to assess if the scale and subscales appeared to be a good measure of the interaction context. This expert review resulted in a slight decrease in the number of items, as some items did not reach consensus amongst the experts. We also pilot tested the scale on 10 users, resulting in changes in the wordings of several items. Users encountered no issue in understanding the item format.

The experimental version of the UX Context Scale was composed of 41 items (Table 2), divided into 6 subscales, as derived from the literature review:

- **Physical Context** (PHYS-CTX) (7 items): this subscale is focused on the conditions of use and the properties of the physical environment (e.g., in terms of noise, temperature and lighting). From an objective point of view, it also gathers data on the level of familiarity related to the location and the type of usage in order to know if the user's context is stable or changing rapidly (i.e., level of dynamism).
- **Social Context** (SOC-CTX) (5 items): this subscale refers to the interactions between the users and other people. Their mere presence and characteristics of individuals will influence users' perceptions and behaviours. From an objective perspective, this subscale informs whether the interaction took place in a public or private place, with or without social interactions. The subjective perspective is mainly focused on the feeling of relatedness, the fulfilment of which is an essential requirement for optimal UX (Hassenzahl, 2010).
- **Internal Context** (INT-CTX) (20 items): this subscale is focused on the internal context of the user, especially his or her mood, motivation and interest in the system. As this internal context has been shown to be highly dynamic depending on the interaction phase, this subscale has been divided into 3 parts, each of them assessing a specific period of interaction. The first one asks users about their feelings and opinion before using the system; the second

assesses the stage during the use, and finally the last part focuses on their opinion after usage. The internal context subscale includes more items than the other subscales composing the UXCS as the internal context of use (i.e., the user) has been highlighted as a major factor impacting UX. In order to cover all aspects related to the internal-context, it was necessary to distinguish different assessment phases and to include several items regarding knowledge, mood, feelings, resources, etc. Moreover, this subscale was also intended to allow for independent assessment of this aspect only.

- **Technical Context** (TEC-CTX) (1 item): despite the importance of the technical context involved in the interaction (i.e., the quality of the technical device and environment supporting the interaction), technical-related aspects are very difficult to evaluate from the users' point of view, as users are generally not IT experts. We therefore decided to only ask final users whether they have encountered any technical issues with the system, without asking them to define precisely what type of issue they had to cope with. This question also allowed our scale to remain at a generic level, compatible with the assessment of various types of systems or products.
- **Task Context** (TASK-CTX) (4 items): this subscale is not focused on tasks per se - as this would have implied to considerer quite as many tasks as users - but rather on the focus devoted to the task. Multitasking (competing tasks) and task interruptions (attentional shift away from the task) are of particular interest here.
- **Temporal Context** (TEMP-CTX) (4 items): this subscale refers to users' perceived temporal aspects of the interaction, illustrated by the level of time pressure or the adequacy of the moment when it took place. Duration of last interaction and regularity of use are also explored as objective contextual factors.

Social, physical, task, temporal and technical context dimensions are part of the core components as identified by Jumisko-Pyykkö and Vainio (2010) in their extensive literature review. Regarding the internal context, it has been pinpointed by Korhonen et al. (2010) as missing in most context studies since it describes the users' characteristics (Jumisko-Pyykkö & Vainio, 2010; Roto, 2006). However, following Dey (2001), users were considered in our study as "entities" in the same way as location or artefacts, and the assessment of their internal state was therefore included as the internal context subscale in the UXCS. Please note however that the internal context subscale might be used also independently in the case one wants to assess this factor only.

2.2. OBJECTIVE AND SUBJECTIVE CONTEXTUAL FACTORS

The UXCS was designed in order to encompass both objective and subjective (i.e., perceived) parameters of the user's context. Each aforementioned subscale therefore includes an assessment of objective contextual factors and perceived contextual factors. On the one hand, the objective items have no affective valence, neither positive nor negative; they

are meant to describe factual information (e.g., being in a public or private place). On the other hand, subjective items are tinted by a value judgment (e.g., being in an unpleasant vs. pleasant location), one extreme of the scale being negative, the opposite being positive.

Objective items will be used to learn more about the situation in which the user is placed and may also be used from a research point of view to assess the impact of specific objective situations on UX. They will therefore be used as informative variables, helpful to interpret users' opinion and feelings toward a system. Subjective items will be averaged per subscale to constitute the evaluation of the perceived context qualities. They will therefore constitute core factors to study. Principal component analyses have been performed on subjective items to check for the validity of the UXCS structure while reliability has been assessed through separate analyses.

2.3. RESPONSE CATEGORIES AND RATING SCALES

When designing a measurement scale, the number and type of response categories has to be carefully defined, as it may influence the psychometrical data quality. As stated by Schaffer and Presser (2003, p. 78), "*the choice of the number of categories represents a compromise between the increasing discrimination potentially available with more categories and the limited capacity of respondents to make finer distinctions reliably and in similar ways*". Standard advice had been to use five to nine categories (Schaeffer & Presser, 2003). The UXCS has therefore been designed using 7-points bipolar anchors (also called "semantic differentials" by Osgood, Suci, & Tannenbaum, 1957). This scale format was consistently used throughout the whole questionnaire. This choice was consciously made in order to allow for easy comparison between contextual factors and UX-related aspects (as it follows the form of the AttrakDiff scale, Hassenzahl et al., 2003). Moreover, from the user's point of view, this format seems easy to understand as it has been used within hundreds of UX studies with a few reported issues only regarding response options.

The overall structure and internal consistency of our UX Context Scale and subscales has been tested through an online study aimed at assessing users' experiences with the online professional network LinkedIn ®.

Following are pairs of words/sentences that describe the context of interaction. Each pair represents a contrast of ideas about the context in which the interaction occurred. Your mission is to rate each item on a 7-points scale. This is not an evaluation of the quality of the system but an attempt to assess your environment and feelings.

Physical context		
While using the system:		
PHYS01	In an unfamiliar place	In a familiar place
PHYS02	**In an unpleasant location**	**In a pleasant location**
PHYS03	In a noisy place	In a quiet place
PHYS04	In a moving/vibrating environment	In a steady environment
PHYS05	**Unpleased with the temperature**	**Pleased with the temperature**
PHYS06	**Unpleased with the lighting**	**Pleased with the lighting**
PHYS07	Moving (mobile usage)	Remaining still
Social context		
While using the system:		
SOC01	In a public space	In a private space
SOC02	Not interacting with people	Interacting with people
SOC03	**Feeling alone**	**Feeling related to other people**
SOC04	**Feeling unsupported**	**Feeling supported**
SOC05	***Bothered by others**	**Unbothered by others***
Internal context		
Before using the system:		
INT_BF01	***I was in a bad mood**	**I was in a good mood***
INT_BF02	I had no expectations about the system	I had high expectations about the system
INT_BF03	I had no specific tasks to achieve	I had specific tasks to achieve
INT_BF04	I had no previous experience with the system	I already had experience in using the system
INT_BF05	I had no information about it	I had information about it
INT_BF06	**I had a bad opinion about it**	**I had a good opinion about it**
INT_BF07	**I felt pressed to use it**	**I felt free to use it**
While using the system:		
INT_WH01	**Demotivated**	**Motivated**
INT_WH02	**Not interested**	**Interested**
INT_WH03	**In a bad mood**	**In a good mood**
INT_WH04	**Unsatisfied**	**Satisfied**
INT_WH05	**I had insufficient skills to use it**	**I had sufficient skills to use it**
INT_WH06	**I did not have enough time to spend on it**	**I had enough time to spend on it**
INT_WH07	**I had the need to be helped or trained**	**I had no need to be helped or trained**
INT_WH08	**Powerless over my environment**	**In control of my environment**
After using the system:		
INT_AF01	**I am in a bad mood**	**I am in a good mood**
INT_AF02	**I feel unsatisfied**	**I feel satisfied**
INT_AF03	**I am not willing to use it again**	**I am willing to use it again**
INT_AF04	**I feel repelled by the system**	**I feel attracted by the system**
INT_AF05	***My expectations are not satisfied**	**My expectations are satisfied***
Technical context		
While using the system:		
TEC01	Technical problems were encountered	No technical problems were encountered
Task context		
While using the system:		
TASK01	**I was doing several things simultaneously**	**I was focusing on the task**
TASK02	**I was often interrupted**	**I was never interrupted**
TASK03	**I focused on the product**	**I focused on attaining my goals**
TASK04	***I felt serious**	**I felt playful***
Temporal context		
While using the system:		
TEMP01	I used the system only once	I am using the system regularly
TEMP02	I spent a short time using the system	I spent a long time using the system
TEMP03	**I was interacting at an uncomfortable pace**	**I was interacting at a comfortable pace**
TEMP04	**It was not the right moment to use the system**	**It was the right moment to use the system**

Table 2. Experimental version of the UX Context Scale

Items in bold font are related to perceived contextual factors whereas items in regular font are related to objective contextual factors; items marked with a * were removed from the final version of the scale, see section 4.2 for more details.

3. METHOD

3.1. PARTICIPANTS AND PROCEDURE

The study was broadcast online on the professional network LinkedIn from August, 2013 to October, 2013 and therefore involves a non-probabilistic sampling. The choice of an online study was made in order to assess the interaction context in an ecological setting, and to collect data on a wide variety of interaction contexts. We explained the participants that the main objective of the study was the evaluation of users' experience with LinkedIn ® (https://www.linkedin.com/). The context was not mentioned within the introductory part of the online survey. General instructions were presented as follows: "We welcome your participation in the evaluation of your experience of LinkedIn. This survey is part of a research project focused on the User Experience of interactive systems. Thank you for responding as seriously as possible to the questionnaire, which will take you approximately 10 minutes. Your answers will be recorded and processed anonymously. Thank you for your interest and participation!"

147 complete answers were collected. IP addresses were checked to avoid participants taking part several times. Ten participants having declared a score of English proficiency inferior to 5 out of 7 were filtered out of the study as we considered that a precise understanding of each item was required for this validation study. The remaining 137 valid answers were quite balanced regarding the criteria of gender, with respectively 65 men and 72 women. Average age of participants was 38.9 years. Almost all participants declared feeling at ease with technology (M = 6.53 out of 7; 97.8% having a score higher than 5 out of 7). Thirty-two countries were represented, with USA (34%) and France as the most representative (15%).

The questionnaire consisted of three parts: evaluation of the system using the abridged AttrakDiff scale, assessment of objective and subjective contextual factors using the UXCS, and sociodemographic questions. All materials were in English.

3.1.1. USER EXPERIENCE ASSESSMENT: SYSTEM PERCEPTION AND EVALUATION

First, we asked participants to evaluate pragmatic and hedonic qualities of their interaction with the system. For doing so, we relied on an existing UX questionnaire: the AttrakDiff scale. The AttrakDiff questionnaire has been developed by Hassenzahl et al. (2003) to measure both pragmatic and hedonic qualities of an interactive product. The measurement relies on the format of semantic differentials: the ratings are therefore made on bipolar seven-point anchors (word-pairs are presented to the user, each word in a pair representing the extreme opposite of its counterparts). Evaluated system's qualities are Pragmatic Qualities (ATD_PRAG), Hedonic Qualities (ATD_HEDO) and finally Attractivity (ATD_ATT). In the present study, we chose to use the abridged version of the instrument in order to keep a reasonable total number of items. The abridged AttrakDiff is composed of

10 items (instead of 28 for the unabridged version). However, one item (*cheap-premium*) from this abriged AttrakDiff was proven to be unreliable (Hassenzahl et al., 2010) and has therefore not been included in our study.

We specifically chose this UX questionnaire as it relies on a sound theoretical model (Hassenzahl, Platz, Burmester & Lehner 2000; Hassenzahl 2002). Moreover, the multidimensional structure of the AttrakDiff allows for a distinction between several UX dimensions with regards to the impact of the interaction context. According to Hassenzahl et al. (2003), a product's perceived quality is based on two main attributes: its pragmatic quality and its hedonic quality. Pragmatic quality is instrumental; it is mainly related to how useful and usable the product is in supporting the realization of specific tasks (called "do-goals"). System's clarity, structure, or predictability are illustrations of pragmatic attributes. Hedonic quality is non-instrumental and refers to the Self. It is strongly linked to the user and based on the perceived product's potential to provide pleasure and to fulfil basic human needs, called "be-goals". A system's ability to stimulate the user, to support a feeling of control or relatedness are some examples of hedonic attributes. The combination between perceived pragmatic quality and perceived hedonic quality lead to a global assessment of a product's attractivity, which gives rise to emotional (e.g., joy) and behavioural consequences (e.g., enhanced use). This evaluation of LinkedIn UX plays a central role in our study as it will be used to study the impact of the interaction context on the perceived pragmatic and hedonic qualities of the system.

3.1.2. ASSESSMENT OF OBJECTIVE AND SUBJECTIVE CONTEXTUAL FACTORS

To gather objective and subjective contextual factors, users were presented with the UX Context Scale along with these instructions: "Following are pairs of words/sentences that describe the context of interaction. Each pair represents contrasted ideas about the context in which the interaction occurred. Your mission is to rate each item on a 7-points scale. This is not an evaluation of the quality of the system but an attempt to assess your environment and feelings the last time you used the system".
Evaluated contextual dimensions are Physical Context, Social Context, Technical Context, Task Context, and Temporal Context. The experimental version of the UXCS is composed of 13 items related to objective contextual factors and 28 items related to subjective contextual factors (see Section 2).

3.1.3. SOCIODEMOGRAPHIC VARIABLES

Age, gender, country of residence, mother tongue and familiarity with technology (7-point Likert scale from "I feel not at all at ease with technology" to "I feel completely at ease with technology") were also collected. Finally, participants were asked to evaluate their level of proficiency in English on a 7-point Likert scale from "not proficient at all" to "fully proficient". This question aimed at ensuring an optimal understanding of all items, and eventually filtering out participants being not proficient enough in English.

3.2. DATA ANALYSIS

Data was analysed using the software SPSS Statistics 21. No answer of the UXCS needed to be reverse-coded, as we had chosen to orient responses on a homogeneous continuum from negative to positive, left to right. Item codes will be used for the description of the results (see Table 2 for all item codes). There was no missing data.
In order to compare most satisfying experiences vs. most unsatisfying experiences as it has been done in previous studies (Partala & Kallinen, 2012), participants were divided into 2 groups regarding their global score on the AttrakDiff scale. Participants with an AttrakDiff score inferior or equal to 4 out of 7 were classified into the "unsatisfying UX" group (n = 28) while participants with an AttrakDiff score superior to 4 out of 7 were classified as "satisfying UX" (n = 109).

4. RESULTS

4.1. VALIDATION OF THE UX CONTEXT SCALE

Univariate statistics were run to examine the means and standard deviations of each item as well as to check for possible outliers or entry errors. No outliers or entry errors were found. A principal component factor analysis was run on 28 items initially composing the perceived aspects of the UX Context Scale. The suitability of PCA was assessed prior to analysis. Inspection of the correlation matrix showed that all variables had at least one correlation coefficient greater than 0.3. However, before further analysing the PCA, the correlation matrix allowed detecting very high correlations between some items, therefore suggesting a redundancy. As items INT_AF02 (*I feel unsatisfied/I feel satisfied*) and INT_AF05 (*my expectations are not satisfied/my expectations are satisfied*) (r = .81, p < .001) were very close from a semantic viewpoint and seem to be measuring the same feeling, we decided to delete one of those two items from the scale. The wording of INT_AF05, involving the notion of expectations, was the most confusing (especially in the case a user would not have any expectation about the system) and it was therefore decided to run again the PCA without INT_AF05. The overall Kaiser-Meyer-Olkin (KMO) measure was 0.84, which is 'meritorious' according to Kaiser's (1974) classification of measure values and higher to the .6 value suggested by Tabachnik and Fidell (2001). Bartlett's Test of Sphericity was statistically significant (p < .001), indicating that the data was likely factorizable.

Principal component analysis provided an initial number of six possible factors (based on components with eigenvalues greater than 1), explaining 63.8% of variance in the data. Three items (SOC05, INT_BF01, and TASK04), not loading properly on any of the six factors identified, were deleted from the scale. The final data set was eventually composed of 24 pairs of items. Component loadings and communalities of the rotated solution are presented in Table 3. This six-factor solution explains 67.4 % of variance in the data. A Varimax orthogonal rotation was employed to aid interpretability.

The first factor, related to the *Internal context*, explains 32.7% of the observed variance in the data. The second factor, related to the *Physical context*, explains 12.1% of variance, the remaining four factors explaining between 6.9 and 4.6% of observed variance each. The rotated component matrix shows that almost all items have high loadings on their respective factor and low loadings on all other factors. Exceptions are items INT_BF07 (*I felt pressed to use it/I felt free to use it*) and INT_WH06 (*I did not have enough time to spend on it/I had enough time to spend on it*), which overlap on several dimensions. Except for the *internal context*, which appears to be split up into two different constructs, this six-factors solution reflects a factor structure that adequately relates to our empirical examination of the interrelationships among contextual-related dimensions. Identified subsets share sufficient variation to justify their existence as factors measuring the perceived context of human-computer interactions.

Table 3. Rotated structure matrix for PCA with Varimax rotation.

Scale / Item	Rotated Component Coefficients						Communalities
	1	2	3	4	5	6	
Internal Context							
INT_BF06	**.711**						.581
INT_BF07	**.451**	.360			.325		.525
INT_WH01	**.758**						.682
INT_WH02	**.760**						.645
INT_WH03	**.796**						.732
INT_WH04	**.845**						.773
INT_AF01	**.752**						.690
INT_AF02	**.831**						.739
INT_AF03	**.712**						.596
INT_AF04	**.795**						.707
Physical Context							
PHYS02		**.644**					.550
PHYS05		**.775**					.660
PHYS06		**.880**					.812
Perceived Resources							
INT_WH05	.344		**.765**				.733
INT_WH06			**.532**		.321		.414
INT_WH07			**.756**				.633
INT_WH08			**.649**				.648
Task Context							
TASK01				**.611**			.528
TASK02				**.758**			.755
TASK03				**.832**			.718
Temporal Context							
TEMP03					**.756**		.687
TEMP04					**.804**		.784
Social Context							
SOC03	.301					**.828**	.810
SOC04						**.810**	.771
EigenValue	7.85	2.90	1.67	1.45	1.21	1.09	
% of variance explained	32.7	12.1	7	6.1	5.1	4.5	

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Component loadings <.30 are suppressed. Major loadings for each item are bolded.

As shown in Table 3, four items composing the initial *Internal context* subscale appear to load on a separate subscale. We will call this subscale *Perceived resources*, as all four items are dealing with the user's perceived resources in terms of skills, time, need for training or control over the environment. Differentiating *Perceived resources* from the *Internal context* makes sense if one considers that resources might not only relate to the user's mental state but could also be attributed to an external causality.

Reliability of the subscales has been tested using Cronbach's alpha (Cronbach, 1951) (Table 4). All subscales showed high internal consistency, with Cronbach's alpha ranging from .68 (*Perceived Resources*) to .93 (*Internal Context*). All subscale items are worth retaining as: (1) no single deletion would cause a substantial increase in Cronbach's alpha and (2) all items correlate with their total scale to a high degree. The only exception to this is item INT_WH06, which correlates at a lower degree only (r = .33). Due to its theoretical importance highlighted by the literature review, we choose to retain this item. Overall, the UXCS (including 24 items divided into 6 contextual dimensions) had a Cronbach's alpha = .87. Based on this, we computed a component-based score for each perceived contextual subscale along with a global score for the perceived quality of the context.

Table 4. UXCS subscales: Descriptive Statistics and Reliability Analysis

Perceived Context Subscale	N Valid	Nb of items	Min.	Max.	Mean	SD	Cronbach's alpha
PHYSICAL CONTEXT (PHYS_CTX) PHYS02 / PHYS05 / PHYS06	137	3	3	7	5.88	0.93	.75
SOCIAL CONTEXT (SOC_CTX) SOC03 / SOC04	137	2	1	7	4.53	1.25	.75
INTERNAL CONTEXT (INT_CTX) INT_BF06 / INT_BF07 / INT_WH01 / INT_WH02 / INT_WH03 / INT_WH04 / INT_AF01 / INT_AF02 / INT_AF03 / INT_AF04	137	10	2.80	7	5.34	1.03	.93
PERCEIVED RESOURCES (RES_CTX) INT_WH05 / INT_WH06 / INT_WH07 / INT_WH08	137	4	1.25	7	5.70	1.02	.68
TASK CONTEXT (TASK_CTX) TASK01 / TASK02 / TASK03	137	3	2	7	4.91	1.34	.69
TEMPORAL CONTEXT (TEMP_CTX) TEMP03 / TEMP04	137	2	1	7	5.60	1.06	.74
USER EXPERIENCE CONTEXT SCALE (UXCS_TOTAL)	**137**	**24**	3.43	7	5.33	0.65	**.87**

The possible range of the UXCS is from 24 to 168. In our sample, the minimum score is 89 and the maximum score is 168, with a median score of 130.

4.2. PERCEIVED CONTEXTUAL DIMENSIONS

Participants' ratings using 7-points scales indicated that perceived contextual dimensions were on average positively evaluated (UXCS_TOTAL M = 5.33, SD = 0.65) (Table 4). The physical context and perceived resources subscales were assessed as the most positive contextual dimensions (M = 5.88 and 5.70, respectively), whereas the social context

and task context obtained the lowest ratings (M = 4.53 and 4.91, respectively).
The effects of demographic variables on the perception of contextual dimensions are weak. Independent samples t-tests show no significant gender differences in the evaluation of perceived contextual dimensions. A slight negative correlation was found between age and perceived resources (r = -.24, p = .004): the older participants are, the lower they assessed their perceived resources. Finally, a low positive correlation was found between perceived technology literacy (i.e. how comfortable participants feel with technology in general) and the overall UXCS (r = .21, p = .015). The more the participants report feeling at ease with technology, the better they evaluated the perceived context of the interaction.

Bivariate correlations analyses were performed in order to study the degree of relationship between perceived contextual subscales (Table 5). The physical context is significantly and positively correlated to all other subscales, except the social context. Main relationships are found between physical and temporal context (r = .31, p < .001), task context (r = .31, p < .001) and internal context (r = .28, p = .001). Feeling pleased by the physical surrounding environment is therefore linked to every other contextual dimension but not to the feeling of relatedness and social support. The social context is mainly linked to the internal context (r = .45, p < .001). As the items composing the perceived social context subscale are dealing with the feeling of relatedness and support from other people, one might argue that the fulfilment of this primary psychological need for relatedness has a great impact on the internal context of users (Hassenzahl, Diefenbach, & Göritz, 2010; Sheldon et al., 2001). Except for the task context, the internal context is also correlated to all other subscales, especially to the perceived resources (r = .46, p < .001) and temporal context (r = .38, p < .001). Finally, perceived resources are mainly correlated to the internal context (r = .46, p < .001) and the temporal context (r = .34, p < .001), but not to the task context.

Table 5. Bivariate correlations between UXCS perceived contextual subscales

Perceived Context Subscale		PHYS	SOC	RES.	INT	TASK	TEMP
PHYS	Pearson Correlation	1	,128	,194*	,279**	,306**	,306**
	Sig. (2-tailed)		,137	,023	,001	,000	,000
SOC	Pearson Correlation	,128	1	,189*	,448**	-,013	,189*
	Sig. (2-tailed)	,137		,027	,000	,880	,027
RESOURCES	Pearson Correlation	,194*	,189*	1	,462**	,021	,345**
	Sig. (2-tailed)	,023	,027		,000	,810	,000
INT	Pearson Correlation	,279**	,448**	,462**	1	-,086	,379**
	Sig. (2-tailed)	,001	,000	,000		,318	,000
TASK	Pearson Correlation	,306**	-,013	,021	-,086	1	,286**
	Sig. (2-tailed)	,000	,880	,810	,318		,001
TEMP	Pearson Correlation	,306**	,189*	,345**	,379**	,286**	1
	Sig. (2-tailed)	,000	,027	,000	,000	,001	

4.3. OBJECTIVE CONTEXTUAL DIMENSIONS

As mentioned in Section 2.2, several items from the UXCS aim at informing about the objective contextual situation in which respondents are placed during the interaction with a system. These items were not included in the PCA, as they do not have any affective valence. Objective items are used to

learn more about the situation in which the user is placed and may also be used to assess the impact of specific objective situations on UX. They will therefore be used as informative variables, helpful to interpret users' opinions and feelings toward a system.

The results inform us that, on average, participants were in a familiar place while using LinkedIn (*PHYS01*: $M = 6.29$, $SD = 1.13$). This place was quiet (*PHYS03*: $M = 5.61$, $SD = 1.3$) and the respondents remained still during their interaction (*PHYS07*: $M = 6.41$, $SD = 1.11$). They described their environment as a private space (*SOC01*: $M = 5.66$; $SD = 1.84$) and considered on average that they were not interacting with people (*SOC02*: $M = 2.82$, $SD = 2.02$). 63,5% of respondents declared that they were not interacting with people, while 27% declare that they were interacting with people and 9.5% gave a neutral answer.
Regarding LinkedIn, our respondents declared having some expectations about the system (*INT_BF02*: $M = 4.36$, $SD = 1.48$) and specific tasks to achieve (*INT_BF03*: $M = 4.6$, $SD = 2.05$). A huge majority of participants had previous experience with the system (*INT_BF04*: $M = 6.12$, $SD = 1.45$), with only 2.2% of respondents having no previous experience at all with LinkedIn. Moreover, they report using the system regularly (*TEMP01*: $M = 6.24$; $SD = 1.1$). The time spent on the system during the studied interaction seemed variable (*TEMP02*: $M = 4.29$, $SD = 1.79$) with about half of the respondents considering their interaction time as short while the other half considered it as long. Finally, the high rating on item TEC01 ($M = 6.23$, $SD = 1.4$) reflects that participants have not encountered technical issues during their interaction with the system.

4.4. LINKS BETWEEN OBJECTIVE AND PERCEIVED CONTEXTUAL DIMENSIONS

Our results also show links between the objective reported context of use and perceived quality of the context. Due to space constraints, only main correlations at the item level will be reported here.

Regarding the physical environment, items PHYS01, PHYS03, and PHYS04 are positively correlated to the global perceived context ($r = .32$, .38 and .42 respectively, $p < .001$). This means that the facts of using LinkedIn in a familiar, quiet and steady environment are linked to a better assessment of the perceived context. Item PHYS01 (*I was in an unfamiliar place/in a familiar place*) is also positively correlated to perceived resources ($r = .34$, $p < .001$), suggesting that users in a familiar place perceive their internal resources as better than users in an unfamiliar place. Finally, the correlation between PHYS07 (*I was moving/I was remaining still*) and TEC01 (*technical problems were encountered/no technical problems were encountered*) ($r = .45$, $p < .001$) suggest that more technical problems were encountered when the user was moving (therefore probably using the mobile version of LinkedIn).

Results also show that being in a private space (vs. being in a public space; item SOC01) is related to a better focus on the task (correlation with task context $r = .31$, $p < .001$) and a better assessment of perceived physical context (correlation with physical context $r = .39$, $p < .001$). On the contrary, SOC02 (*I was not interacting with people/interacting with people*) is

negatively correlated to the task context (r = -.38, p < .001) and especially to the item TASK02 (r = -.48, p < .001): the more users were interacting with people, the less they were focused on the task. Unsurprisingly, there is a negative correlation between SOC01 and SOC02 (r = -.28, p < .001), suggesting that the more private the location, the less interaction one has with other people.

Regarding objective internal context, the fact of having information about the system (INT_BF05) is linked to the global perceived context (r = .36, p < .001). The results also show that the more users had specific tasks to achieve, the higher their expectations about the system (correlation between BF03 and BF02; r = .31, p < .001). Unsurprisingly, the correlation between perceived resources and INT_BF04 (r = .21, p = .015) and INT_BF05 (r = .29, p = .001) suggest that having previous experience and information about the system increases the assessment of perceived internal resources. Similarly, INT_BF04 and INT_BF05 are also correlated to TEC01 (r = .28, p = .001 and r = .27, p = .002 respectively): users having more experience and information about LinkedIn encountered less technical problems. Having information about the system (INT_BF05) is also related to the temporal context (r = .35, p < .001): the more users are informed about the system before using it, the less they feel annoyed by time pressure or have the impression that it was not the right moment to use the system. Finally, having specific tasks to achieve (INT_BF03) is linked to the task context (r = .28, p = .001), which means that users having tasks to achieve felt more focused on the task, less interrupted and more goal-oriented.

TEC01 (*technical problems were encountered/no technical problems were encountered*) is moderately correlated to the global perceived context (r = .45, p < .001) and weakly to the physical context (r = .18, p < .05). Besides these two dimensions, TEC01 is not correlated to any other UXCS subscale. This result might be explained by the fact that only few technical issues seem to have occurred during participants' interactions with the website, therefore leading to a high average score on this item (M = 6.23, SD = 1.4). Finally, TEMP01 (*I used the system only once/I am using the system regularly*) is correlated to the internal context (r = .38, p < .001), the perceived resources (r = .27, p = .001) and the global perceived context (r = .38, p < .001). TEMP02 (*I spent a short time using the system/I spent a long time using the sytem*) is related to the internal context (r = .27, p = .002) and the global perceived context (r = .23, p = .007) but no significant correlation was found with perceived resources, social context nor with task context.

4.5. IMPACT OF CONTEXTUAL DIMENSIONS ON USER EXPERIENCE

Results show that, overall, the user experience of the professional network LinkedIn (*ATD_TOTAL*) is evaluated as quite good with an average score of M = 4.82 (SD = 0.91). Attractiveness (M = 5.11, SD = 1.03) and Pragmatic Quality (M = 4.91, SD = 1.05) of the system were assessed as better than its Hedonic Quality (M = 4.51, SD = 1.04). Table 6 presents the descriptive statistics and reliability analyses related to the AttrakDiff scale and subscales. No significant links were found between AttrakDiff

evaluations and demographics like age, gender and technology literacy, suggesting that perceived UX does not depend on the user profile.

Table 6. Abridged AttrakDiff scale: Descriptive Statistics and Reliability Analyses

	Mean	St. Dev.	N	Nb items	Cronbach's Alpha
AttrakDiff Total (ATD_TOTAL)	4,82	0,91	137	9	.90
Pragmatic Quality (ATD_PRAG)	4,91	1,05	137	4	.83
Hedonic Quality (ATD_HEDO)	4,52	1,04	137	3	.79
Attractiveness (ATD_ATT)	5,11	1,03	137	2	.84

What are the impacts of contextual dimensions on user experience? Table 7 shows the bivariate correlations between the UXCS and the AttrakDiff, along with their respective subscales. First, we see that the UXCS is positively correlated to the AttrakDiff (r = .43, p < .001) and to all its subscales (correlations ranging from .37 to .39, p < .001). The more the context is perceived as positive, the more the user experience is reported as positive as well. Overall perceived quality of the interaction context statistically explained 18% of the variability in user experience assessment.

At the subscale level, the main correlation is found between the perceived internal context and the AttrakDiff scale (r = .71, p < .001) and subscales. Perceived internal context is therefore the dimension being the most strongly linked to the global user experience. At the item level, results show a weak correlation between user's mood before using the system (INT_BF01) and the global user experience (r = .22, p = .009) and a stronger correlation between user's mood while or after using the system (INT_WH03 and INT_AF01) and the same assessment of global experience (both correlations being equal to r = .55, p < .001). Regarding the impact of the opinion about the system on the UX, we see a strong correlation between INT_BF06 and the AttrakDiff (r = .62, p < .001) and its subscales (ranging from r = .42 for hedonic quality to r = .61 for pragmatic quality). This confirms previous findings by Raita and Oulasvirta (2011), which showed how favourable product expectations boost subjective usability ratings. Voluntariness to use the system (INT_BF07) is also correlated to the AttrakDiff (r = .40, p < .001) and its subscales. The more users felt free to use LinkedIn, the better the reported user experience. User's motivation (INT_WH01), interest (INT_WH02) and satisfaction during interaction (INT_WH04) are all strongly correlated to the AttrakDiff score (ranging from r = .55 to r = .60). Our results also highlight a positive correlation between the Attrakdiff score and the willingness to use the system again (INT_AF03, r = .45, p < .001). Noteworthy is that the respondents reported a very strong willingness to use LinkedIn again (M = 6.04, SD = 1.16). Finally, feeling attracted by the system (INT_AF04) is also strongly correlated to the AttrakDiff (r = .65, p < .001) and all its subscales.

In addition to the internal context subscale, three other subscales are significantly correlated to the AttrakDiff: social context subscale (r = .33, p < .001), perceived resources subscale (r = .36, p < .001) and temporal context subscale (r = .28, p = .001). The correlations are however weaker than the strong Pearson coefficients observed between the Internal Context and the User Experience evaluation performed through the AttrakDiff. The social context subscale is especially linked to the perceived hedonic quality

(r = .31, p < .001) and attractiveness (r = .33, p < .001), whereas the perceived resources subscale is more correlated to the pragmatic quality of the system (r = .38, p < .001) than to its hedonic quality (r = .26, p = .002) or attractiveness (r = .25, p = .003). The positive correlations between the temporal context and AttrakDiff subscales (ranging from .217 to .285) confirm previous findings by Partala and Kallinen (2012), who witnessed a bigger level of time pressure during most unsatisfying experiences compared to most satisfying experiences reported by users.

The two remaining UXCS subscales, namely physical context and task context, are not correlated to the global AttrakDiff rating nor to any of its subscales. Noteworthy is that two of the correlations' p. values linking task context and AttrakDiff scores are close to the threshold of .05 with a 2-tailed significance test and would have been significant with a 1-tailed test. For these last two subscales, we therefore explored the item level to better understand their link with LinkedIn reported User Experience. Item PHYS02 (*I was in an unpleasant location / I was in a pleasant location*) is positively but weakly correlated with the AttrakDiff (r = .18, p = .037) and its pragmatic quality subscale (r = .21, p = .012). Regarding the task context, only item TASK02 (*I was often interrupted / I was never interrupted*) is significantly and negatively correlated to the AttrakDiff (r = -.27, p = 001) and its subscales (r = -.28, p = .001 for pragmatic quality; r = -.20, p = .021 for hedonic quality; r = -.21, p = .013 for attractiveness). At first glance, this negative link seems puzzling: the more a user was interrupted during his interaction with LinkedIn, the better the reported user experience. However, the aforementioned negative correlation between TASK02 and SOC02 (*I was not interacting with people / I was interacting with people*) (r = -.47, p < .001) might explain this result as interruptions are related to the presence of other people perceived as positive.

Table 7. Bivariate correlations between UXCS subscales and AttrakDiff (N = 137)

Perceived Context Subscale		ATD TOTAL	ATD PRAG	ATD HEDO	ATD ATT
UXCS_TOTAL	Pearson Correlation	,431**	,373**	,379**	,386**
	Sig. (2-tailed)	,000	,000	,000	,000
PHYS_CTX	Pearson Correlation	,093	,103	,086	,030
	Sig. (2-tailed)	,279	,229	,317	,725
SOC_CTX	Pearson Correlation	,327**	,248**	,312**	,328**
	Sig. (2-tailed)	,000	,003	,000	,000
INT_CTX	Pearson Correlation	,713**	,632**	,587**	,670**
	Sig. (2-tailed)	,000	,000	,000	,000
RESOURCES_CTX	Pearson Correlation	,356**	,379**	,262**	,251**
	Sig. (2-tailed)	,000	,000	,002	,003
TASK_CTX	Pearson Correlation	-,155	-,161	-,122	-,106
	Sig. (2-tailed)	,071	,060	,157	,218
TEMP_CTX	Pearson Correlation	,281**	,217*	,285**	,251**
	Sig. (2-tailed)	,001	,011	,001	,003

This phenomenon is also visible when comparing participants having experienced a satisfying UX (participants having reported a mean AttrakDiff rating superior to 4 out of 7, n = 109) to participants having experienced an unsatisfying UX (participants having reported a mean AttrakDiff rating inferior or equal to 4 out of 7, n = 28). Figure 1 shows the differences in the mean evaluation of perceived contextual factors between

users satisfied with UX and users unsatisfied with the UX of LinkedIn. Inter-group differences were tested through independent-samples t-tests and the sets of data differ significantly for all subscales, except the Physical Context subscale. In the case of satisfying UX, the perceived context (*UXCS_TOTAL*) is reported as better than in the case of unsatisfying UX (M = 5.41 vs M = 5), *diff*=0.41, $t(135) = 3.06$, $p = .003$. The same holds true for the social context, perceived resources, internal context and temporal context. Unsurprisingly, the main difference is found for the internal context (M = 5.62 vs. M = 4.28), which is strongly linked to the AttrakDiff evaluation (*diff*=1.34, $t(135)=7.15$, $p < .001$). The only subscale presenting higher ratings in the case of unsatisfying UX is the task context, *diff*=0.5, $t(135) = -1.78$, $p = .038$, which confirms previous observations and will be discussed in the next section of the paper.

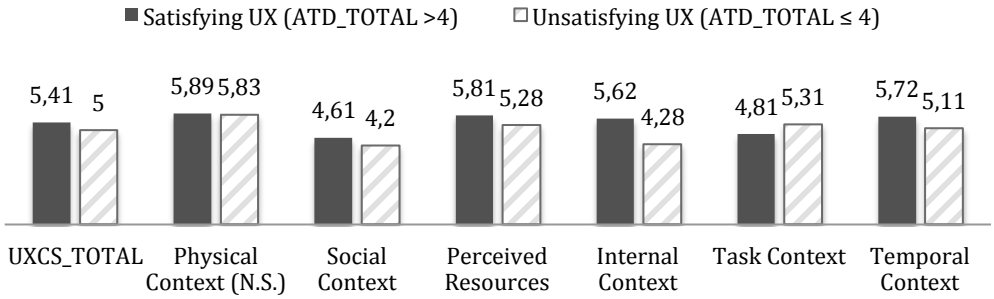


Figure 1. Salience of perceived contextual factors in satisfying and unsatisfying user experiences

A linear regression performed with UXCS_TOTAL as an independent variable and ATD_TOTAL as a dependant variable show that context properties predicted UX scores, Bêta = .43, $t(136) = 5.55$, $p < .001$, and that 18.6% of the total variation in the reported UX can be explained by the perceived quality of the context, $R^2 = .19$, $F(1, 135) = 30.76$, $p < .001$. When only entering INT_TOTAL as independent variable, we see that this variable predicts UX scores, Bêta = .71, $t(136) = 11.8$, $p < .001$, and explains 50.8% of the total variation in the reported AttrakDiff rating, $R^2 = .51$, $F(1, 135) = 139.34$, $p < .001$. Perceived context, and especially internal context is useful as a predictor of UX assessed by the AttrakDiff scale.

It is also interesting to study the links between objective contextual parameters and User Experience to see how the interaction context impacts the felt experience. While there are no significant correlations between objective items of the physical context (PHYS01, PHYS03, PHYS04 and PHYS07) and the AttrakDiff scale and subscales, significant correlations link items of the objective social context and the reported LinkedIn User Experience. SOC01 (*I was in a public space/in a private space*) is negatively correlated to the AttrakDiff global score ($r = -.17$, $p = .048$) whereas SOC02 (*I was not interacting with people/interacting with people*) is positively correlated to the AttrakDiff ($r = .22$, $p = .009$). Being in a private space is slightly correlated with a decrease in the assessment of UX. Furthermore, results show that having higher expectations about the system (INT_BF02) is correlated with a better AttrakDiff rating ($r = .24$, $p = .005$). Except for this

item, all other objective items of the internal context (INT_BF03, INT_BF04, INT_BF05) are not correlated to the AttrakDiff scale or subscales. The fact of reporting a technical issue or not (TEC01) is not correlated to the AttrakDiff score either. However, as the average rating for TEC01 is very high (meaning that almost none of our respondents reported any issue during the interaction), the absence of link between technical issue and User Experience should be explored further in future studies. Finally, TEMP01 (r = .22, p = .011) and TEMP02 (r = .25, p = .004) are correlated to the AttrakDiff, suggesting that there is a link between the fact of using the system regularly or using the system during a long time span and the quality of the felt experience.

5. DISCUSSION

5.1. DEVELOPMENT OF THE UXCS

The results of this study confirm the interest of the UXCS for the measurement of both objective and perceived contextual factors involved in User Experience. Regarding perceived contextual factors, principal components analysis and reliability analyses have shown the coherence of the current structure of the scale including 24 items divided into 6 contextual dimensions. As compared to the initial envisioned structure, four items were deleted (INT_AF02, SOC05, INT_BF01 and TASK04) and four items of the Internal Context subscale that were loading on a separate factor now compose the Perceived Resources subscale. All UXCS subscales show a high internal consistency, with Cronbach's alpha ranging from .68 (*Perceived Resources*) to .93 (*Internal Context*).

The Internal Context subscale shows a very high Cronbach's score (.93), perhaps reflecting a redundancy between some items. This subscale was intended to encompass the temporal dynamics involved during the interaction. This is why the subscale is built around 3 temporal steps "before", "during" and "after" and includes more items than the other subscales. The number of items also reflects the importance of the internal context in the study of UX. In the current study, all items of the Internal Context subscale were answered at the same time and this might explain the high correlation between some items (the assessment after the interaction of the user's mood before interaction might be biased by memory processes). According to us, this does not affect the quality of the subscale as the most important UX is the remembered UX (Karapanos, Martens, & Hassenzahl, 2012). However, we aim to use this subscale more efficiently in further studies by asking participants to answer the items separately before, during and after the interaction.

In the design of the UXCS, the distinction between objective and perceived contextual dimensions allows for a better understanding of the context of use and its impact on the felt experience. In the case of LinkedIn, the 13 items focused on the objective use context provide us with valuable information. We learned that the interaction generally takes place in a quiet, familiar place and that the users are majorly not interacting with other people. The participants had previous experience with the system, which is unsurprising considering the fact that this online study was broadcast

directly on the professional network. The users declared having some expectations about the system and specific tasks to achieve. This information might seem trivial and one might argue that user research (e.g., interviews or observation) could provide the same results. However, in the case of an online study, UX practitioners rarely have access to such valuable information to characterize the context of interaction. In addition, as these 13 items can easily be used as independent variables in statistical analyses, the UXCS allows for the evaluation of the impact of these contextual parameters on UX. As an illustration, we saw that being in a private space is correlated with a decrease in the assessment of LinkedIn UX whereas interacting with people is correlated with an increase in UX rating. We also saw that more technical problems were encountered when the user was moving, which suggests that the mobile version of LinkedIn might be less effective than the standard version.

Following Grimm and Yarnold's (1995) recommendations, validating the construction of our scale through principal component analyses required a representative sample, sized minimum 5 times bigger than the amount of items used (and a minimum sample size of 100 participants in any case regardless of the ratio). This prerequisite is called "subjects-to-variables (STV) ratio". Having 137 participants in our sample, we were able to adequately analyse the results of the 24 pairs of items composing the perceived context of the UXCS. Nevertheless, we are aware that the validation of a summated rating scale such as the UXCS requires the consolidation of results through several studies involving large samples of participants and various use cases.

5.2. EXPLORATION OF THE LINKS BETWEEN UX AND CONTEXTUAL DIMENSIONS

Beyond aspects related to the construction of the scale, this study also contributes to exploring the links between UX and the several context dimensions composing the UXCS. First, we saw that the UXCS is positively correlated to the AttrakDiff ($r = .43$, $p < .001$) and to all its subscales. The more the context is perceived as positive, the more the user experience is reported as positive as well.

At the subscale level, the internal context, which was already identified as a key dimension in the literature, is significantly and strongly correlated to every AttrakDiff's subscale and even to AttrakDiff global score (ranging from .59 to .71). This subscale explains 50.8% of the total variation in the reported UX score assessed by the AttrakDiff scale. Hedonic and pragmatic qualities as well as attractiveness might therefore be partly explained by the internal context of the user. This is not really surprising considering the fact the UX is focused on the user and his/her feelings and emotions during the interaction while the internal context subscale includes items related to user's interest, motivation, satisfaction or attraction to the system. As mentioned in the introduction of this paper, the user is one of the pillars influencing UX, along with the system and the context. In this way, studying the internal context as a distinct dimension may seem redundant since the user (including his internal context) is already part of the UX. This assumption would support Hassenzahl's understanding of the context and

the way his research work on UX deals with context. Instead of focusing on the situation (context), he argued to focus on the meta-motivational state the situation triggers, so called "usages modes" (Hassenzahl, Kekez, & Burmester, 2002). His claim is that these usage modes may be easier to define and more general than the complex description of the situation. In other words, it seems to play no role whether some contextual factors are positive or negative unless this somehow impacts the user's internal state, which in turn impacts product perception and evaluation. This internal context would "summarize" external attributes of the environment. Usages modes were experimentally manipulated in some of his studies (Hassenzahl, 2007; Hassenzahl, Schöbel, & Trautmann, 2008) and a "Usage Mode Scale" was even featured in one of them (Hassenzahl et al., 2002). We believe that data provided in the present study tend to support Hassenzahl's assumption claiming that what is crucial is the internal state of mind, triggered by the context.

More surprisingly, we observed a negative correlation between TASK02 and the AttrakDiff scale (r = -.27, p < .001) and subscales: the more a user was interrupted during his interaction with LinkedIn, the better the reported user experience. When looking at the differences in the mean evaluation of perceived contextual factors between users satisfied with UX and users unsatisfied with the UX of LinkedIn, results confirmed that the only subscale presenting higher ratings in the case of unsatisfying UX was the task context (*diff*=0.5, $t(135)$ = -1.78, p = .038). Paradoxically, the effects of interruptions in the interaction on UX are described as majorly negative in the literature. An empirical study in an office environment showed that interrupting the user while he was carrying out a task on his computer (document editing, web-searching or watching a media) had a significant effect on frustration, annoyance, time pressure and mental effort (Adamczyk & Bailey, 2004). By trying to understand the most satisfying and unsatisfying experiences, Partala and Kallinen (2012) also showed a tendency of more distractions during unsatisfying experiences. In positive psychology, the concept of flow (Czikszentmihaly, 1990), often described as the "ultimate experience", seems to support the need for focus on the task to reach an optimal user experience. In essence, flow is characterized by complete absorption in what one does; "intense and focused concentration on the present moment" being one of the six factors encompassing an experience of flow. Therefore, in contradiction with the assumption that focus is necessary to achieve a great UX, LinkedIn experience seems to be better when users were interrupted and multitasking. We argue that the effect of interruptions might depend on the nature of the task. In the case of a strongly goal-oriented interaction (as it is the case in a work environment, for example), interruptions might affect the interaction whereas they might be perceived as positive in the case of an interaction only guided by a free exploration, as it might be the case when using a social or professional network. The main drawback of multitasking lies in the decrease of performance in the primary task (Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005; Wickens, 2008), which might not negatively impact the interaction if the primary task is explorative and does not include a precise objective to achieve.

Results also showed that interruptions were correlated in our study with the presence of social interactions, as suggested by the correlation between

SOC02 and TASK02. As stated in the UX White Paper (2011), "UX is rooted in a social and cultural context. It is not about just an individual using a system in isolation." (p. 6) Other people - being physically present or through the mediation of technology - therefore influence the UX of an interaction. In a qualitative study using description of personal experiences narratives, Tuch, Trusell, and Hornbæk (2013) found that negative user experiences were more likely to occur at home and when users are alone, whereas positive experiences occur more frequently in the presence of other people. The fulfilment of the human primary need for relatedness (Sheldon et al., 2001) might therefore also explain the positive impact of interruptions on the user experience. It is noteworthy that the impact of the social context is positive only when the presence of other people is perceived as desirable: when social context is perceived as disturbance from surrounding people (vs. perceived as relatedness), it may have a negative influence on user experience. Wac et al. (2011) showed that in a mobile usage context UX was evaluated better when the user was alone and could better focus on the application being used. Social context here was impeding the task context, which in turn led to a deterioration of the quality of UX.

Our results show an absence of correlation between the physical context and the AttrakDiff scale or any of its subscales, suggesting that the physical environment is not linked to the felt user experience. In the present study, the very high average rating and small standard deviation related to the Physical Context subscale (M = 5.88, SD = 0.93) does not allow for a distinction between users being placed in a deteriored physical environment versus users being placed in a comfortable physical environment. Previous research has shown that characteristics of the physical context such as the physical location or user's mobility affected task performance (Jumisko-Pyykkö & Hannuksela, 2008; Lin, Goldman, Price, Sears, & Jacko, 2007; Mustonen, Olkkonen, & Häkkinen, 2004). Hence, it seems that different aspects of physical context indeed play a great role in the quality of an interaction, and unfavourable contexts have been compared in the literature to temporary physical or cognitive impairments (Barnard, Yi, Jacko, & Sears, 2007). This line of research is related to the concept of "situationally-induced impairments and disabilities" (Sears, Lin, Jacko, & Xiao, 2003; Sears, Young, & Feng, 2003). In this view, users' abilities are constrained by contextual factors (e.g., one-hand operation, vibrations, noise, temperature, lighting, demands for the user's attention, stress), which temporarily impede reaching an optimal level of performance in the interaction. As these situationally-induced impairments are temporary, the user is less likely to develop a coping strategy. This fact strengthens the need for devices and interfaces that adapt automatically depending on a given physical context (Sears et al., 2003). To explain the absence of correlation between physical context and the perceived UX in the present study, our assumption is that the physical context may act as a "hygiene factor", which would not cause a positive user experience when environmental conditions are good but would significantly reduce the perceived quality of the experience in case conditions are bad or disturbing. This is very similar to Herzberg's two-factor theory (Herzberg, Mausner, & Snyderman, 1959) stating that factors causing job satisfaction are distinct from factors causing dissatisfaction. This dual-factor theory distinguishes between motivators that give positive satisfaction and hygiene factors, which do not give positive satisfaction but

lead to dissatisfaction when they are not given. Applied to UX needs fulfilment, Hassenzahl et al. (2010) named this phenomenon a "deficiency need", which he defines as a "need that creates negative affect if blocked, but not necessarily strong positive feelings if fulfilled" (p. 358). Similarly, the physical context would negatively impact the UX if perceived as uncomfortable but not necessarily create a positive UX if perceived as comfortable. The same assumption could be relevant to test regarding technical issues a user might encounter during interaction (item TEC01). Not encountering any issue would not be seen as particularly satisfying, whereas encountering technical problems could strongly impede the quality of the interaction and reported UX.

Finally, we also observed some correlations between perceived contextual dimensions, for example between social and internal Context (r =.45, p = .001). As relatedness is described as one of the basic psychological needs (Sheldon et al., 2001; Hassenzahl, 2010) that every human aims to fulfil, we assume that a positive social context will have a positive impact on the internal context through the fulfilment of this relational need.

5.3. LIMITATIONS AND FUTURE WORK

The present research entails some limitations that require discussion. First, the UXCS has been tested in the present study on a single sample involving a single use case (LinkedIn). The sample size used to validate the structure of the scale is at the lower limit of the recommended subjects-to-variables (STV) ratio and it will be necessary to replicate these results on a larger representative sample to ensure the stability and validity of the UXCS structure.

Moreover, as highlighted in the literature and confirmed by our results, the impacts of context dimensions on user experience are themselves dependent on the interaction context. As the application of the UXCS was limited here to the particular situation of an interaction with a professional network, the generalizability of our results linking UX and context dimensions might be limited. In future studies, it will therefore be necessary to better explore the relationships between context and UX and their variations according to this context. In order to do so, we plan to apply the UXCS to different use cases encompassing various usage modes, and to pay a special attention to the four general properties of the context as described by Jumisko-Pyykkö and Vainio (2010). The impacts of the level of magnitude, level of dynamism or type of patterns on the UX should be explored, as well as the typical combinations of contextual dimensions (i.e., which different context elements are usually present simultaneously).

Finally, it is also worth mentioning that the sequence of delivery between the UXCS and the AttrakDiff scale was not balanced in the present study, the AttrakDiff being always the first scale to be answered by participants. As a potential sequence effect might have impacted some findings, this is a potential limitation of this study that should be controlled for in further work.

6. CONCLUSION

The development of the UXCS aims at providing a holistic measurement of contextual factors for UX assessment. The creation of this tool contributes to filling the gaps in UX Evaluation questionnaires, which are currently lacking a self-declared measure of the quality of the context. The observations brought by our results confirm the relevance of the UXCS as a distinct measure of the context especially useful when one needs to control for the influence of the context or simply wants to know more about contextual aspects involved in a situation. Being able to isolate the portion of variance due to each contextual dimension is one of the benefits of the UXCS. The interest of this tool has been confirmed through this study, which also showed that the format of the UXCS makes it an easy to use, easy to deploy and easy to answer questionnaire. From a psychometrical perspective, the current UXCS structure was confirmed by a PCA showing good internal consistencies of its subscales. Of course, additional use cases involving other experimental situations will provide empirical validation of our scale. By trying to build a quantitative and self-reported measurement tool to assess contextual factors, we aimed at easing and promoting the integration of the context as a core dimension of every UX evaluation. In the assessment of context, the UXCS scale can be used both for the design of an interactive system and for its evaluation. In the first case, the focus will be on the process (formative evaluation) allowing product designers to better take into account contextual factors. In the second case, the focus will be on the product itself (summative evaluation), and the UXCS will collect valuable information on the influence of context on the perceived quality of interaction. Both applications might be especially relevant for highly context-dependent products, as for example mobile devices.

In this study, we found strong evidence that contextual factors impact the assessment of UX, which confirms previous findings. But most of all, we gained better insight on how this impact of context on UX works. Correlations between UX measured by the AttrakDiff questionnaire and contextual factors measured by the UXCS provide us with valuable information to better understand how context impacts UX.

Finally, we would like to highlight that, as suggested by Dey (2001), we believe that the relevance and impact of each contextual dimensions for a particular system or use situation is itself dependent on context. The results obtained in this study on LinkedIn regarding the impact of a specific dimension of context on UX are therefore to be confirmed on similar systems and explored on different systems in order to gain a better vision of how context impacts UX. We think that the UXCS will be a valuable tool to achieve this research objective.

Appendix A. Final version of the UXCS

Following are pairs of words/sentences that describe the context of interaction. Each pair represents contrast of ideas about the context in which the interaction occurred. Your mission is to rate each item on a 7-points scale. This is not an evaluation of the quality of the system but an attempt to assess your environment and feelings.

Physical context

While using the system, I was:

In an unfamiliar place	○ ○ ○ ○ ○ ○ ○	In a familiar place
In an unpleasant location	○ ○ ○ ○ ○ ○ ○	**In a pleasant location**
In a noisy place	○ ○ ○ ○ ○ ○ ○	In a quiet place
In a moving / vibrating environment	○ ○ ○ ○ ○ ○ ○	In a steady environment
Unpleased with the temperature	○ ○ ○ ○ ○ ○ ○	**Pleased with the temperature**
Unpleased with the lighting	○ ○ ○ ○ ○ ○ ○	**Pleased with the lighting**
Moving (mobile usage)	○ ○ ○ ○ ○ ○ ○	Remaining still

Social Context

While using the system, I was:

In a public space	○ ○ ○ ○ ○ ○ ○	In a private space
Not interacting with people	○ ○ ○ ○ ○ ○ ○	Interacting with people
Feeling alone	○ ○ ○ ○ ○ ○ ○	**Feeling related to other people**
Feeling unsupported	○ ○ ○ ○ ○ ○ ○	**Feeling supported**

Internal context

Before using the system,

I had no expectations about the system	○ ○ ○ ○ ○ ○ ○	I had high expectations about the system
I had no specific tasks to achieve	○ ○ ○ ○ ○ ○ ○	I had specific tasks to achieve
I had no previous experience with the system	○ ○ ○ ○ ○ ○ ○	I already had experience in using the system
I had no information about it	○ ○ ○ ○ ○ ○ ○	I had information about it
I had a bad opinion about it	○ ○ ○ ○ ○ ○ ○	**I had a good opinion about it**
I felt pressed to use it	○ ○ ○ ○ ○ ○ ○	**I felt free to use it**

While using the system, I felt:

Demotivated	○ ○ ○ ○ ○ ○ ○	**Motivated**
Not interested	○ ○ ○ ○ ○ ○ ○	**Interested**
In a bad mood	○ ○ ○ ○ ○ ○ ○	**In a good mood**
Unsatisfied	○ ○ ○ ○ ○ ○ ○	**Satisfied**
I had insufficient skills to use it	○ ○ ○ ○ ○ ○ ○	**I had sufficient skills to use it**
I did not have enough time to spend on it	○ ○ ○ ○ ○ ○ ○	**I had enough time to spend on it**
I had the need to be helped or trained	○ ○ ○ ○ ○ ○ ○	**I had no need to be helped or trained**
Powerless over my environment	○ ○ ○ ○ ○ ○ ○	**In control of my environment**

After using the system,

I am in a bad mood	○ ○ ○ ○ ○ ○ ○	**I am in a good mood**
I feel unsatisfied	○ ○ ○ ○ ○ ○ ○	**I feel satisfied**
I am not willing to use it again	○ ○ ○ ○ ○ ○ ○	**I am willing to use it again**
I feel repelled by the system	○ ○ ○ ○ ○ ○ ○	**I feel attracted by the system**

Technical Context

While using the system,

Technical problems were encountered	○ ○ ○ ○ ○ ○ ○	No technical problems were encountered

Task Context

While using the system,

I was doing several things simultaneously	○ ○ ○ ○ ○ ○ ○	**I was focusing on the task**
I was often interrupted	○ ○ ○ ○ ○ ○ ○	**I was never interrupted**
I focused on the product	○ ○ ○ ○ ○ ○ ○	**I focused on attaining my goals**

Temporal Context

Overall,

I used the system only once	○ ○ ○ ○ ○ ○ ○	I am using the system regularly
I spent a short time using the system	○ ○ ○ ○ ○ ○ ○	I spent a long time using the system
I was interacting at an uncomfortable pace	○ ○ ○ ○ ○ ○ ○	**I was interacting at a comfortable pace**
It was not the right moment to use the system	○ ○ ○ ○ ○ ○ ○	**It was the right moment to use the system**

REFERENCES FOR PAPER C

Adamczyk, P. D., & Bailey, B. P. (2004). If Not Now, When?: The Effects of Interruption at Different Moments Within Task Execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI'04. 6(1), 271–278. doi:http://dx.doi.org/10.1145/985692.985727

Alonso-Ríos, D., Vázquez-García, A., Mosqueira-Rey, E., & Moret-Bonillo, V. (2010). A context-of-use taxonomy for usability studies. *International Journal of Human-Computer Interaction, 26*(10), 941-970.

Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2007). Capturing the effects of context on human performance in mobile computing systems. *Personal Ubiquitous Computing*, 11(2), 81–96. doi:10.1007/s00779-006-0063-x

Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behaviour and Information Technology, 13*, 132–145.

Bradley, M., & Lang, P. (1994). Measuring emotion: the Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.

Bradley, N. A., & Dunlop, M. D. (2005). Toward a multidisciplinary model of context to support context-aware computing. *Journal of Human-Computer Interaction, 20*(4), 403-446. doi:10.1207/ s15327051hci2004_2

Crilly, N., Moultrie, J. & Clarkson, P. J. (2004). Seeing things: consumer response to the visual domain in product design. *Design Studies, 25*, 547-577.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper & Row.

Dey, A. (2001). Understanding and using context. *Personal Ubiquitous Computing*, 5(1), 4-7. doi:10.1007/s007790170019

Dey, A. K. & Abowd, G. D. (1999). *Towards a better understanding of context and context-awareness*. Technical Report GIT-GVU-99-22, Georgia Institute of Technology.

Dourish, P. (2004). What we talk about when we talk about context. *Personal Ubiquitous Computing, 8*(1), 19-30. doi:10.1007/s00779- 003-0253-8

Forlizzi. J., & Ford. S. (2000). The building blocks of experience: An early framework for interaction designers. *Proceedings of Designing Interactive Systems*, 419-423. New York: ACM Press.

Grimm, L. G., & Yarnold, P. R. (Eds.). (1995). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.

Hassenzahl, M. (2002). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*, 479-497.

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

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff : Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In : J. Ziegler & G. Szwillus (Eds.) *Mensch & Computer 2003: Interaktion in Bewegung*. Teubner : Stuttgart, Germany, pp. 187-196.

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers, 22*(5), 353-362. doi:10.1016/j. intcom.2010.04.002

Hassenzahl, M., Kekez, R., & Burmester, M. (2002). The importance of a software's pragmatic quality depends on usage modes. In H.Luczak, A. E. Cakir, & G. Cakir (Eds.). In *Proceedings of the 6th international conference on Work With Display Units* (WWDU 2002) (pp. 275-276). Berlin: ERGONOMIC Institut für Arbeits- und Sozialforschung.

Hassenzahl, M., Platz, A., Burmester, M., Lehner, K. (2000). Hedonic *and* Ergonomic Quality Aspects Determine a Software's Appeal. In: T. Turner, & G. Szwillus (Eds.), *Proceedings of the CHI 2000 Conference on Human Factors in Computing*. New York: ACM, 201 – 208.

Hassenzahl, M., Schöbel, M., & Trautmann, T. (2008). How motivational orientation influences the evaluation and choice of hedonic and pragmatic interactive products: The role of regulatory focus. *Interacting with Computers, 20* (4-5): 473-479. doi:10.1016/j.intcom.2008.05.001

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

Hassenzahl, M., & Ullrich, D. (2007) To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, 19(4), 429-437. doi:10.1016/j.intcom.2007.05.001

Herzberg, F., Mausner, B., & Snyderman, B. (1959). *The Motivation to Work* (2nd ed.). New York: John Wiley.

International Organization for Standardization. (1998). *ISO 9241–11:1998, Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs), Part 11: Guidance on Usability*. Geneva, Switzerland: International Organization for Standardization.

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

International Organization for Standardization. (2010). ISO 9241-210:2010, *Ergonomics of human-system interaction -- Part 210: Human-centred design for*

interactive systems. Geneva, Switzerland: International Organization for Standardization.

Izard, C.E. (1977). *Human emotions*. New York: Plenum Press.

Jumisko-Pyykkö, S., & Hannuksela, M. M. (2008). Does context matter in quality evaluation of mobile television? In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* MobileHCI 08 (pp. 63–72). ACM. doi:10.1145/1409240.1409248

Jumisko-Pyykkö, S., & Vainio, T. (2010). Framing the Context of Use for Mobile HCI - Review paper about mobile contexts of use between 2000-2007. *International Journal of Mobile Human-Computer-Interaction*, 3(4).

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 32-36.

Kaltz, J.W., Ziegler, J., & Lohmann, S. (2005). Context-aware Web Engineering: Modeling and applications. *Revue d'Intelligence Artificielle, 19* (3): 439–458. doi:10.3166/ria.19.439-458.

Karapanos, E., Martens, J.-B., & Hassenzahl, M. (2012). Reconstructing experiences with iScale. *International Journal of Human-Computer Studies*, *70*(11), 849–865. doi:10.1016/j.ijhcs.2012.06.004

Korhonen, H., Arrasvuori, J., & Väänänen-Vainio-Mattila, K. (2010). Analysing user experience of personal mobile products through contextual factors. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia* - MUM '10, 1–10.

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

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.): HCI and Usability for Education and Work USAB 2008, LNCS 5298, pp. 63-76. Berlin/Heidelberg: Springer.

Law, E., Roto, V., Hassenzahl, M., Vermeeren, A. & Kort, J. (2009). Understanding, scoping and defining UX: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'09*, Boston, USA.

Lin, M., Goldman, R., Price, K. J., Sears, A., & Jacko, J. (2007). How do people tap when walking? An empirical investigation of nomadic data entry. *International Journal of Human-Computer Studies,* 65(9), 759-769. doi:10.1016/j.ijhcs.2007.04.001

Maissel, J., Dillon, A., Maguire, M., Rengger, R., & Sweeney, M. (1991). *Context Guidelines Handbook*. MUSiC Project Deliverable IF2.2.2. Teddington, UK: National Physical Laboratory.

Maguire, M. (2001). Context of use within usability activities. *International Journal of Human-Computer Studies, 55*, 453–483. doi:10.1006/ijhc.2001.0486

Mahlke, S., & Thüring, M. (2007). Studying Antecedents of Emotional Experiences in Interactive Contexts. In *Proceedings of CHI 2007* (pp. 915-918). New York: ACM Press.

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

Mustonen, T. Olkkonen, M. & Häkkinen, J. (2004). Examining mobile phone text legibility while walking. Extended abstracts of the 2004 Conference on Human factors in Computing Systems, pp. 1243-1246. New York: ACM Press.

Nacke, L.E., Drachen, A., Kuikkaniemi, K., Niesenhaus, J., Korhonen, H.J., …, & de Kort, Y.A.W. (2009) Playability and Player Experience Research. DiGRA 2009, Breaking New Ground: Innovation in Games, Play, Practice and Theory, Brunel University, United Kingdom.

Osgood, C.E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Chicago: University of Illinois Press.

Oulasvirta, A. (2009). Field experiments in HCI: Promises and challenges. In Saariluoma, P., & Isomaki, H. (Eds.), *Future Interaction Design II*. New York: Springer. doi:10.1007/978-1-84800-385-9_5

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* CHI'05 (pp. 919-928). ACM.

Partala, T., & Kallinen, A. (2012). Understanding the most satisfying and unsatisfying user experiences: Emotions, psychological needs, and context. *Interacting with Computers, 24*(1), 25-34. doi:10.1016/j.intcom.2011.10.001

Raita, E., & Oulasvirta, A. (2011). Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers*, 23(4), 363-371. doi:10.1016/j.intcom.2011.04.002

Roto, V. (2006). *Web Browsing on Mobile Phones - Characteristics of User Experience*. Unpublished doctoral dissertation, Helsinki University of Technology, Finland.

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (2011) User Experience White Paper: Bringing clarity to the concept of user experience. Result from *Dagstuhl Seminar on Demarcating User Experience*, Finland.

Russell, J. A., Weiss, A., & Mendelsohn, G. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology 57*(3), 493–502.

Ryan, R. (1982). Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. *Journal of personality and social psychology 43*(3), 450–461.

Schaeffer, N. C., & Presser, S. (2003). The Science of Asking Questions. *Annual Reviews Sociology, 29*, 65-88.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York: Oxford University Press.

Schilit, B., Adams, N. & Want, R. (1994). Context-aware computing applications. *Proceedings of the 1st International Workshop on Mobile Computing Systems and Applications*. Santa Cruz, CA: IEEE Computer Society.

Sears, A., Lin, M., Jacko, J., & Xiao, Y. (2003). When Computers Fade...Pervasive Computing and Situationally-Induced Impairments and Disabilities. In C. Stephanidis (Ed.), *Human-Computer Interaction: Theory and Practice (Part II)*. (pp. 1298–1302). Hillsdale: Lawrence Erlbaum Associates.

Sears, A., Young, M., & Feng, J. (2003). Physical disabilities and computing technologies: an analysis of impairments. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 482–503). Lawrence Erlbaum Associates.

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*, 325–339.

Spector, P.E. (1992). *Summated rating scale construction. An Introduction*. Newbury Park: Sage.

Spool, J. M. (2005, Jan 04). Putting Context Into Context [Blog post]. Retrieved from http://www.uie.com/articles/putting_context_into_context

Suchman, L. A. (1987). *Plans and Situated Actions: the Problem of Human-Machine Communication*. New York: Cambridge University Press.

Tabachnik, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn and Bacon.

Tuch, A., Trusell, R., & Hornbæk, K. (2013). Analyzing users' narratives to understand experience with interactive products. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2079-2088). Paris, France: ACM.

Wac, K., Ickin, S., Hong, J.-H., Janowski, L., Fiedler, M., & Dey, A.K. (2011). Studying the experience of mobile applications used in different contexts of daily life. In *Proceedings of the first ACM SIGCOMM workshop on Measurements up the stack* (W-MUST '11). ACM, New York, NY, USA, 7-12. doi:10.1145/2018602.2018605

Watson, D., Tellegen, A., Clark, L. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*, 6, 1063-1070.

Wickens, C. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449–455. doi:10.1518/001872008X288394.

Wigelius, H., & Väätäjä, H. (2009). Dimensions of Context Affecting User Experience in Mobile Work. In *Proceedings of INTERACT*, *2*, 604–617.

UX CARDS – EXPERIENCE DESIGN THROUGH PSYCHOLOGICAL NEEDS

D

Carine Lallemand

Submitted for journal publication

Keywords: user experience, UX cards, design, evaluation, card-based method

ABSTRACT

While UX has become a key concern of product development, designing for UX or evaluating UX still remains a challenge. To close the gap between UX research and industry, practical methods need to be developed and current UX research methods need to be adapted to the requirements of industrial UX development.

The psychological needs-driven UX approach is a well-explored area in UX research and appears to be a powerful framework for the design of more experiential interactive systems. However, the transfer from UX research to practice is difficult and slow and this specific approach is not yet widely used by UX practitioners. As card-based methods have been shown to support designers in both the generation of ideas and the evaluation of their designs, we created the UX cards as a pragmatic tool able to support psychological needs-driven UX evaluation and design.

In this paper, we present the iterative development of the UX cards, which might be used for design, evaluation or training purposes. We report on an experiment involving 33 UX experts and aimed at validating the use of the UX cards for UX evaluation. We also present two idea generation techniques to be used for UX design using the UX cards. Our findings suggest that the UX cards are a valuable tool able to support psychological needs-driven UX design or evaluation.

1. CARD-BASED METHODS IN HCI AND DESIGN

When talking about a method involving cards in the field of Human-Computer Interaction (HCI), first thoughts usually point towards the popular card-sorting method (Nielsen, 1995; Spencer, 2009), where users are asked to arrange a set of cards in order to generate a meaningful structure. However, many other card-based methods have been developed during the last decade and serve different purposes in the design process (Wölfel & Merritt, 2013). Some card-sets are rather generic, as the IDEO Methods Cards (2003), which include 51 design methods, or the Experience Design 1 Cards by Shedroff (2009), which broadly cover the topic of experience design by presenting several key questions to address during the design process. Others are more specific, as the Eco Innovators' Design Play Cards (2013), focusing on the design for sustainability, or the Positive Emotional Granularity Cards by Yoon, Desmet, and Pohlmeyer (2013), based on a typology of positive emotions. Just like this latter example, some card-based tools rely on theoretical frameworks and act as tangible translations of this theoretical knowledge. This is for instance the case of the PLEX Cards (Lucero & Arrasvuori, 2010), based on the Playful Experiences (PLEX) framework, a categorization of playful experiences. Other categories of tools are derived from design practice: the 77 Design heuristics instructional cards by Yilmaz et al. (2014) are for instance based on empirical evidence from successful designs. Card-based tools in HCI also vary according to their intended purpose, their format, the numbers of cards they entail, the inclusion of instruction cards or rules of play, or their subdivision into several sub-sets.

In a willingness to provide designers with an overview of available card systems, Wölfel and Merritt (2013) reviewed 18 card-based design tools. They classified each card-set along 5 dimensions: intended purpose and scope, duration of use and placement in the design process, system or methodology of use, customization, and formal qualities. The analysis of their classification suggested three main patterns of card-based design tools:

- General purpose/repository cards: this type of tool offers either a method repository (e.g., IDEO methods cards, 2003) or aim to stimulate inspiration and lateral thinking (e.g., Design to connect by Bleuzé et al., 2014).
- Customizable cards: this type of tool is inherently customizable to some degree, and the card-sets included in this category are often used as a participatory design tool. Examples of customizable card-sets are the Instant Card Technique (Beck, Obrist, Bernhaupt, & Tscheligi, 2008) or Ideation decks (Golembewski & Selby, 2010).
- Context specific cards: this type of tool focuses on a specific design agenda or context, as the Emotional Granularity Cards by Yoon et al. (2013) or the PLEX Cards by Lucero and Arrasvuori (2010).

The advantages of card-based tools are numerous. First, they support creativity and lateral thinking (Bleuzé et al., 2014; Halskov & Dalsgaard, 2006). Card-based methods have been shown to support both novices and experienced designers in generating creative ideas (Yilmaz, Daly, Christian et al., 2014). Both the quantity and the quality of generated ideas are

improved with a higher number of ideas generated and bigger design solutions spaces. Thanks to their physical form, cards are tangible objects easy to manipulate (e.g., shuffling, drawing and sorting) (Chen, Liang, & Chiang, 2011; Hornecker, 2010). They support the design process by literally turning ideas into tangible objects and supporting the focus of design activities (Halskov & Dalsgaard, 2006). Card-based design tools also represent a summary of theories and ideas that might be easily used by designers, who tend to underuse the existing body of knowledge produced in UX or design research (Hassenzahl et al., 2012). As card tools are primarily pragmatic and generally involve no training to be used, they are easy to integrate into the design process. Finally, they are also powerful collaborative tools that may help initiating discussions on a topic and foster collaboration between projects' stakeholders (Halskov & Dalsgaard, 2006; Hornecker, 2010) by serving as a communication tool (Beck et al., 2010), a "low-tech common language that supports rich communication among the stakeholders" (Lafrenière, Dayton, & Muller, 1999, p. 151). Their main limitation is that their content is majorly static (except in the case of customizable cards).

2. PSYCHOLOGICAL NEEDS-DRIVEN UX EVALUATION AND DESIGN

While UX has become a key concern of product development, designing for UX or evaluating UX still remains a challenge. To close the gap between UX research and industry, practical methods need to be developed and current UX research methods need to be adapted to the requirements of industrial UX development (Väänänen-Vainio-Mattila, Roto, & Hassenzahl, 2008). Amongst relevant theories in UX research, the psychological needs-driven UX approach (Kim, Park, Hassenzahl, & Eckoldt, 2011) allows for adaptation to the requirements of industry and translation into practical tools and methods. Card-based tools, which are often used to translate theories into tangible and pragmatic design tools, offer this possibility.

One of the main assumptions in UX research is that an interactive system providing a positive user experience will trigger an engaging experience (Pine & Gilmore, 1998). Many findings (Csikszentmihalyi, & Lefevre, 1989; Hassenzahl, Diefenbach, & Göritz, 2010; Sheldon, Elliot, Kim & Kasser, 2001) inspired by the Self-Determination Theory (Deci & Ryan, 2002; Sheldon, Ryan, & Reis, 1996) suggest that the fulfilment of basic human psychological needs could act as one of the main drivers for a positive experience. It is therefore assumed that a system able to fulfil the need for relatedness, the need for competence or the need for autonomy will support an optimal and engaging user experience. Eight to ten basic needs are usually described within the literature (Sheldon et al., 2001), the three aforementioned (i.e., relatedness, competence and autonomy) being considered as central to the Self-Determination Theory.

The psychological needs-driven UX approach is a well-explored area in UX research (Hassenzahl, 2010; Kim, Park, Hassenzahl, & Eckoldt, 2011; Tuch, Trusell, & Hornbæk, 2013) and appears to be a powerful framework for the design of more experiential interactive systems. Nevertheless, the transfer from UX research to practice is difficult and slow (Odom & Lim, 2008) and

this specific approach is not yet widely used by UX practitioners (Lallemand, Koenig, & Gronier, 2014). Often, novel UX approaches developed in UX research need to be adapted to the requirements of industrial settings (Väänänen-Vainio-Mattila, Roto, & Hassenzahl, 2008). Starting from the assumption that basic human needs constitute a relevant framework to evaluate and design for UX, we created the UX cards as a pragmatic tool able to support psychological needs-driven UX evaluation and design.

3. THE DEVELOPMENT OF UX DESIGN AND EVALUATION CARDS

In order to translate the psychological needs identified within the literature into tangible supports that could be used for the design and evaluation of UX, we designed seven UX Cards. UX Cards were designed both in English and French. The UX Cards set is composed of: a cover card, seven UX needs cards and instruction cards.

3.1. DESIGNING THE UX CARDS

Amongst the ten needs identified by Sheldon et al. (2001), we selected seven candidate needs (Table 1).

Table 1. The seven needs represented on the UX Cards (adapted from Sheldon et al., 2001)

Need	Definition
Relatedness	Feeling that you have regular close contact with people who care about you rather than feeling lonely and uncared of.
Competence	Feeling that you are very capable and effective in your actions rather than feeling incompetent or ineffective.
Autonomy	Feeling like you are the cause of your own actions rather than feeling that external forces or pressure are the cause of your action.
Security	Feeling safe and in control of your life rather than feeling uncertain and threatened by your circumstances.
Pleasure	Feeling that you get plenty of enjoyment and pleasure rather than feeling bored and understimulated by life.
Meaning	Feeling that you are developing your best potentials and making life meaningful rather than feeling stagnant and that life does not have much meaning.
Popularity	Feeling that you are liked, respected and have influence over others rather than feeling like a person whose advice or opinion nobody is interested in.

Each card is composed of:

- a title,
- a definition of the need (adapted from Sheldon et al., 2001),
- linked terms (synonyms or keywords),
- real-life examples of elements or situations able to trigger the fulfilment of the need,
- main scientific references related to the need (on the back side of each card).

Five pictures representing each need were also included on each UX card to enhance visual design and attractiveness, as well as to support the use of the cards by triggering a special context. Moreover, pictures play an important role in the ideation process (Chen et al., 2011). According to Hornecker (2010), images have to be evocative, to provide inspiration while leaving space for interpretation.
Pictures were chosen by the main author and validated through a short user test in order to ensure that each set of pictures represented on a UX Card really triggers the idea of the specific need. For each set of pictures, we asked seven participants to think about the first five words that came into their mind when looking at the pictures. Pictures sets were validated only if amongst those five words, the title or related terms representing the need was cited. We selected the examples on the cards to cover several aspects or expressions of each need. They are based on a review of scientific literature – mainly from the field of psychology - on each need. However, the examples are not exhaustive and meant to support illustrating each need rather than defining or completely describing it.
Two senior UX researchers (both with a Psychology background) reviewed the UX cards for face validity. Then, we asked a small sample of users to assess the understandability of the content and implemented some minor adaptations based on their feedbacks.

An example of the UX Card "Security / Control" is presented in Figure 1.

Figure 1. Example of the UX Card "Control / Security"

The UX Cards are presented along with a generic instruction card reminding the experts that UX is holistic and that their evaluation should not focus on aspects of the interface solely. This instruction card also provides guidance by suggesting a list of 12 elements that could be used during the evaluation: usability, visual design, brand/marketing, features, interface design, content, accessibility, interoperability, technical support, service experience, information design, and interaction design. The purpose is not to be exhaustive regarding the elements that could be identified through UX assessment, but rather to ensure that experts will have a broad understanding of UX before starting to assess a system or product.

A more specific instruction card serves as a basis for UX evaluation and two others as a basis for UX design (cf. section 5).

3.2. UX CARDS PURPOSE

The UX Cards were designed to support three main purposes:
(a) the design of UX,
(b) the evaluation of UX,
(c) the training of UX designers and researchers.

3.2.1. UX CARDS FOR UX DESIGN

In accordance with Stolterman's belief that "designers can be prepared-for-action but not guided-in-action" (Stolterman, 2008), UX Cards are not prescriptive; they rather prepare for action and are a repertoire of concepts to be used as inspiration in future design situations. The knowledge on UX needs provided through the UX Cards does not predefine the interaction design process in any way, but it can be used as inspirational tool to generate design ideas.
The generation of design ideas might be done by the design team alone or might involve users in participatory design sessions. Being summarized and understandable representations of a UX theory, UX Cards might also be used as a tool for communication with the different stakeholders involved in a project. They might help the design team communicating the concept underlying a specific project. Finally, the fact that they represent a theory of UX might be a mean to convince a client or project manager of the benefits and credibility of generated ideas.

3.2.2. UX CARDS FOR UX EVALUATION

While card-based tools are generally employed in design-related activities, their use as evaluation tools is less frequent. Nevertheless, some authors have shown that cards tools – and their underlying frameworks – might provide anchor points for evaluators to reflect during expert evaluations (Lucero et al., 2013).
In the case of UX evaluation, expert evaluators might use our UX Cards on their own or in presence of potential end-users. Regarding their use by experts, UX Cards should not be considered as heuristics. They provide experts with some knowledge about basic psychological needs that should be fulfilled to produce a positive UX. However, they do not encompass a comprehensive list of dimensions and sub-dimensions to check against

when verifying if an interface complies with these guidelines. The goal of the UX Cards is not to "debug" a system but to assess how well it might support the fulfilment of human needs with regard to triggering a positive experience.
The UX Cards might be used to conduct a UX evaluation at any stage in the design process and do not regard fully marketed products only. Ideally, as applies to any evaluation technique used in an iterative design process, a UX evaluation should occur throughout the design life cycle, with the results of the evaluation feeding back into modifications to the design (Dix, Finlay, Abowd, & Beale, 2004).

3.2.3. UX CARDS FOR UX TRAINING

Finally, UX cards might also be used for training purposes, as it supports the discovery of relevant psychological basic needs in a pragmatic way. Similarly to Desmet's idea (2012) of using his Positive Emotional Granularity Cards (Yoon, Desmet, & Pohlmeyer, 2013) to support the development of designers' emotional granularity (i.e., the ability to characterize one's emotional state with specificity), the UX Cards could support the development of UX designers' sensitivity for psychological theories. As recommended by Stolterman (2008), student designers could therefore develop "a useful repertoire of design ideas or concepts to be used in future design situations".
While this topic will not be discussed in the present paper, it would be important to assess the quality of the UX cards as a training tool for future UX professionals. Case studies involving design students and the processes underlying design training are common in design research. With regards to card-based design tools, Kramer et al. (2014) have studied the use of their Design Heuristics in a cross-disciplinary design course. In an upcoming study, we intend to explore the following research questions: how well do the cards communicate the seven basic human needs? How well are the cards understood and used by students? Is the card-set more effective in providing learning outcomes than a more traditional course about the needs? Beyond assessing the relevance of the UX cards as a training tool, we also aim at using the results of this study to provide pedagogical recommendations on how to use the UX cards in design or HCI classes.

4. VALIDATION OF THE UX CARDS: USE CASE OF UX EVALUATION IN AN EXPERIMENTAL SETUP

In order to assess the potential of the UX cards as an expert evaluation tool, we conducted an experiment involving 33 UX experts (see Paper XX for a detailed description of the methodology and results). This allowed us to evaluate several properties of the UX cards, to reflect on the strengths and weaknesses of this tool in the context of supporting expert evaluations, and to elaborate suggestions for improvement.

4.1. PARTICIPANTS

The profiles of the UX experts composing our sample are described in Table 2. About two thirds of the participants are consultants or

practitioners working in Industry (n = 20) while the other third are researchers or students working in Academia (or between Industry and Academia). Experts were mainly educated in Psychology or Social Sciences (n = 19).

Table 2. Experts' background

Background Variable		Frequency (n = 33)	Valid Percentage
Domain	Industry	20	60.6 %
	Academia	5	15.2 %
	Both or between	8	24.2 %
Role	Researcher	9	27.3 %
	Consultant / Practitioner	20	60.7 %
	Student	4	12.1 %
Education	Design	5	15.2 %
	Psychology / Social Sciences	19	57.6 %
	Cognitive Sciences	3	9.1 %
	Technology / Software	4	12.1 %
	HCI	2	6.1 %

Individual experts' profiles are also detailed in Table 3. We assigned an ID to each participant in order to quote experts' comments in this paper.

Table 3. Experts' profiles and interview IDs

Interview # ID	Gender	Age	Role	Education	Experience in HCI (years)
1	F	27	Researcher	Psychology	4
2	F	27	Researcher	Psychology	2
3	F	25	Consultant	Psychology	1
4	F	34	Researcher	HCI	3
5	M	40	Practitioner	Design	7
6	F	31	Consultant	Psychology	4
7	M	29	Practitioner	Psychology	4
8	M	41	Consultant	Psychology	15
9	M	27	Consultant	HCI	1
10	F	26	Consultant	Psychology	1
11	F	26	Consultant	Cog. Sciences	2
12	M	27	Student	Psychology	3
13	M	30	Consultant	Psychology	3
14	F	28	Researcher	Psychology	4
15	M	24	Student	Psychology	1
16	F	32	Researcher	Technology	7
17	M	26	Consultant	HCI	1
18	M	32	Researcher	Psychology	3
19	M	27	Practitioner	Cog. Sciences	1
20	F	33	Practitioner	Design	8
21	M	23	Consultant	Cog. Sciences	1
22	M	31	Practitioner	Design	10
23	F	25	Student	Medicine	1
24	M	31	Consultant	Technology	5
25	F	43	Manager	Psychology	19
26	F	41	Practitioner	HCI	14
27	F	35	Consultant	Design	13
28	M	29	Consultant	Psychology	3
29	M	27	Researcher	Design	3
30	F	42	Consultant	Medicine	10
31	M	26	Researcher	Psychology	2
32	F	40	Manager	Psychology	14
33	M	40	Consultant	Technology	14

4.2. PROCEDURE

During individual 2-hours evaluation sessions, we asked the experts to evaluate the UX of four interactive systems, using the UX cards. Their assessment was therefore based on a psychological needs-driven theory of UX. The four use cases were similar for each expert and the duration of the task (15 minutes per system) was standardized. Experts were first familiarized with the UX cards by reading them with the author.

The four interactive systems inspected during the experiment were:

- the game Angry Birds on IPhone 5S
- the online e-commerce website Amazon
- the social network Facebook
- an Olympus digital camera (SZ-16 compact model).

We chose four strictly different examples of interactive systems in order to maximize the diversity of HCI elements and of potential need fulfilment coverage.

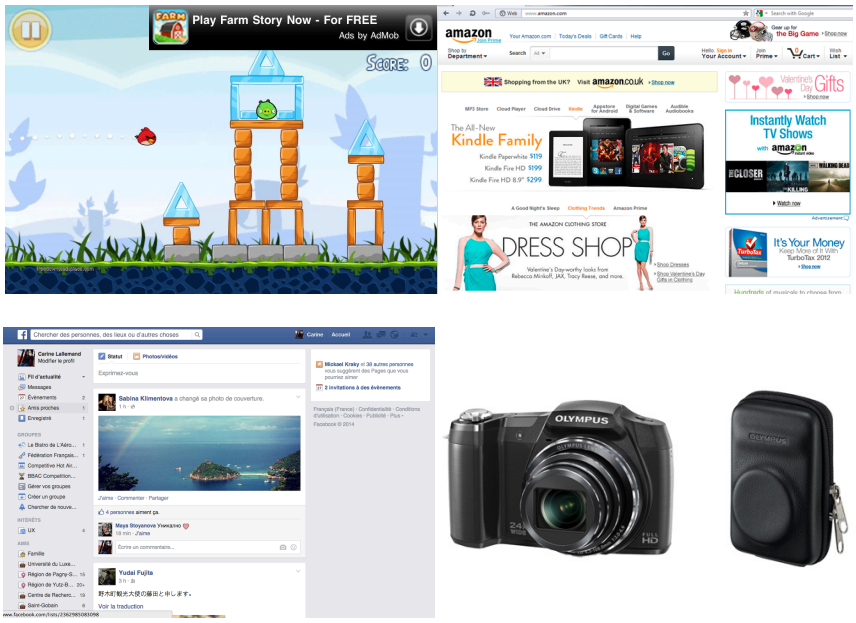


Figure 2. The four interactive systems used during our experiment. From left to right: (1st line) Angry Birds, Amazon (2nd line) Facebook, Olympus digital camera.

We provided participants with a paper-based grid to report UX elements while conducting their UX evaluation (Table 4). The grid was composed of three columns: identified element, UX need(s) positively impacted by this element and UX need(s) negatively impacted by this element. To ensure that all instructions were clearly understood before starting the evaluation task, the experimenter went through a first example (using Apple.com website) by showing the participant how to report elements and related needs in the assessment.

Table 4. Paper-based grid reporting tool

Identified elements	UX need(s) fulfilled thanks to this element	UX need(s) hindered by this element
Ex: camera visual design	*Pleasure (aesthetics)* *Security (doesn't look fragile)*	
Ex: absence of tactile screen		*Pleasure* *Effectiveness*
…		

The understandability and operationalizability (i.e., imagined ease of using the cards in the context of a UX evaluation) of the UX cards have been tested before the evaluation task, whereas the perceived usefulness of the cards has been evaluated after each 2 hours expert evaluation session. Quantitative data on the use of the cards was also collected during the experiment. Finally, we conducted semi-structured interviews to better understand how the experts perceived the cards.

4.3. RESULTS

4.3.1. UNDERSTANDABILITY OF THE UX CARDS

Before conducting a UX expert evaluation, we read each card and asked each participant to rate on 7-points scales the level of understandability of the cards and the imagined difficulty to use the card in the context of a UX evaluation (anticipated operationalizability). It is worth noting that a wide majority of experts in our sample was not aware of the psychological needs-driven approach linking basic human needs to UX.

Before starting the task, participants assessed the overall understandability of the UX cards as very good with an average score of 6.35 (SD = 0.54) and a minimum score (yet fully acceptable) of 5.97 for the card "Competence/Effectiveness". Anticipated operationalizability of the cards (i.e., imagined ease of using the cards in the context of a UX evaluation) was assessed as satisfactory (M = 5.78, SD = 0.65). However, the need for Self-Actualizing was assessed as much harder to operationalize than the others (M = 3.82, SD = 1.8). Understandability and operationalizability ratings for each card are presented in Table 5.

Table 5. Understandability and anticipated operationalizability of the UX Cards (N = 33)

UX Cards	Understandability				Operationalizability			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Security / Control	5	7	6.64	.60	5	7	6.36	.78
Influence / Popularity	5	7	6.58	.61	3	7	6.36	.93
Relatedness/Belongingness	5	7	6.42	.61	4	7	6.15	.83
Autonomy / Independence	5	7	6.39	.70	3	7	6.09	1
Pleasure / Stimulation	3	7	6.39	.93	4	7	6.15	.94
Self-Actualizing / Meaning	3	7	6.03	.92	1	7	3.82	1,8
Competence/Effectiveness	3	7	5.97	1.1	2	7	5.55	1.6

Background variables (age, gender, seniority, level of familiarity with UX or level of familiarity with psychological needs theory) do not significantly impact the assessment of understandability or operationalizability of the cards.

4.3.2. USE OF THE UX CARDS

Overall, the experts identified 1794 elements, which corresponds to an average of 54.4 elements per expert and 13.6 elements per assessed system. Experts linked these identified elements to a total of 3455 UX needs. 2277 needs were cited as positive (66%) and 1179 needs cited as negative (34%). Experts were more focused on interactive elements able to fulfill UX needs than on elements having a negative impact on needs. This is compliant with the common statement that "UX focuses on positive aspects of users' interaction with interactive products" (Bargas-Avila & Hornbæk, 2011, p. 2690).

Results show a significant order effect impacting the number of needs cited for each system: experts cited less UX needs during the evaluation of the first system than for the next three systems (*diff* = -3.1, $t(32) = 2.12$, $p < .05$) (Figure 3). As all experts were using the UX cards for the first time, the evaluation of the first system acted as a learning phase, where the performance was not fully optimal. After the first evaluation however, we observed acceleration in the evaluation speed and a stabilization of the number of needs cited. The duration of the first evaluation was 15 minutes; it thus seems that the appropriation time required for the UX Cards is relatively short.

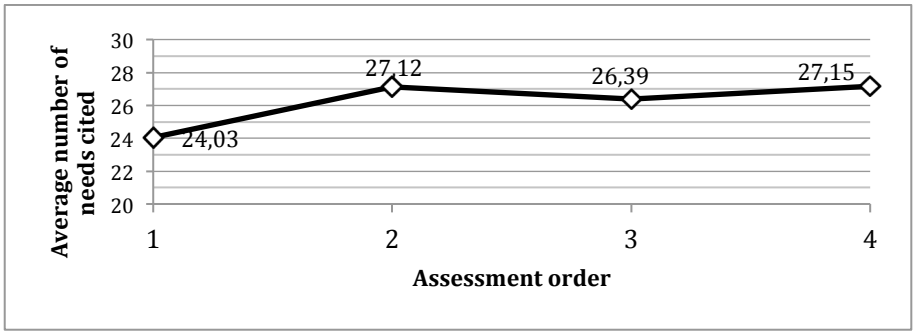


Figure 3. Average number of cited needs according to the order of the evaluation

As interactive systems do not equally cover the fulfilment of all needs, some cards were used more often than others. Most cited needs (Figure 4) were Security (22%, 771 citations) and Pleasure (23%, 784 citations), while least cited needs were Influence (8%, 266 citations) and Self-Actualizing (6%, 211 citations).

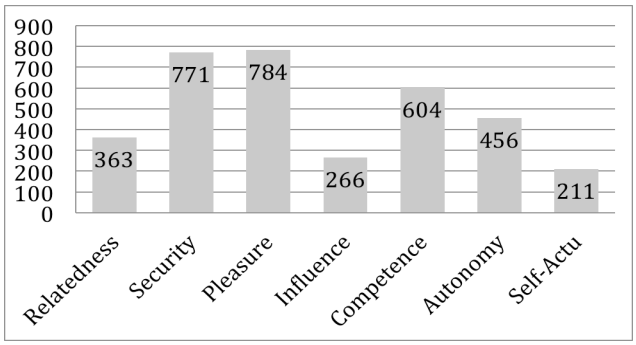


Fig 4. Total number of cited needs (considering both positive and negative) during the UX evaluation task

Nevertheless, except for expert #25 who never used the self-actualizing card, each expert used each card at least once during the evaluation. On average, experts used from 6 (for the digital camera use case) to 6.73 (in the case of Facebook) cards out of seven during each evaluation.

Our results on the use of the cards suggest that UX experts encountered no blocking issues in conducting an UX expert evaluation by using the UX Cards.

4.3.3. PERCEIVED USEFULNESS OF THE UX CARDS

Perceived usefulness of the UX Cards was assessed at the end of the evaluation session using four 7-points Likert scales ranging from 1 (not useful) to 7 (highly useful). Participants found the UX Cards highly useful for both practitioners (M = 6.45, SD = 1) and researchers (M = 5.91, SD = 1.59). Similarly, they believed the UX Cards to be potentially useful for both the design (M = 6.15, SD = 1.37) and the evaluation of interactive systems (M = 6.55, SD =0.62). No significant differences were found between the perceived usefulness of the UX Cards with regard to background variables.

4.3.4. WHICH CARDS ARE THE MOST DIFFICULT / EASY TO USE?

During the interview, we asked each expert to cite the cards that were the easiest to use (a maximum of 3 cards could be cited) and the cards that were the most difficult to use (Table 6). The cards "Pleasure" and "Security" were globally cited as easy to use, whereas the card "Self-Actualizing" and "Autonomy" were cited as harder to use.

Table 6. Cards cited as the most difficult or the most easy to use (N = 33)

UX Cards	Cited as difficult to use	Cited as easy to use
Relatedness	7	13
Security	6	**24**
Pleasure	3	**27**
Competence	6	13
Autonomy	**13**	4
Influence	10	6
Self-Actualizing	**30**	1

The experts explained that it was harder to envision how an interactive system could provide users with a feeling of self-actualizing, as this feeling is both more abstract and harder to attain, even in real-life. On the contrary, identifying elements related to the need of security was assessed as easy because this concern is already included in usability evaluation and therefore refers to something that UX experts are used to assessing. These results are in line with quantitative data on the evaluation task, which shows that most cited needs were Security (22%, 771 citations) and Pleasure (23%, 784 citations), while least cited needs were Influence (8%, 266 citations) and Self-Actualizing (6%, 211 citations).

We also noticed that some cards were more ambivalent, being considered by some experts as difficult to use and by some others as easy to use. This is for instance the case for the "Influence" card, which was cited 10 times as one of the hardest cards to use and 6 times as one of the easiest cards to

use. Participant #6 explained that she had trouble distinguishing between the needs fulfilments of systems that make the users feel influent versus those having an influence on the user.
Finally, some difficulties were reported during the identification task, especially on the way to categorize the impact of some interactive elements on the felt experience. Expert #13 wondered for instance, "*Angry Birds is mind-numbing, but does it have a negative impact on the need for competence or the need for self-actualizing, or maybe both?*" Experts generally solved these issues by reading each card again, and then relied on their own appreciation.
The links between the cards were also a matter of concern. The pleasure was also often described as a transversal need that could be derived from the fulfilment of other needs. "*I wonder whether the cards might be used in combination: for instance pleasure might be derived from the fulfilment of any other need*" (Expert #32). Other needs were also linked: "*Might a lack of control result in a negative autonomy feeling?*" (Expert #28).

4.3.5. ELEMENTS USED DURING THE EVALUATION

During the interview, we also asked each expert to explain which elements on the card were the most useful during the evaluation task and which ones were the least useful (Table 7).

Table 7. Usefulness of the elements (N = 33)

UX cards elements	Cited as one of the most useful element	Cited as one of the least useful element
Title	**33**	0
Definition	1	**27**
Examples	**25**	3
Keywords	9	16
Pictures	6	**21**

Experts used card's title and the real-life examples as main supports for their evaluation. Surprisingly, definitions were cited as one of the least useful elements during the evaluation. However, the experts explained that definitions were only useful to clearly understand the meaning of each need, when they discovered the cards for the first time. After that, they only relied on the title to conduct their evaluation (some experts reported having read some definitions a second time when they felt unsure about the meaning of a need during the task). Expert #11 explained, "*the definitions were useful at the beginning when I discovered the cards. Once I was sure to understand each need correctly, I never read them again*".

Pictures had an ambivalent status. While they were not assessed as very useful to perform the task, the experts highlighted their importance with regards to the visual appeal of the cards. "*I haven't used the pictures during the task. However, I truly feel that this is an important element on the cards because the images trigger a specific impression and also make the cards much more attractive than they would be with only textual information*" (Expert #7). "*I think the pictures also act as a visual confirmation to check if the concept was clearly understood. It is reassuring.*" (Expert #28) These results are consistent with Chen et al.'s findings (2011) showing that bigger pictures on their IoT cards attracted and stimulated their participants more than textual information.

Finally, the bibliographical references (at the back of each card) were not assessed here, as they were not used during the evaluation task. However, while they do not intend to use it regularly, a majority of experts found this information useful. "*The references will not serve on a daily basis, that's sure! However, it helps to understand what's behind each need and could constitute good readings to enhance my professional skills. But most importantly, I could use the references to argue for the solutions I propose to the client. Very often I witness long debates about why a solution is good or not and whether it is worth implementing it. Here the solutions will emerge out of a sound theoretical framework, and it will support the credibility and acceptance of the proposal"* (Expert #11).

4.3.6. PARTICIPANTS' FEEDBACK ON THE CARDS AND HOW TO IMPROVE THEM

During the interview, our participants all showed a strong interest for the UX cards and agreed on the fact that this seems a promising approach to support an expert assessment of UX.

> "*The UX Cards are an intelligible way to talk about UX, the needs are clearly presented and the card-set constitutes a good framework to support UX design or evaluation*" (Expert #31)
> "*The cards are good notes to support the acquisition and use of the embedded theoretical knowledge*" (Expert #11)
> "*The cards are easy to understand and pleasant to use. They help me to structure my evaluation*" (Expert #12)
> "*This is a powerful tool to foster creativity*" (Expert #22)

The ease of using the cards was also highlighted as an important point:

> "*The cards are great because we don't need a 3-days training to use them*" (Expert #31)
> "*The UX Cards are a good support for junior experts or to conduct a participatory evaluation of UX*." (Expert #8)
> "*It was hard during the first ten minutes. Then I truly felt that I was able to use the cards effectively*" (Expert #28)

And the presentation of the UX needs approach under the form of a card-set was appreciated:

> "*I like them, it's more appealing and inventive than a checklist*" (Expert #7)
> "*The format of the cards is pleasant*" (Expert #10)
> "*The format is practical and easy to manipulate*" (Expert #11)

Beyond expressing their general feeling about the tool, experts also provided numerous comments and suggestions, especially on how they would improve the cards.

4.3.6.1. DESIGN AND CONTENT OF THE CARDS

Experts' opinions regarding the design and content of the cards were diverse. At a global level, several participants (#4, #11, #16, #22, #29) wished we had designed each card using a distinct color: "*Each card should have a different color in order to differentiate them easily and to categorize the results according to the colors*" (Expert #11).

The type, size and location of pictures were often commented. Globally described as useful and attractive, our participants essentially regretted that the pictures were not used in a powerful, immersive or emotional manner. Three participants also mentioned the fact that some pictures did not fit well to the content of the card.

> "*The images are too small, we cannot immerse into the feeling that they trigger*" (Expert #11)
>
> "*The pictures are interesting, even if I admit that I haven't used them during the evaluation task*" (Expert #7)
>
> "*I think there are too many pictures. I would prefer a single image with a strong emotional impact.*" (Expert #14)
>
> "*Some pictures do not add anything to the card, this is the case for instance for the card Security where the pictures are too banal*" (Expert #8)
>
> "*I have the feeling that images could be chosen more carefully. Some of them do not truly reflect the content of the card*"(Expert #16)

As the content of each card should be rapidly and unequivocally understood by the participants (Halskov & Dalsgaard, 2007), we might think about an alternative design of the UX cards involving for instance a single bigger picture (Figure 5). It would also allow to suppress the most problematic pictures that were described as "too banal" or "not truly reflecting the content of the card". This of course requires additional studies on how well that picture covers the intended concept.

The usefulness of the keywords was rarely questioned and several experts suggested making them more visible: "*The keywords should be more visible, with a bigger font size*" (Expert #12).

Experts also made suggestions on how to use both sides of the cards:

> "*I suggest presenting the examples and definition at the back of each card and only having the title and keywords on the front side*" (Expert #9)
>
> "*As the definitions are only used the first time, one should put them on the back side. Additional space on the front side could be used to add a bigger image*" (Expert #20)

Figure 5. Alternative design of the UX Card "Relatedness" based on experts' suggestions

A first source of disagreement between experts was related to the examples. Each UX card entails examples (from 5 to 8) of how the need might be fulfilled in the real life, or what kind of feeling this need might trigger. These examples are rather generic and do not relate to specific examples in the field of HCI. Our purpose was to keep the tool at a generic level, open for interpretation as well as for use in another domain of application; the instructions cards being however specific to UX design and evaluation. The question raised by our participants was therefore: should the UX cards be more explicitly oriented towards the evaluation of UX, by encompassing examples related to interactive systems? At first sight, having HCI examples seemed attractive to our experts, but after giving careful thought to it, they generally agreed that this would not be a suitable approach:

> "*One should have access to more HCI related examples. In their current form, the cards are too psychological*" (Expert #27)
> "*I don't think you should include HCI examples because the cards would then become very specific and this is likely to impede the appropriation of the tool by each expert. Moreover, the examples will quickly become obsolete due to fast technological evolutions. In their actual form, the cards might be used by designers, engineers, or even marketers*" (Expert #12)
> "*While I would not add HCI examples to the UX cards, one could think of an online database where practitioners would share their examples*" (Expert #17)

Another source of disagreement relates to the level of guidance provided by the cards. Some experts tried to compare the cards to usability heuristics and therefore suggested reformulating the examples under the form of questions to be answered by the evaluator. Still, a majority of our participants recognized that the interest of the tool lies in its openness.

> "*One could replace the examples provided on the cards by questions. It could add concreteness and facilitate the evaluation process*" (Expert #10)
> "*When I read a card, I have the impression that each identified element on the system should fulfill all examples given on the card. Maybe one should create more cards by dividing the UX needs into sub-needs*" (Expert #18)
> "*What is a bit confusing about the cards is that there are several levels of interpretation of each need, from a micro-level (e.g., personalization as a way to create autonomy) to a macro-level (e.g., being the cause of my action)*" (Expert #12)
> "*I don't think we should distinguish needs and sub-needs because then the cards would look like heuristics and evaluators would feel the need to check every single example. The cards are only an idea in which direction you could go to attain the fulfillment of a specific need.*" (Expert #20)

Finally, many experts mentioned the fact that some cards are conceptually related, and that we could use this distance between the cards to orient the design of the cards.

> "*We could study the conceptual distance between the cards. I guess that some needs are closer to each other.*" (Expert #13)
> "*It would be interesting to map the needs according to their proximity. We would have a cartography of needs along with their relationships.*" (Expert #28)

4.3.6.2. INSTRUCTIONS

Regarding the instructions, some experts expressed their wish to have more guidance for the process of UX evaluation, perhaps under the form of an additional instruction card (Expert #32) or a website (Expert #18).

> "*One could add a link to a website providing additional guidance and the publications related to the UX cards*" (Expert #18)

Experts' highest concern was to be able to adopt the users' perspective and to forget about their own experience. Some experts reported during the task that they felt as having identified elements that disturb their own user experience, but might not be disturbing for a typical user (Expert #25).

> "*There is a risk that experts using the cards will evaluate the UX of a system according to their own experience*" (Expert #32)
>
> "*There is an issue regarding the objectivity of the assessment. The cards do not really help in adopting the users' perspective*" (Expert #11)

A way to deal with this issue could be to combine the UX Cards with methods providing contextual information and supporting empathy for users.

> "*The cards should be used in combination with other methods providing contextual information, such as personas*" (Expert #17)

Another comment regarding the instructions was to make evaluators think about the experience as something holistic, and to help them in conducting a comprehensive UX evaluation. For example, one could remind experts to think about the market and concurrent systems or products as well, in order to know what elements could be missing (and therefore could lead to users' disappointment).

> "*The cards should perhaps provide step-by-step instructions so that we don't forget anything in our evaluations. I think I forgot some important things, because I was too focused on details and not on the experience as a whole.*" (Expert #29)
>
> "*After having finished the evaluation task, I realized I forgot a lot of relevant things. But this frustration is related to the experiment. I mean, in my daily job I would just go back to the grid and add additional info*" (Expert #30)

4.3.6.3. REPORTING TOOL

Beyond the card-set, experts also made suggestions on how to improve the reporting tool; if it was to be provided with the card set. In the current form (see Table 4), the reporting grid was described as lacking guidance, which was convenient for the experiment but might be seen as not practical enough for a use in real conditions. The advantages of an online reporting tool were also highlighted.

> "*The reporting of the results is very open, which is good, however, it might lack a bit of organization and I am afraid that the analysis of the results will be time-consuming*" (Expert #11)
>
> "*I prefer an online reporting tool. During the task, I would thus be able to copy-paste some screenshots to illustrate my evaluation.*" (Expert #17)
>
> "*One should be attentive to the identification of false positive (i.e., problems identified by experts but not problematic for users). The reporting tool should not look like a heuristic checklist*" (Expert #21).

"*The grid can be rearranged to facilitate the counting of observations. Each of the seven needs could be presented as one column, so that one just has to rate each element using + or – signs*" (Expert #12) (Table 8)

Table 8. Adapted reporting grid following the suggestions of Expert #12

Identified elements	**Security**	**Autonomy**	**Relatedness**	**Influence**	**Pleasure**	**Competence**	**Self-Actu.**
Ex: camera visual design	**(+)** *not fragile*				**(+)** *aesthetics*		
Ex: absence of tactile screen					(–) *expected as standard feature*	(–) *less effective*	
…							
Total positive	1				1		
Total negative					1	1	

Some experts pointed out small elements to ease the reporting process:

"*One should add a code on each card to report the results more easily*" (Expert #13)

Two main concerns regarding the reporting tool were highlighted. The first was related to the impossibility to assign different weights to the reported observations, resulting in a lack of differentiation between small details and important information.

"*I think I would personally add a score of importance or criticality for each identified element*" (Expert #16)

"*During the task, I identified some elements and reported them on the grid. However, I thought that some of them were less relevant than others*" (Expert #2)

The second concern was linked to the analysis and presentation of the results.

"*I now wonder how I could present the results of my evaluation to a client*" (Expert #11)

"*It would have been great to provide, along with the card-set, some examples of how the results might be analyzed and presented*" (Expert #6)

Finally, expert #12 made a mischievous suggestion for improving of the cards: "*What about using the cards to enhance the cards themselves?*"

5. PERSPECTIVES

5.1. UX EVALUATION

Based on the results collected during this first experiment on UX expert evaluation, we foresee several concerns to be addressed in future studies and propose suggestions for improvement.

5.1.1. PROVIDING MORE GUIDANCE

During the evaluation task, we noticed that some experts felt a bit lost because of their complete freedom / lack of constraints for the evaluation. This choice was made in order to see how the experts would use the UX cards on their own, without a lot of instructions. Most of them felt somewhat uncomfortable deciding what kind of elements they should identify and we could observe that they often described on their report grid the same type of elements: features, contents, or usability-related elements seemed to be the more obvious ones. Despite the presence of a specific instruction card listing a lot of elements that could be identified, experts seem to rely on their own evaluation routines and to adopt a restricted view of UX. Moreover, in some cases, apparently important UX elements were forgotten during the evaluation task. The experts often recalled these elements later on, during the overall UX evaluation. As an example, some experts did not mention in their report any relatedness elements supported by the digital camera. However, when assessing the overall impact of the digital camera on the need for relatedness, they suddenly remembered that taking pictures of family or friends could have a positive impact on relatedness.

Based on these observations, we believe that UX experts need more guidance to using the UX Cards for an evaluation purpose. We propose a 4-steps guidance to support experts in providing a thorough and overall UX assessment, which would not be limited to pragmatic elements or category of elements only:
(1) Experts will be advised to first think about the UX of the assessed system at a very generic level (e.g., concept, brand, associations, visual design)
(2) Second, they will be instructed to assess the system from a functional perspective (e.g., features, interoperability, interaction design)
(3) Third, the evaluation should focus on detailed user interface elements (e.g., content, information design, usability issues)
(4) Finally, we will invite the experts to reflect on missing elements, which would be required to provide the desired UX or to satisfy user expectations.

This optional guidance could help UX experts to improve the accuracy and coverage of their evaluation and also support the standardization of their practice in case of multiple experts assessing the same system or product. This could therefore reduce the evaluator effect, described by Hertzum and Jacobsen (2003) in the context of usability evaluation as the fact that multiple evaluators assessing the same interface with the same method detect different sets of problems (p. 183). Future research work will focus on the effectiveness of these evaluation guidelines to support UX expert evaluation using the UX Cards.

5.1.2. SUPPORTING A HOLISTIC AND EMPATHIC APPROACH

Another observation made during the experiment is that experts were less used to assessing subjective and hedonic aspects of UX and therefore mentioned more pragmatic aspects in their report. It seems that making an informed guess of what users are likely to feel during an experience seems

harder than estimating users' likelihood to succeed or fail in performing a task. On this point, experts also mentioned that the use of the second person singular in the UX Cards (e.g., "feeling that you get plenty of enjoyment…") disturbed them in adopting an expert perspective rather than a user perspective. We therefore suggest reformulating the UX Cards using the third person singular, referring to "users".

Moreover, as we wanted to stay at a generic level, we had asked the experts to evaluate how each identified element would impact the UX of a "regular user" or of "the majority of users". However, research shows that UX is unique to an individual and influenced by several individual and contextual factors. Hornbæk (2010) criticized studies on usability evaluation methods for considering usability issues as stable, independent from circumstances and users. The same way we cannot claim for a stable number of usability problems existing in an interface, we cannot consider that a stable number of elements will impact UX needs or that this impact will be the same for any user involved in any context of use. As mentioned by Expert #13, "*some features are either negative or positive depending on the target user, the temporality or the context of use*". To help experts adopting an end-user perspective, we therefore suggest combining the use of UX Cards with methods providing contextual information, such as scenarios of use or personas (Pruitt & Adlin, 2010). Moreover, despite the fact that the systems assessed here were general use products, we also expect that a domain/application expert might be required to assist the UX expert in the case of business-specific systems requiring a deep understanding of users' tasks and objectives.
Another way of reducing personal biases during expert evaluation would be to have several experts conducting the evaluation together. Following the experimental setup implemented by Lucero et al. to test the relevance of the PLEX Cards for expert evaluation (2013), this co-evaluation using the UX Cards could involve three major steps: first, the experts will interact with the assessed system (either individually or together), then they will pick the UX Cards one by one and discuss whether each kind of experience is triggered or hindered by the system, and how (e.g., what kind of elements or dynamics support the experience), finally they will produce the UX report by rating each of the seven needs, highlighting positive and negative aspects of the system, and suggesting design alternatives to improve the system. Whenever possible, an additional step should be included to the expert co-evaluation process: the observation of users interacting with the system.

Finally, the cards could also be used during creativity or evaluation workshops involving final users. "*I can imagine using the cards during meetings with project's stakeholders that are not experts in UX, or with final users to help them express their expectations about the product*" (Expert #13). In this case, the UX cards will serve as a basis to discuss what kind of experience is triggered or hindered by the system under study.

5.1.3. SUPPORTING A PRAGMATIC REPORTING OF UX EVALUATION RESULTS

While the reporting tool might not be considered as part of the card-set, it seems relevant for a UX evaluation task to take into account how the results will be reported. We therefore intend to explore several alternatives to

support a pragmatic and attractive reporting of UX evaluation results. This is nonetheless a challenging topic, since our goal is to keep the open-minded nature of the cards. The reporting tool should not constrain the evaluation. A relevant idea could be to compute a summary of the information provided by the expert during his/her evaluation. Using this summary, the expert would be able to select the information regarded as relevant and to create graphical representations of the results. The tool will therefore require, on purpose, the experts' input and validation before presenting the results.

As a conclusion, we saw through this first experimental study that the use of the UX cards for UX expert evaluation seems to be a relevant approach. The UX card-set guided the UX expert evaluations of several kinds of interactive systems effectively and was described by our expert participants as easy and enjoyable to use. As a tool, the UX cards comply with several of the requirements defined by Väänänen-Vainio-Mattila et al. (2008) for practical UX evaluation methods. It is applicable for various types of products and suitable for different product lifecycle phases. The level of expertise required to master the tool is low. It is a fast, lightweight, and cost-efficient approach. Further research is nevertheless needed to investigate the validity and reliability of our approach. Our next step in this project will be the enhancement of the UX card-set and associated method, before conducting additional experiments involving UX experts.
The second part of this paper is focused on the use of UX cards for UX design.

5.2. UX DESIGN

The UX cards were used as an ideation technique during several design workshops involving UX practitioners and UX students from 2013 to 2014. We used the observations made during these workshops to iteratively improve and refine the tool (especially the instructions cards). These informal experiments were fruitful for the design of the cards and led to the development of two additional techniques supporting the ideation process.

5.2.1. TECHNIQUES SUPPORTING THE GENERATION OF IDEAS WHILE USING THE UX CARDS

To accompany the use of the UX cards for UX design, we created two instructions cards describing brainstorming techniques aimed at supporting the triggering of ideas and outside-the-box thinking processes. This idea is not novel and some other cards-sets already propose scenarios to support the generation of ideas using the cards. This is for instance the case for the PLEX Cards (Lucero & Arrasvuori, 2010), which include two related idea generation techniques called PLEX Brainstorming ("for many ideas, fast") and PLEX scenarios ("for more elaborate ideas") (see Lucero and Arrasvuori, 2010 for more details).

The first technique that we propose for the UX Cards in UX design is a 2-steps brainstorming technique that we named *COMBINE* (Figure 6). First, designers are instructed to use each relevant card to think freely about as many design ideas as possible related to the need. Next, to extend the scope

of the brainstormed ideas or to elaborate further on generated ideas, the instruction card "UX is holistic" will be used. This card entails 12 categories of elements that are part of a user experience with an interactive system. Designers will be instructed to combine the needs cards with the element card in order to develop more ideas. For instance, they will have to generate as many ideas as possible on how the element "visual design" might be able to support the fulfilment of the need for "relatedness". Then how the element "features" might be able to support this need, and so on for each of the twelve mentioned elements.

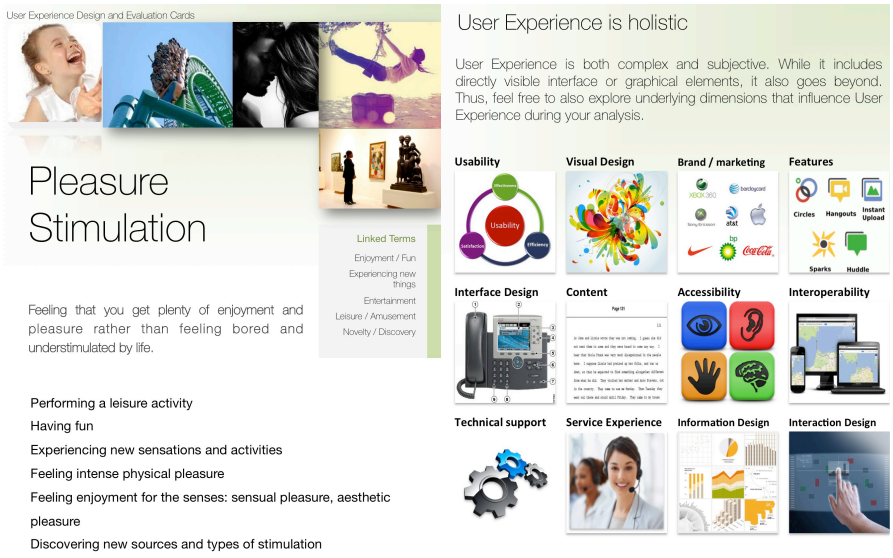


Figure 6. COMBINE idea generation technique where designers generate ideas by using both the UX need card and the elements described on the "UX is holistic" instruction card

The second technique that we propose for the UX Cards in UX design is called *ANALOGY*. Based on a selected UX card and the existing (however non exhaustive) list of real-life examples, designers will be instructed to think about everything in the real-life that impacts positively or negatively the fulfilment of a need. For instance, they will think about as many as possible situations or objects that trigger a feeling of *security* or on the contrary create a feeling of *insecurity*. If personas (Pruitt & Adlin, 2010) are used in the design project, they might use the personas to think empathetically about what triggers a specific feeling for the end users. Then, they will be instructed to think by analogy (Blanchette & Dunbar, 2000; Visser, 1996) about how they might transpose these specific situations or object characteristics into their design to trigger the same experience. In the case of real-life situations impacting negatively the fulfilment of a need (e.g., triggering a feeling of insecurity), they will try to see how to solve this kind of problem through the designed system. Design by analogy has been described as an effective ideation approach to generate high quality novel ideas and reduce design fixations (Moreno et al., 2014).

We can illustrate this design thinking process through the example of the Philips Glo Nightlight (Figure 7). This object is a colour-changing nightlight with portable glowing balls. What kind of UX needs might the nightlight fulfil? Here is the story: final users of this lamp are children and children are

feeling especially insecure during night. They are often afraid of the dark, and therefore often afraid of moving from their bed to the toilet. To solve this security issue, the light offers portable glowing balls that can be transported by the kid in order to feel reassured while crossing the dark corridor. We can say that this lamp provides its users with a positive reassuring UX and it has been designed to solve a real-life issue regarding one of the fundamental human needs.

Figure 7. Philips Glo Nightlight fulfils children's need for security by providing portable glowing balls

We will use metrics for measuring ideation effectiveness described in the design literature (Nelson, Wilson, Rosen, & Yen, 2009; Shah, Vargas-Hernandez, & Smith, 2003) to assess if the UX cards and the two aforementioned ideation techniques really facilitate ideation.

5.2.1. EXPLORING USERS PSYCHOLOGICAL NEEDS BEFORE USING THE UX CARDS

Finally, it is worth reminding that the generation phase is not the only crucial design phase. According to Cross (2008), the design process is composed of four phases: Exploration, Generation, Evaluation and Communication. In the case of the UX cards, it is not only useful to *generate* several ideas for each needs, but one should first *explore* what experience we want to design for. An interactive system or product will generally not be designed to fulfill all seven UX needs (as it would be very challenging and would involve heterogeneous design strategies). A first design choice before generating ideas related to a specific need will consequently be to actually select which need(s) have to be fulfilled through the designed product or system. As highlighted by user-centred design (ISO 9241-210, 2010), user research is an essential preliminary step in a design project to investigate what are the main motivations of final users. With regards to psychological needs-driven design, one should therefore try to understand what is the need that is most important to the users when using the kind of product we are designing, in the specific usage situation it is intended to be used.

In order to support the exploration of relevant needs before using the UX cards for idea generation, we designed a simple tool: empathic user stories. The format of these user stories is quite similar to the one adopted when using personas (Pruitt & Adlin, 2010): the picture of a fictive final user

along with a quotation explaining his/her particular motivation of why he/she actually uses a specific kind of system or product. This format was adopted in order to put users into a realistic setting and therefore increasing the potential for identification. User stories are presented on a 2-pages leaflet. Each user story is followed by three questions to be rated on a scale from 1 to 5 (Figure 8).

Tom uses the application because he loves to feel related to his colleagues and peers. It is important to him to feel that he belongs to a community and that he has regular contact with people sharing his interests.

How close do you feel to this person?

Not close at all ☐ ☐ ☐ ☐ ☐ Very close

How important is his/her motivation for you?

Not important ☐ ☐ ☐ ☐ ☐ Very important

How important is his/her motivation for your peers?

Not important ☐ ☐ ☐ ☐ ☐ Very important

Figure 8. Empathic user stories to explore users' main psychological needs before using the UX cards

The use of empathic user stories is followed by a debriefing interview. The analysis of the results will inform the designer on which UX needs are the most relevant from the users' perspective. This will therefore contribute to the formulation of primary design goals.

6. CONCLUSION

In this paper, we introduced the UX Cards as a pragmatic tool able to support psychological needs-driven UX evaluation and design. The cards aim at supporting practitioners in both the generation of ideas and the evaluation of their designs. High levels of perceived usefulness and high amounts of cited elements and needs during the evaluation task evidence the effectiveness of this approach from an expert perspective. In further studies, we intend to experimentally assess the use of the UX cards for UX design by analysing how the cards impact the emergence of design ideas, from both a quantitative and a qualitative viewpoint. Relevance and usefulness of generated ideas will be assessed along several criteria such as their quantity, novelty, variety or quality (Shah, Vargas-Hernandez, & Smith, 2003). Designers' feelings about the UX cards will also be collected through think aloud methods and interviews. Acting as a tangible translation of a UX research framework, the UX Cards contribute to filling the gap between research and practice.

REFERENCES FOR PAPER D

Beck, E., Obrist, M., Bernhaupt, R., & Tscheligi, M. (2008). Instant card technique: how and why to apply in user-centered design. *In Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, pp. 162-165. New York, NY: ACM.

Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory & Cognition*, *28*(1), 108-124.

Bleuzé, T., Ciocci, C., Detand, J., & De Baets, P. (2014). Engineering meets creativity: a study on a creative tool to design new connections. *International journal of design creativity and innovation*, *2*(3), 1–21.

Chen, Y.-H., Liang, R.-H., & Chiang, C.-D. (2011). IOT Deck: A Digital Card-Based Ideation Game to Inspire Internet of Things Design. In Proceedings of the 4th World Conference on Design Research (IASDR2011), Netherlands, Oct. 31 – Nov. 4, 2011.

Cross, N. (2008). Engineering Design Methods. Chichester: John Wiley & Sons Ltd.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology, 56*, 815-822.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.

Desmet, P. M. A. (2012). Faces of product pleasure: 25 positive emotions in human-product interactions. *International Journal of Design, 6*(2), 1-29.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (2004). *Human-Computer Interaction*. Prentice Hall.

Eco Innovators (2013). Design Play Cards: designing for sustainability. Melbourne: Eco Innovators.

Friedman, B., & Hendry, D. (2012) The Envisioning Cards: A Toolkit for Catalyzing Humanistic and Technical Imaginations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'12). New York, NY: ACM, 1145-1148. doi:10.1145/2207676.2208562

Golembewski, M., & Selby, M. (2010). Ideation decks: a card-based design ideation tool. In Proceedings of the 8th Conference on Designing Interactive Systems (DIS 2010), pp. 89-92, New York, NY: ACM. doi: 10.1145/1858171.1858189

Halskov, K., & Dalsgaard, P. (2006). Inspiration card workshops. *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*. New York, NY, USA: ACM, pp. 2–11. doi: 10.1145/1142405.1142409

Halskov, K., & Dalsgaard, P. (2007). The emergence of ideas: the interplay between sources of inspiration and emerging design concepts, *CoDesign*, *3*(4), 185–211. doi:10.1080/15710880701607404

Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*, *3*(1), 1-95. doi: 10.2200/S00261ED1V01Y201003HCI008

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – facets of user experience, *Interacting with Computers*, *22*, 353-362. doi:10.1016/j.intcom.2010.04.002

Hassenzahl, M., Heidecker, S., Eckoldt, K., Diefenbach, S., Hillmann, U. (2012). All You Need is Love: Current Strategies of Mediating Intimate Relationships through Technology. *ACM Transactions on Computer-Human Interaction*, *19*(4). doi: 10.1145/2395131.2395137

Hertzum, M., & Jacobsen, N. E. (2003). The evaluator effect: a chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, *15*(1), 183-204.

Hornbæk, K. (2010). Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology*, 29(1), 97-111. doi:10.1080/01449290801939400

Hornecker, E. (2010). Creative idea exploration within the structure of a guiding framework: the card brainstorming game (pp. 101-108). In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction* (TEI'10). New York, NY, USA: ACM. doi: 10.1145/1709886.1709905

IDEO (2003). *IDEO Method Cards: 51 ways to inspire design*. San Francisco: William Stout Architectural Books.

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

Kim, J., Park, S., Hassenzahl, M., & Eckoldt, K. (2011). The Essence of Enjoyable Experiences: The Human Needs - A Psychological Needs-Driven Experience Design Approach. In A. Marcus (Ed.) *Lecture Notes in Computer Science: Vol. 6769. User Experience, and Usability. Theory, Methods, Tools and Practice* (pp. 77-83). doi: 10.1007/978-3-642-21675-6_9

Kramer, J., Daly, S. R., Yilmaz, S., & Seifert, C. M. (2014). A case-study analysis of Design Heuristics in an upper-level cross-disciplinary design course. *Annual Conference of American Society of Engineering Education (ASEE)*, June 16-18, Indianapolis, IN.

Lafrenière, D., Dayton, T., Muller, M. (1999). Variations of a theme: card-based techniques for participatory analysis and design. Extended Abstracts

on Human Factors in Computing Systems (CHI EA '99). New York, NY, USA: ACM. doi: 10.1145/632716.632807

Lallemand, C., Koenig, V., & Gronier, G. (2014). How Relevant is an Expert Evaluation of User Experience based on a Psychological Needs-Driven Approach? In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction*. New York, NY: ACM (pp. 11-20). doi:10.1145/2639189.2639214

Lucero, A., & Arrasvuori, J. (2010). PLEX Cards: a source of inspiration when designing for playfulness. In *Proceedings of the 3rd International Conference on Fun and Games (Fun and Games '10)*, 28-37. New York, NY, USA: ACM Press. doi: 10.1145/1823818.1823821

Lucero, A., Holopainen, J., Ollila, E., Suomela, R., & Karapanos, E. (2013). The playful experiences (PLEX) framework as a guide for expert evaluation. In *Proceeding of the 6th International Conference on Designing Pleasurable Products and Interfaces (DPPI '13)*, 221-230. New York, NY, USA: ACM. doi: 10.1145/2513506.2513530

Moreno, D., Hernández, A., Yang, M., Linsey, J., & Wood, K. (2014). A Step Beyond to Overcome Design Fixation: A Design-by-Analogy Approach. In *Proceedings of Design Computing and Cognition Conference (DCC'14)*. Berlin: Springer.

Nelson, B. A., Wilson, J. O., Rosen, D., & Yen, J. (2009). Refined metrics for measuring ideation effectiveness, *Design Studies*, *30*(6), 737-743. doi:10.1016/j.destud.2009.07.002

Nielsen, J. (1995). *Card Sorting to Discover the Users' Model of the Information Space*. Blog article published on May 25, 1995. Available at http://www.nngroup.com/articles/usability-testing-1995-sun-microsystems-website/

Odom, W., & Lim, K.-Y. (2008). A Practical Framework for Supporting "Third-Wave" Interaction Design. *Proceedings of alt.chi 2008, Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM Press.

Pine, B.J., & Gilmore, J.H. (1998). Welcome to the Experience Economy: work is theatre & every business a stage, *Harvard Business Review*, July-August 1998.

Pruitt, J., & Adlin, T. (2010). *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. San Francisco: Morgan Kaufmann Publishers.

Shah, J. J., Vargas-Hernandez, N., & Smith, S. M. (2003). Metrics for measuring ideation effectiveness. *Design studies*, *24*, 111-134. doi:10.1016/S0142-694X(02)00034-0

Shedroff, N. (2009). *Experience Design 1 Cards*. New Riders. Available at http://www.experiencedesignbooks.com/EXP1C/index.html

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*, *89*, 325-339. doi: 10.1037//O022-3514.80.2.325

Sheldon, K. M., Ryan, R. M., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, *22*, 1270-1279. doi: 10.1177/01461672962212007

Spencer, D. (2009). *Card sorting: designing usable categories*. New York, USA: Rosenfeld Media.

Stolterman, E. (2008). The Nature of Design Practice and Implications for Interaction Design Research. *International Journal of Design*, *2*(1), 55-65.

Tuch, A., Trusell, R., & Hornbæk, K. (2013). Analyzing users' narratives to understand experience with interactive products. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2079-2088). Paris, France: ACM. doi:10.1145/2470654.2481285

Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008). Towards practical user experience evaluation methods. *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland.

Visser, W. (1996). Two functions of analogical reasoning in design: A cognitive-psychology approach. *Design Studies*. Special Issue "Design Cognition and Computation", *17*, 417-434.

Wölfel, C., & Merritt, T. (2013). Method card design dimensions: a survey of card-based design tools. In P. Kotzé et al. (Eds.) *Lecture Notes in Computer Science: Vol. 8117*. INTERACT 2013, Part I (pp. 479-486). doi: 10.1007/978-3-642-40483-2_34

Yilmaz, S., Daly, S. R., Christian, J. L., Seifert, C. M., & Gonzalez, R. (2014). Can experienced designers learn from new tools? A case study of idea generation in a professional engineering team. *International Journal of Design Creativity and Innovation*, *2*(2), 82-96. doi:10.1080/21650349.2013.832016

Yilmaz, S., Daly, S. R., Seifert, C. M., & Gonzalez, R. (2014). Design Heuristics as a tool to improve innovation. *Annual Conference of American Society of Engineering Education* (ASEE), June 16-18, Indianapolis, IN.

Yoon, J., Desmet, P., & Pohlmeyer, A. (2013). Embodied Typology of Positive Emotions; The Development of a Tool to Facilitate Emotional Granularity in Design. In *Proceedings of the International Congress of International Association of Sciences of Design Research (IASDR '13)*, 1195-1206.

HOW DOES USER EXPERIENCE ALTER ESTABLISHED HCI EVALUATION METHODS? EXPERT AND LABORATORY EVALUATION

Carine Lallemand, Vincent Koenig, Guillaume Gronier

Submitted for journal publication

Keywords: user experience, evaluation, expert evaluation, laboratory evaluation, user testing, psychological needs

ABSTRACT

In the "third wave" of Human-Computer Interaction, the emergence of User Experience (UX) as a key concept has opened up both exciting perspectives and hard challenges. The broadened scope of UX (as compared to usability) is inevitably changing the way we evaluate the quality of an interaction.

To investigate how UX alters established HCI methods, we selected two widely used methods - expert evaluation and user testing in a laboratory setting - and applied them to the evaluation of UX using a psychological needs-driven approach. Two complementary experiments are described in this paper. In the first experiment, we asked thirty-three UX experts to perform a UX expert evaluation on four given interactive systems. For each system, we collected data on the predicted fulfilment of seven UX needs as assessed by the experts. In the second experiment, seventy users were asked to evaluate their experience with two systems, by filling out the AttrakDiff scale and a UX needs fulfilment questionnaire. For both experiments, qualitative data was collected through think aloud protocols and debriefing interviews. We then compared the results of both studies in order to assess the quality and limitations of each method for the evaluation of UX. Our results are discussed from a methodological perspective by analysing the challenges and limitations of established HCI methods for the evaluation of UX. By highlighting why and how UX alters widely used evaluation methods, this paper offers a discussion on the way we evaluate UX and reviews alternative paradigms for UX evaluation.

1. INTRODUCTION

Following the "era" of usability and user-centered design, the HCI field recently entered the "era" of user experience (UX) and experience design (Hassenzahl, 2010). This conceptual shift to a more comprehensive and emotional view of human-computer interactions has been accompanied by the development of new or adapted methods for the design and evaluation of interactive systems (Roto, Obrist, Väänänen-Vainio-Mattila, 2009; Vermeeren et al., 2010). These novel methods mainly aspire at coping with the complexity and subjectivity of UX, as compared to the more objective view of usability. However, a majority of these new methods need more time for consolidation and are only slowly transferred into practice (Odom & Lim, 2008). At the moment, established HCI evaluation methods such as ex-situ user testing or expert evaluation therefore tend to remain standard practices in both research and practice (Alves, Valente, & Nunes, 2014). In this paper, we use the term "established" to refer to widely used and accepted methods that were proven to conform with accepted standards in our field. A majority of current established user-centred evaluation methods were developed as usability evaluation methods; yet several studies show that these usability methods are now used by extension for the evaluation of UX (Alves et al., 2014).

As established HCI evaluation methods are still in use, we can wonder how the shift to UX impacts these methods: what are the challenges that experts are facing when evaluating UX using the "usual" methods they were trained to use? What practices remain unchanged yet effective, valid and reliable? What are the new requirements for UX evaluation?

In the present study, we used a psychological needs-driven approach to experimentally investigate how UX alters two well-established HCI evaluation methods, namely expert evaluation and user testing in a laboratory setting. In the first section of the paper, we describe how UX raises new challenges and questions about the topic of evaluation. We then report on the two experiments we conducted to address the research questions and compare the results to assess the appropriateness of each method for UX evaluation. Finally, we discuss the results from a methodological perspective.

2. FROM THE EVALUATION OF USABILITY TO THE EVALUATION OF USER EXPERIENCE

Since its inception, the field of HCI has been primarily concerned with the design of usable systems whose evaluation was mainly focused on rather instrumental concerns such as effectiveness, efficiency or learnability (ISO 9241-11, 1998). The most widely used usability evaluation methods were usability testing and inspection methods (Cockton, 2014).

During the last decade, the emergence of UX as a key concept opened up both exciting perspectives and hard challenges. Concurrently with numerous attempts at scoping and defining UX (Law et al., 2009), a broad discussion on UX evaluation and measure rapidly appeared on the research

agenda (Hassenzahl & Tractinsky, 2006; Law et al., 2008). However, the diversity of definitions and interpretations of what constitutes UX (Lallemand, Gronier, & Koenig, 2015) along with the complexity of UX attributes and consequences makes it difficult to select appropriate UX evaluation methods (Bevan, 2008). Despite sharing common grounds with the concept of usability, UX spans further by also including emotional, subjective and temporal aspects involved in the interaction (Roto et al., 2011). UX is more holistic and thus more complex. Researchers generally agree that UX is subjective, holistic, situated, temporal, and has a strong focus on design (Bargas-Avila & Hornbæk, 2011; Roto et al., 2011). Topics such as interaction meaning, temporal dynamics of an experience, or needs fulfilment through the use of technology challenge the evaluation of UX to an extreme.

To account for the richness and complexity of experiences, UX research attempts at producing viable alternatives to traditional HCI methods. Researchers have thus responded to UX underlying challenges by developing new methods. Nearly eighty of them have been identified and categorized in 2010 (Roto, Obrist, Väänänen-Vainio-Mattila, 2009; Vermeeren et al., 2010) and many more methods have been developed during the last four years. Regrettably, novel UX evaluation methods are rarely validated (Bargas-Avila & Hornbæk, 2011) and are only slowly transferred into practice (Odom & Lim, 2008). This is partly due to the fact that novel UX methods are demanding and still need to be adapted to the requirements of evaluation in an industrial setting (Väänänen-Vainio-Mattila, Roto, & Hassenzahl, 2008). UX being commonly understood by practitioners as an extension of usability (Lallemand et al., 2015), established usability evaluation methods remain standard practice for the evaluation of UX (Alves et al., 2014). Bargas-Avila and Hornbæk (2011) reviewed 66 empirical studies on UX and concluded that the most frequent UX evaluation pattern is a combination of "during and after measurements – similar to traditional usability metrics, where users are observed when interacting and satisfaction is measured afterwards" (p. 2694).

2.1. UX AND LABORATORY EVALUATION PRACTICES

A laboratory evaluation refers to the evaluation of human-computer interactions in a controlled environment where the evaluator monitors the use of a system, observes users' actions and reactions, and assesses users' feelings about the quality of the interaction. Laboratory evaluations are generally opposed to in-situ (also called "field" or "in-the-wild") evaluations, which involve assessing the interaction in its real context of use. Laboratory evaluation sessions generally involve the use of a combination of methods (aka. mixed-methods), the more typical being: scenarios of use to observe how users operate (both in a non interfering way and a posteriori based on video and sound recording of user behaviour), think aloud protocols to capture users' immediate experience, questionnaires to provide a standardized quantitative measure of factors of interest, log file analysis, and finally debriefing interviews.

During the third wave of HCI (Bødker, 2006), new topics such as UX or ubiquitous computing have shaken up established design and evaluation

methods. While controlled experiments used to be the gold standard in many disciplines, a recent trend in our field claims for more naturalistic evaluation approaches (Rogers, 2011; Schneiderman, 2008; see also Crabtree and al., 2013 "the turn to the wild" TOCHI special issue). A passionate debate notably animated the Ubicomp community following the publication of Kjeldskov et al.'s intentionally provocative paper "It is worth the hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field" (2004), where the authors claim that field studies bring not much added value to the usability evaluation process. In the field of UX, the laboratory setting has been described as less effective for evaluating UX than it is for evaluating usability (Benedek & Miner, 2002). With the acknowledgment of the temporal and contextual factors underlying UX, the "turn to the wild" movement has gained influence in research (Rogers, 2011).

Surveys on UX practice show that field studies are considered the most important practice, though they are not widely used (Vredenburg et al., 2002). Laboratory evaluations therefore remain common practice, even if more sophisticated tools have now stepped into the lab to support the evaluation of human-computer interactions. The development of psycho-physiological measurements such as eye-tracking, skin conductance activity or facial expression analysis software and devices allow for an in-depth investigation of human cognitive and emotional processes involved in UX. HCI researchers can of course take advantage of these new methods, though they have to be aware of their limitations and pitfalls (Park, 2009), especially linked to data misinterpretation. Besides these technological tools, new self-reported evaluation scales and questionnaires have been developed (or imported from other fields) to assess several facets involved in UX, such as emotions (Bradley & Lang, 1994; Desmet, 2003; Huisman, 2009), hedonism (Hassenzahl, Burmester, & Koller, 2003), aesthetics (Lavie & Tractinsky, 2004), values (Friedman & Hendry, 2012), desirability (Benedek & Miner, 2002) or psychological needs (Hassenzahl, 2010; Sheldon et al., 2001).

To overcome the limitations of controlled ex-situ experiments, Kjeldskov et al. (2004) proposed to enhance the realism of laboratory setups by arranging the space so as to recreate realistic contexts of use. They successfully recreated a healthcare context to test the usability of a portable working device. While appealing, this idea quickly becomes limited when considering large-scale environments or mobility practices. Technological tools might be used in the lab to cope with the issue, for instance through the use of simulators or augmented reality devices (Kjeldskov & Skov, 2007).
To summarize, it seems that UX rekindles discussions on field versus laboratory studies. It has also changed laboratory evaluations by promoting the multiplication of measuring devices. The scope of this study however is not to discuss additionally deployed tools but rather the impact of UX on the general principles of laboratory studies.

2.2. UX AND EXPERT EVALUATION PRACTICES

Though the involvement of users in design and evaluation processes is relentlessly stressed as essential, Alves et al. (2014) showed in a recent paper

on the state of UX evaluation practice that final users are involved in less than 50% of reported UX evaluation cases. This basically suggests that in half of the cases, UX evaluations solely rely on experts. These results are consistent with previous findings on UCD practices, where users' involvement was often described as selective and punctual rather than systematic and where heuristic evaluations were reported as frequent (Venturi & Troost, 2004; Vredenburg et al., 2002).

Expert-based evaluation has a long history in HCI, since the development in the 1990's of inspection methods, consisting in the inspection of the user interface by one or several evaluators (Nielsen, 1993). Unlike user-based methods, where the evaluation relies on the (non-interfering) observation of users performing a set of tasks while actually interacting with a system, the evaluation of a system through expert-based methods relies solely on the expertise and judgment of the evaluator (Dillon, 2001). Sometimes labelled as discount usability engineering methods (Nielsen, 1989), inspection methods enjoyed some popularity with usability practitioners, who appreciated them for being cheap, fast and easy to use (Nielsen, 1993; Vredenburg et al., 2002). Heuristic evaluation (Nielsen & Molich, 1990) and cognitive walkthrough (Wharton, Rieman, Lewis, & Polson, 1994) are the most common usability inspection methods and have been extensively used by HCI practitioners for more than three decades (Dillon, 2001; Vredenburg et al., 2002). While these methods provide benefits for detecting and fixing usability issues, several studies suggest that they fail at supporting an evaluation of UX. In their study focusing on Youtube ®, Silva and Dix (2007) showed for instance that the worldwide successful video-sharing website Youtube ® failed when assessed with traditional usability heuristics. The authors even suggested that the issue might be more fundamental, in the way that "conventional usability [could] neglects or conflicts with the more ludic aims of the site".

With the conceptual shift from usability to UX, we could have expected either that the use of expert-based evaluation would decrease to the benefit of evaluation methods involving user participation, or that novel or adapted methods supporting the evaluation of UX would arise. However, while studies on UX practice (Alves et al., 2014) show evidence that the use of expert-based evaluation is still common practice, scientifically grounded methods supporting an expert evaluation of UX are rare. Out of their collection of 96 UX evaluation methods, Vermeeren et al. (2010) identified only 13 expert methods of which six require users or groups of users in addition to the expert. Only seven methods are therefore described as purely expert-based. Amongst these, some methods are derived from expert evaluation, like Jordan's property checklists (2000) described as a structured way to conduct an expert evaluation; or Jordan's concept of immersion where the investigator himself uses the system in real contexts and evaluates it. More recently, Wilson (2011) suggested transferring perspective-based inspection from the study of usability to the study of UX. He defines it as "a user interface evaluation method where the evaluators are asked to adopt a specific perspective as they examine a product for problems".

Following the evolutions of the HCI field, several heuristic sets have also been developed to take into account new concerns beyond usability: playability heuristics (Korhonen & Koivisto, 2006) or heuristics for human-

environment interaction in virtual worlds (Sutcliffe & Gault, 2004) for instance. Regarding UX, Väänänen-Vainio-Mattila and Wäljas (2009) developed UX heuristics for Web 2.0 services. Their initial set was composed of seven heuristics, but surprisingly only a single heuristic (H6 "general UX-related issues") dealt with hedonic or subjective aspects of the experience. In the revised version of these heuristics, H6 is replaced by a heuristic on web services usability and a trust and safety heuristic, thus restraining the scope of the evaluation to mainly pragmatic issues. More recently, two sets of generic UX heuristics were developed for the design and evaluation of systems providing users with a positive experience (Arhippainen, 2013; Colombo & Pasch, 2010). Arhippainen (2013) highlights the need for fast and cost effective UX evaluation methods that can be used during early stages of product design. Willing to adopt a more comprehensive approach to UX than focusing only on optimal experiences as Colombo and Pasch (2010), she proposed the "Ten UX Heuristics", based on empirical UX studies. While interesting, many of these approaches have not yet been sufficiently studied empirically. We therefore do not know how effective and appropriate these methods are for the evaluation of UX.

3. RESEARCH OBJECTIVES

This study investigates how UX alters established HCI methods by thoroughly analysing the processes and outcomes of two types of UX evaluations: expert evaluation and laboratory testing. We specifically chose these two methods as they are frequently mentioned as the more commonly used user-centred evaluation methods (Venturi & Troost, 2004; Vredenburg et al., 2002).

Based on our findings, we aim at identifying the respective strengths and weaknesses of expert evaluation and laboratory testing when it comes to the evaluation of UX. Knowing more about the new set of challenges we have to address when assessing UX will allow us to suggest ways of adapting research methods and evaluation practices to the particular characteristics of UX.

With regards to the expert evaluation, our research questions sound: what differences can we observe between expert evaluation based on usability frameworks as compared to that based on UX? How do experts conduct UX expert evaluation? What factors are impacting the evaluation? How close are expert evaluations to users' experiences? Similarly, with regards to the laboratory evaluation: what differences can we observe between a laboratory deployment for usability analysis as compared to the deployment for UX analysis? What factors are impacting the evaluation?

4. MATERIALS AND METHOD

Two experiments are described in this paper. To test how UX alters established HCI methods, we selected two widely used HCI methods - expert evaluation and user testing in a laboratory setting - and applied them to the evaluation of UX using a psychological needs-driven approach.

In this approach, the fulfilment of human psychological needs is thought to be a main trigger of positive experiences with technologies (Hassenzahl, 2010; Kim et al., 2011). Within the field of positive psychology, an extensive

amount of studies have demonstrated that psychological needs are particular qualities of experience that all people require to thrive (Deci & Ryan, 2000; Sheldon et al., 2001, Sheldon, Ryan, & Reis, 1996): they are both necessary inputs and driving motives (Sheldon, 2011). The transfer of this assumption to the field of HCI has led to psychological needs-driven UX approaches (Hassenzahl, 2010; Kim et al., 2011; Tuch, Trusell, & Hornbæk, 2013) where designers should consider interactive systems as means to fulfil needs ("be-goals") and not only means to achieve task oriented "do-goals" (Hassenzahl, 2010). Needs provide categories of experiences, such as "competence experiences" or "relatedness experiences" (Hassenzahl, 2013) that UX practitioners should seek to design.

Based on the literature on fundamental human needs (Sheldon, 2011; Sheldon et al., 2001) and on studies linking these needs to the UX of interactive systems (Hassenzahl et al., 2010; Hassenzahl et al., 2013; Kim et al., 2011; Tuch, Trusell, & Hornbæk, 2013), we selected seven candidate needs supposed to be relevant in the context of UX design and evaluation (Table 1). In our experiments, we assessed the fulfilment of these seven needs as metrics for the quality of UX.

Table 1. The seven needs represented on UX Cards and their definition (Sheldon et al., 2001)

Need	Definition
Relatedness	Feeling that you have regular close contact with people who care about you rather than feeling lonely and uncared of.
Competence	Feeling that you are very capable and effective in your actions rather than feeling incompetent or ineffective.
Autonomy	Feeling like you are the cause of your own actions rather than feeling that external forces or pressure are the cause of your action.
Security	Feeling safe and in control of your life rather than feeling uncertain and threatened by your circumstances.
Pleasure	Feeling that you get plenty of enjoyment and pleasure rather than feeling bored and understimulated by life.
Meaning	Feeling that you are developing your best potentials and making life meaningful rather than feeling stagnant and that life does not have much meaning.
Popularity	Feeling that you are liked, respected and have influence over others rather than feeling like a person whose advice or opinion nobody is interested in.

In the first experiment, UX experts assessed four interactive systems by conducting a UX expert evaluation: the social network Facebook ®, the e-commerce website Amazon ®, the game Angry Birds ® on IPhone ® and finally an Olympus ® digital compact camera. For each system, we collected data on predicted fulfilment of seven UX needs as gauged by the experts.

Next, we selected two out of the four systems evaluated by the experts, namely Amazon and the camera, and designed a user testing session in a laboratory setting. Facebook was excluded from the laboratory experiment due to privacy reasons, whereas Angry Bird was not selected because of its primary focus on the need for pleasure. We created task scenarios based on the main tasks assessed by the experts in the first experiment. We also let the users freely explore each system in order to encourage a realistic user experience. Seventy users were asked to evaluate their experience with the

two systems, by filling out the AttrakDiff scale and a UX needs questionnaire (adapted from Sheldon et al., 2001).
Finally, we compared the results obtained in both experiments by looking at analogous metrics (i.e., overall evaluation of needs fulfilment). All details regarding the metrics are explained at the beginning of each study section. We used this comparison as a basis for discussing the respective strengths and weaknesses of UX expert and laboratory evaluation.

5. STUDY 1: EXPERT EVALUATION

5.1. PARTICIPANTS

Thirty-three UX experts (16 women et 17 men) participated in this study. They were recruited among personal contacts and through an advertisement on social networks. Mean age of the sample was 31 years (*Min* 23, *Max* 43, *SD* = 5.96). Table 2 shows the background of the participants. About two thirds of the participants were consultants or practitioners working in Industry (n = 20) while the other third were researchers or students working in Academia (or between Industry and Academia). Experts were mainly educated in Psychology or Social Sciences (n = 19). The average level of expertise with expert evaluation (or heuristic evaluation), self-assessed on a 7-points Likert scale, is 5.24 (*SD* = 1.39). The average familiarity with UX at a theoretical level is 5.21 (*SD* = 1.78) while the average familiarity with psychological needs theories was self-assessed to be much lower (*M* = 3.97, *SD* = 1.89). On a 7-points Likert scale ranging from 1 "not familiar" to 7 "very familiar", experts also reported their level of familiarity with each use case: Facebook (*M* = 6.09, *SD* = 1.72), digital cameras (*M* = 5.73, *SD* = 1.35), Amazon (*M* = 5.7, *SD* = 1.49), and Angry Birds (*M* = 4.45, *SD* = 2.12).

Table 2. General profiles of the experts

Background Variable		Frequency (n = 33)	Valid Percentage
Domain	Industry	20	60.6 %
	Academia	5	15.2 %
	Both or between	8	24.2 %
Role	Researcher	9	27.3 %
	Consultant / Practitioner	20	60.7 %
	Student	4	12.1 %
Education	Design	5	15.2 %
	Psychology / Social Sciences	19	57.6 %
	Cognitive Sciences	3	9.1 %
	Technology / Software	4	12.1 %
	HCI	2	6.1 %

5.2. PROCEDURE AND MATERIAL

The study took place in several locations, most of the time at the workplace of each participant. Each individual session lasted approximately 2 hours. Participants received a 50€ shopping voucher in compensation for their time spent. We provided experts with seven UX cards, a tool to support a psychological needs-driven UX expert evaluation (Lallemand, Koenig, & Gronier, 2014). The UX cards are designed to be used as an expert evaluation technique, but should not be considered as heuristics. The UX

cards provide experts with some knowledge about basic psychological needs that should be fulfilled to produce a positive UX; however they do not encompass a comprehensive list of dimensions and sub-dimensions to check when verifying if an interface complies with these guidelines. The goal of the UX cards is not to "debug" the system but to assess how well it might support the fulfilment of human needs, leading to a positive experience.

5.2.1. INTERACTIVE SYSTEMS

Four interactive systems were inspected during the experiment: (1) the social network Facebook ® (2) the online e-commerce website Amazon ® (3) the game Angry Birds ® on IPhone ® 5S and finally (4) an Olympus ® digital compact camera. We chose four varied examples of interactive systems in order to maximize the diversity of HCI elements and also the diversity of potential need fulfilment coverage, while still providing a common ground for comparison across the experts, which would have been compromised if experts were allowed to freely choose their example. Facebook was for example expected to show a high proportion of relatedness elements, while Angry Birds was assumed to encompass more pleasure related elements. In addition to screen-based interfaces, we also decided to ask experts to inspect a tangible object, namely a digital camera. Before the task, experts reported their level of familiarity with each of the four systems.

5.2.2. GENERAL INSTRUCTIONS AND PRELIMINARY SURVEY

After welcoming each participant, we first explained the main goals of the study and the psychological needs-driven UX theory. We then invited each expert to fill in a preliminary survey, indicating gender, age, educational background, professional background (domain, role, job title) and experience (in the HCI field, with expert evaluation, with the concept of UX, and with psychological needs theories). Then, we familiarized the experts with the UX cards and the underlying theory; we read each card and answered questions if necessary. Several metrics related to the quality of the UX Cards for UX evaluation (e.g., understandability, operationalizability, usefulness) were collected and are reported in Lallemand, Koenig, and Gronier (2014).

5.2.3. UX EVALUATION TASK

We asked the experts to evaluate each of the four systems during 15 minutes. The four systems were presented in a counterbalanced and rotated order to avoid sequence biases by distributing practice effects equally across conditions. We instructed the experts to identify, within each assessed system, elements able to have a positive or negative impact on one or several UX needs. Neutral observations were not written down. Experts were completely free in their evaluation, so that they could for example relate several needs to a single element, as well as identify the same need as both positive and negative for the same element. Complete freedom was also given to the experts regarding the type of elements they could identify. They were able to identify anything they thought could impact UX, from

elements as broad as the system's brand or the system's concept to elements as precise as features, interface design or content.

Overall need fulfilment evaluation: For each system, after having identified elements impacting UX during a 15 minutes timespan, we asked the experts to provide an overall UX assessment of the system. This overall assessment relied on 7-points Likert scales (one scale per need) to answer the question: "Overall, what is the impact of (name of the assessed system) on the fulfilment of the need for (name of the need)?" The scales ranged from "negative" to "positive".

Reporting tool: We provided experts with a paper-based grid to report UX elements during the evaluation. The grid was composed of three columns: identified element, UX need(s) positively impacted by this element and UX need(s) negatively impacted by this element. To ensure that all instructions were clearly understood before starting the evaluation task, the experimenter went through a given example by showing the participant how to report elements and related needs in the assessment. After task completion, we collected experts' opinions during a debriefing interview.

5.3. RESULTS OF THE EXPERT EVALUATION EXPERIMENT

5.3.1. UX EVALUATION TASK: IDENTIFICATION OF ELEMENTS AND RELATED UX NEEDS

Overall, the experts identified 1794 elements, which corresponds to an average of 54 elements per expert and 13 elements per assessed system. The number of cited elements did neither significantly differ according to the order the systems were presented in, nor according to the assessed system. Similarly, background variables (age, gender, seniority, level of familiarity with UX or level of familiarity with psychological needs theory) did not significantly impact the number of cited elements.
Experts linked these identified elements to a total of 3455 UX needs. 2277 needs were cited as positive (66%) and 1179 needs cited as negative (34%). Experts were thus more focused on interactive elements able to fulfill UX needs than on elements having a negative impact on needs. Despite the counterbalancing of use cases across participants, results show a significant order effect impacting the number of needs cited for each system: experts cited less UX needs during the evaluation of the first system (M = 24.03) than for the next three systems (*diff* = -3.1, $t(32) = 2.12$, $p < .05$) (with means ranging from M = 26.39 to M = 27.15). It thus seems that the appropriation time required to master the evaluation tool is relatively short (about 15 minutes).

Most cited needs (Figure 1) were security (22%, 771 citations) and pleasure (23%, 784 citations), while least cited needs were influence (8%, 266 citations) and self-actualizing (6%, 211 citations). Experts declared that self-actualizing was the hardest need to evaluate, as not many interactive systems succeed in fulfilling such a need.

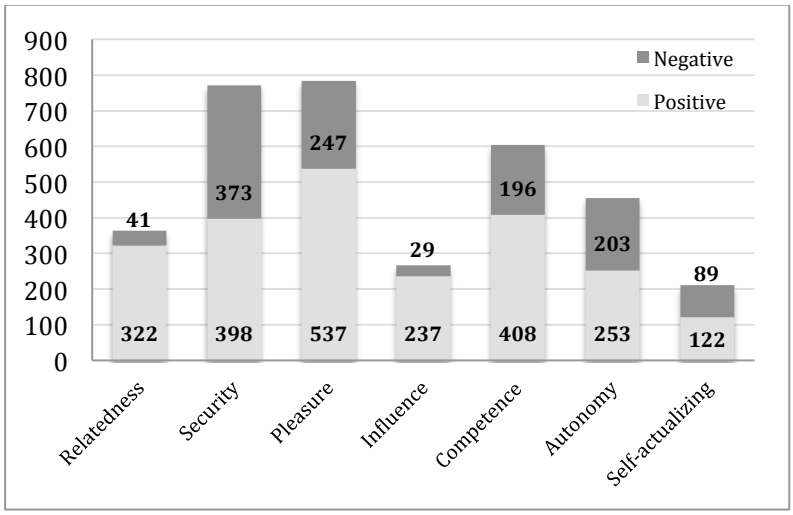


Fig. 1. Total number of cited needs during the UX evaluation task

We also observed some patterns in the way needs were cited as either positive or negative. We can distinguish three categories of needs:

- needs almost only cited as positive with an approximate ratio of 90/10 (i.e., 90% cited as positive and 10% cited as negative): the needs for relatedness or influence
- needs frequently cited as positive but also sometimes as negative with an approximate ratio of 70/30: the needs for pleasure or competence
- needs almost equally cited as positive and negative with an approximate ratio of 45/55: the needs for security, autonomy and self-actualizing

These patterns suggest that the seven needs do not impact UX in a homogeneous way. For instance, it is not likely that an identified element negatively impacts the fulfilment of relatedness or influence. It seems that these feelings might be absent during the interaction but very rarely negatively impacted by the interaction. On the opposite, the needs for security, autonomy or self-actualizing are equally impacted negatively than positively, which means that elements of the interaction are both able to fulfil these needs or to hinder them.

Before starting the evaluations, experts were varyingly experienced with the four use cases. Their level of familiarity significantly correlates with some evaluation variables, especially in the case of Facebook: the experts' level of familiarity with Facebook positively correlates with the number of elements identified ($r = .36$, $p < .05$) and with the total number of cited needs ($r = .49$, $p < .05$). Interestingly, it is also related to the number of needs cited as positive ($r = .48$, $p < .05$), but not to the number of needs cited as negative. In other words, the more familiar experts are with Facebook, the more positive needs they are likely to cite. In the case of Angry Birds, the only significant link relates the familiarity level to the number of needs cited as positive. No significant correlations were however found between the familiarity level and the number of elements or needs cited regarding the use cases Amazon and camera.

Regarding the age of our participants, the only significant correlation shows that the younger an expert was, the more needs he cited on average (r = -.36, p < .05), especially positive needs (r = -.41, p < .05). As age and seniority are of course strongly related (r = .90, p < .01), results also show that senior experts tend to cite significantly less positive needs than less experienced practitioners (r = -.36, p < .05). Self-reported familiarity with UX, familiarity with UX needs theories and the level of expertise with heuristic evaluation do not significantly impact the total number of elements or needs cited by the experts. The same holds for gender, domain, role or education.

5.3.2. OVERALL UX EVALUATION OF THE SYSTEMS

In order to understand how UX experts actually assess the UX of interactive systems, we compared the identification task (number of cited elements and needs) to the overall evaluation (7-points Likert scales) conducted after each evaluation. This comparison allows us to understand how the overall UX evaluation made by the experts relates to the elements they have identified during the 15-minutes evaluation time.
Figure 2 presents the overall need fulfilment evaluations of the four interactive systems, as derived from the Likert scales ratings. An overall UX evaluation score has been computed by averaging the scores of the seven individual needs.

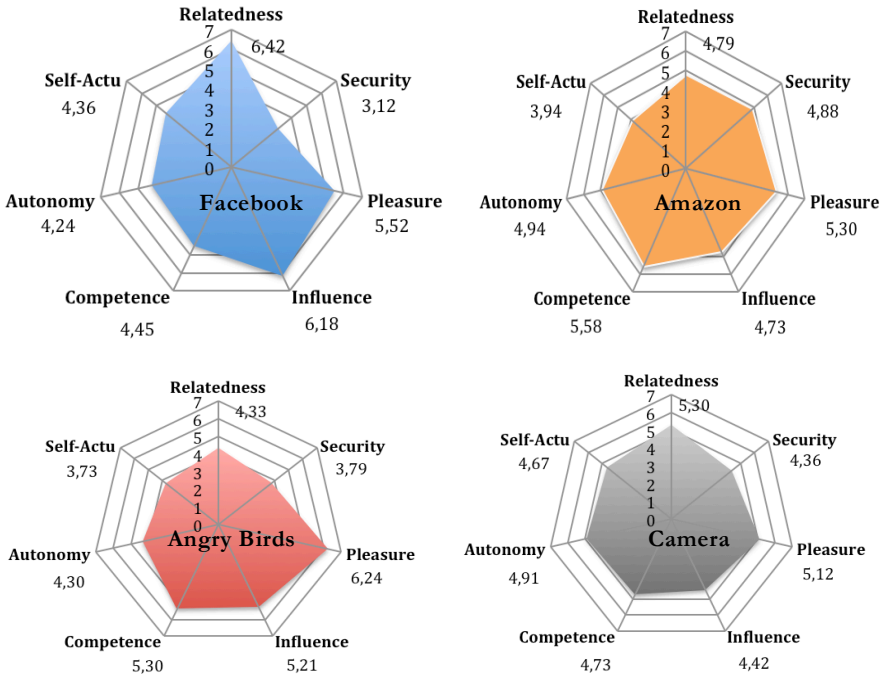


Fig. 2. Overall UX Evaluation of the interactive systems. Average ratings are presented under each need.

Background variables such as age, seniority, familiarity with UX, familiarity with needs theories or number of heuristic evaluations performed are not significantly correlated to the overall assessment of our uses cases, except for the camera. In this case, age and seniority are negatively correlated to

the overall UX assessment (r = -.67 and r = -.53 respectively, p < .05). The assessment of the camera also differs significantly according to the level of expertise with HE (r = -.47, p < .05). The assessment of tangible objects, which obviously become obsolete after a period of time, seems to be impacted more by personal characteristics than other types of systems.
The level of familiarity with a system is positively correlated with the global assessment in all cases, except the camera. The more an expert is familiar with Facebook, the more he is likely to assess the system as globally positive (r = .46, p < .01). The same holds for Amazon (r = .45, p < .01) and Angry Birds (r = .37, p < .05). Implications of these results for the validity of the UX expert evaluation method will be discussed in the next section.

By exploring the links between the overall need fulfilment assessment of the systems and the number of elements identified during the evaluation task, results show no significant links between those two factors, except for Amazon (r = .39, p < .05). This suggests that the overall a posteriori assessment is globally not influenced by the number of elements an expert has identified. This might be explained by the strict 15-minutes time limit that imposed a general identification limit across all the participants, thus reducing the variance in their quantitative production.
Systematic significant links were however observed between the overall need fulfilment assessment and the number of needs cited as positive or negative. In all cases, the overall assessment is positively correlated to the number of positive needs cited and negatively correlated to the number of negative needs cited. Table 3 presents bivariate Pearson's correlation coefficients for each use case.

Table 3. Correlations between overall UX assessment and number of needs cited as positive or negative

Overall UX assessment	Number of needs cited as positive	Number of needs cited as negative
Facebook	.55**	-.41*
Amazon	.62**	-.41*
Angry Birds	.57**	-.55**
Camera	.71**	-.56**

* significant at p < .05 level; ** significant at p < .01 level

This observation suggests that the overall assessment is mainly based on the identification task, which means that experts base their judgment on their identification of elements and related needs. We could therefore consider the whole (overall assessment) to be coherent with the sum of its parts (single elements and needs identified) in the case of a UX expert-based evaluation. However, a closer look at the results for each need partially invalidates this assumption.
In the case of Facebook, the overall need fulfilment assessment is significantly correlated only to the number of positive and negative needs for security (r = .54 and r = -.30 respectively, p < .05), autonomy (r = .51 and r = -.41 respectively, p < .05) and self-actualizing (r = .46 and r = -.32 respectively, p < .05). The links between global evaluation (for each need) and the number of times the related needs were cited as positive and negative, are not systematically significant. In the case of Facebook, the evaluation of relatedness, pleasure and competence did not rely on the identification task, whereas this was the case for the other needs. This means that the overall assessment of needs might sometimes be based on

other factors than the actual number of elements and needs identified. The level of familiarity with Facebook is for instance positively correlated to the specific assessment of the needs for security ($r = .42$, $p < .05$), autonomy ($r = .31$, $p < .05$) and self-actualizing ($r = .47$, $p < .05$).

5.3.3. CATEGORIES OF ELEMENTS ON WHICH EXPERTS' EVALUATIONS WERE BASED

Table 4 presents the categories of elements on which experts' evaluations were based. We coded each element identified by experts in the reporting grid as belonging to one or several categories of elements. On average, experts mainly identified elements related to the concept/content (28 %), to the features (22 %), the usability (18 %) or the design (18 %) of the use cases. They rarely reported on more holistic elements such as the marketing/brand (4 %), interoperability (2 %), or service experience (3 %). The proportion of identified elements per category differs according to the use case: Facebook and Amazon were for instance mainly assessed through its concept or features, whereas the camera was mainly assessed through its design and usability.

Table 4. Elements on which experts' evaluations were based

	Marketing / brand	Concept / content	Design	Usability	Features	Interoperability	Service experience	Adverts
Facebook	30	259	38	79	188	14	10	28
	5 %	40 %	6 %	12 %	29 %	2 %	2 %	4 %
Amazon	34	179	63	81	164	12	34	10
	6 %	31 %	11 %	14 %	28 %	2 %	6 %	2 %
Angry Birds	18	143	114	96	52	17	12	65
	3 %	28 %	22 %	19 %	10 %	3 %	2 %	13 %
Camera	17	69	200	156	92	5	13	0
	3 %	13 %	36 %	28 %	17 %	1 %	2 %	0 %
Total	99	650	415	412	496	48	69	103
	4 %	28 %	18 %	18 %	22 %	2 %	3 %	5 %

6. STUDY 2: LABORATORY EVALUATION

6.1. PARTICIPANTS

We conducted user-testing sessions in a usability laboratory. Seventy participants (36 males, 34 females) were recruited through several channels: mailing list, social networks and advertisement in public places. Participants received 30 euros in compensation for their time spent. The sample's mean age was 29 ($Min = 18$, $Max = 48$). All materials were in French. Fluency in French was a prerequisite to participate in the study. However, we also asked the participants to report their native language in order to study potential language- or culture-related effects. Amongst our 70 participants, 38 (54.3%) were French native speakers and 32 (45,7%) had another mother tongue. Regarding their employment status, 50% were employed, 48.6% were students and 1.4% unemployed. Almost all participants declared feeling at ease with technology ($M = 5.84$, $SD = 1.22$).
Regarding the use cases, 83% of the participants are registered on Amazon for more than a year. The average level of familiarity with Amazon's website on a 5-points scale is relatively high ($M = 3.74$, $SD = 1.09$). Participants'

opinions about Amazon assessed on a 5-points scale are positive (M = 4.10, SD = 0.74). Participants' average level of familiarity with digital cameras (in general) was also assessed as relatively high (M = 3.41, SD = 1). Fifty participants out of seventy own a camera, however 16 participants do not know what type of camera they own. Amongst camera owners, 30% of participants use their camera less than once a month (n = 21), only 7.1% use it several times a week and only 1 participant out of 50 uses it every day.

6.2. PROCEDURE AND MATERIAL

First, we welcomed the participants and explained them the functioning of the laboratory. The participants were then made familiar with our strict ethical requirements and signed an informed consent form. After having presented the experiment's general instructions, we asked them to fill a preliminary survey including variables such as age, gender, mother tongue, employment status, and familiarity with technology (7-point Likert scale from "I feel not at all at ease with technology" to "I feel completely at ease with technology").

6.2.1. USE CASES AND SCENARIOS

During the user test, the participants had to assess two interactive systems: the e-commerce website Amazon.fr ® and an Olympus ® digital compact camera. The two systems were presented in a counterbalanced order to avoid sequence biases by distributing practice effects equally across conditions. Half of the participants interacted with Amazon first and then with the camera (Order 1) whereas the other half interacted with the camera first and then with Amazon (Order 2).
In order to stimulate the exploration of the systems, we defined scenarios with tasks to achieve. The scenarios were chosen to represent the main actions performed by users on such systems. Five scenarios were related to Amazon: exploring featured recommendations on the home page and adding one item to a wishlist, searching for a book and looking inside to read some pages, consulting customers' reviews, adding the book to the shopping cart and finally browsing the shop for a pair of shoes. To enhance the ecological validity of the assessed experience, we asked the participants to log in using their own Amazon account. The suggestions made by Amazon's recommender system were therefore real suggestions based on prior items viewed or bought by each user. Seven scenarios were related to the Camera: taking a picture in Auto mode, making a short movie, entering the gallery view, deleting the movie, taking a picture in Magic mode, setting image size, and finally exploring the Help menu. Participants were aware that the performance was not assessed and were instructed to work through the scenarios without time or failure pressure. Participants were asked to freely explore each system before starting the scenarios and to think aloud to provide their impressions. After task completion, we collected participants' opinions during a debriefing interview.

6.2.2. SYSTEM EVALUATION

After having achieved the scenarios of each use case, participants were asked to answer the AttrakDiff scale. The AttrakDiff questionnaire has

been developed by Hassenzahl, Burmester, and Koller (2003) to measure both pragmatic and hedonic quality of an interactive product. The measurement items are presented in the format of 28 semantic differentials. The ratings are therefore made on bipolar seven-point anchors (word-pairs are presented to the user, each word in a pair representing the extreme opposite of its counterpart). Evaluated system's qualities are Pragmatic Qualities, Hedonic Qualities (subdivided into Hedonic-Stimulation and Hedonic-Identification) and finally Attractiveness.
As all materials were in French, we used the French version of the AttrakDiff (Lallemand, Koenig, Gronier, & Martin, in press). After having checked the reliability of each AttrakDiff subscale (independently for both use cases), we computed mean scale values for each subscale by averaging the respective items for each participant, and a global AttrakDiff score (ATD_TOTAL) by averaging the values of the four subscales. Regarding Amazon, Cronbach's alpha values range from .67 [Hedonic-Identity] to .86 [Attractiveness] (scale inter-correlations from .00 to .66 with an average of .40). Regarding the camera, Cronbach's alpha values range from .70 [Hedonic-Identity] to .90 [Attractiveness] (scale inter-correlations from .00 to .69 with an average of .46).

6.2.3. NEED FULFILMENT

Evaluation of needs fulfilment using the UX needs scale: Need fulfilment was assessed using an adapted and translated version of the scale developed by Sheldon et al. (2001). Thirty items divided into seven subscales were used to assess the fulfilment of seven basic needs: Competence, Autonomy, Security, Pleasure, Relatedness, Influence, Self-Actualizing. After having checked the reliability of each UX need subscales for both use cases independently, we computed mean scale values for each need by averaging the respective items for each participant, and a global need fulfilment score by averaging all items.

Overall evaluation of needs fulfilment using needs definitions: We also asked users to assess, at the end of the questionnaires, the fulfilment of the needs at a global level. While the UX needs scale basically breaks each need into several items, our global needs fulfilment variable presents the users with a description of each need (as defined by Sheldon et al., 2001) and ask them to evaluate how present each need was during the interaction on a 5-points Likert scale from 1 (not at all) to 5 (completely). The beginning of the sentence "When using Amazon, I have the feeling that" was therefore followed by seven definitions to rate. We called this variable *Need definition*.

Importance of needs fulfilment: Finally, based on the assumption that some needs might be perceived as more important than others depending on the system and the context, we asked participants to report on a 5-points Likert scale how important they assessed each need in the context of the interaction with either Amazon or a digital camera.

6.3. RESULTS OF THE LABORATORY EXPERIMENT

We used univariate statistics to examine the means and standard deviations of each item as well as to check for possible outliers or entry errors. No

outliers or entry errors were found. We used SPSS v22 software to perform statistical analyses. An assessment of the normality of data has been conducted and this condition has been verified before the computation of independent-samples t-tests, bivariate Pearson's correlations and one-way ANOVAs.

6.3.1. SCENARIOS, ORDER EFFECTS AND DEMOGRAPHICS

On average, participants worked through the Amazon test scenarios in 11 minutes (*Min* = 5, *Max* = 26, *SD* = 4). The time spent on a task is generally associated with the usability of a system. As expected, the duration of the task correlates negatively with the AttrakDiff pragmatic scale (r = - .28, p = .018). However, there is no correlation with AttrakDiff's hedonic or attractiveness subscales. Similarly, there is no correlation between the time spent on the Amazon scenarios and UX needs scale or subscales: the duration of the tasks does apparently not influence the perceived fulfilment of basic needs. Regarding the digital camera, participants achieved the scenarios in 11 minutes on average (*Min* = 5, *Max* = 29, *SD* = 4). The time spent on the testing scenarios is negatively correlated to the AttrakDiff's pragmatic subscale (r = -.24, p = .044) and positively correlated to the hedonic-stimulation subscale (r = .24, p = .047). These results suggest that the more time a user spent on the scenarios, the less the camera is assessed as pragmatic, but the more it is assessed as stimulating. Similar to the Amazon use case, we found no correlation between time spent on the camera scenarios and UX needs scale or subscales.

Independent-samples t-tests were conducted to compare the effects of sequence order on the evaluation of UX and fulfilment of needs. Several significant order effects were observed. When participants interacted with Amazon as a second use case (Order 2), they assessed it more positively on the AttrakDiff scale (M = 5.03, SD = 0.67) than in the condition were they interacted with Amazon first (Order 1; M = 4.73, SD = .57), *diff* = -.30, $t(68)$ = -2.06, p = .043. At the subscale level, interacting with the camera first and Amazon afterwards lead to a better evaluation of Amazon's pragmatic quality (M = 5.74, SD = 0.72) than in the case where Amazon was experienced first (M = 5.22, SD = 0.79), *diff* = -.53, $t(68)$ = -2.91, p = .005. The same tendency was observed for the reported fulfilment of UX needs (M = 2.80 for Order 1 vs. M = 3.23 for Order 2), *diff* = -.43, t(68) = -2.87, p = .005. Similarly, participants assessed the camera's pragmatic quality higher in the Order 2 condition (M = 4.72) than in the Order 1 condition (M = 4.22), *diff* = -.50, $t(68)$ = -1.79, p = .039 one-tail.

6.3.2. UX ASSESSMENT USING THE ATTRAKDIFF SCALE

Overall, Amazon was positively assessed on the AttrakDiff scale (M = 4.88, SD = 0.64), especially regarding its pragmatic quality (M = 5.48, SD = 0.79) and attractiveness (M = 5.13, SD = 0.89) (Table 5). The UX of the digital camera was also positively assessed (M = 4.35, SD = 0.83), with the highest rating on attractiveness (M = 4.71, SD = 1.13) and the lowest rating on hedonic-stimulation (M = 3.85, SD = 1.11). Independent samples t-tests show no significant differences between owning a camera (n = 50) or not (n = 20) on the AttrakDiff evaluation and no effect on need fulfilment.

Table 5. AttrakDiff scale: descriptive statistics and reliability analyses (N = 70)

	Amazon website					Digital camera				
AttrakDiff scores	Min	Max	M	SD	Cronbach's alpha	Min	Max	Mean	SD	Cronbach's alpha
AttrakDiff global score	3.50	6.46	4.88	0.64	.88	2.39	5.96	4.35	0.83	.92
Pragmatic	3.14	6.86	5.48	0.79	.75	1.57	6.71	4.47	1.18	.87
Hedonic-stimulation	2.29	6.43	4.16	0.89	.74	1.86	6.29	3.85	1.11	.87
Hedonic-identification	2.57	6.71	4.76	0.77	.67	2	6.14	4.36	0.85	.70
Attractiveness	2.71	7	5.13	0.89	.86	2.14	6.71	4.71	1.13	.90

For both use cases, we observed no significant gender differences (independent-samples t-test) or age differences (bivariate correlations) with regards to the evaluation of UX. Self-reported level of familiarity with Amazon is positively correlated to the AttrakDiff global rating ($r = .44$, $p < .001$) and its subscales, especially the hedonic-identity subscale ($r = .47$, $p < .001$). The more familiar I am with Amazon, the more I feel the experience is positive. As familiarity level is predictably linked to opinion about Amazon ($r = .50$, $p < .001$), we also observe a correlation between opinion and the AttrakDiff evaluation ($r = .34$, $p < .001$). No significant correlations were found between familiarity with technology and Amazon's UX evaluation. In the case of the digital camera, familiarity with technology is negatively correlated to the evaluation of the hedonic-stimulation quality of the device ($r = -.24$, $p = .041$): the more familiar I am with technology, the less stimulated I am by the camera. No significant correlations were found between familiarity with digital cameras and the UX evaluation.

6.3.3. FULFILMENT OF UX NEEDS

The fulfilment of UX needs assessed using the UX needs scale is presented in Table 6. Results show that the need that is best fulfilled by Amazon is the need for security ($M = 3.93$, $SD = 0.7$), whereas the need that is least fulfilled by Amazon is the need for relatedness ($M = 2.12$, $SD = 0.95$). Regarding the digital camera, the need that is best fulfilled is also the need for security ($M = 3.44$, $SD = 0.92$) and the need that is least fulfilled is the need for influence ($M = 2.04$, $SD = 0.95$). As expected, the fulfilment of UX needs is strongly correlated to the perceived UX assessed through the AttrakDiff scale and this holds true both for Amazon ($r = .50$, $p < .001$) and the camera ($r = .65$, $p < .001$).

Table 6. UX Needs subscales: descriptive statistics and reliability Analyses (N = 70)

UX needs subscales	Nb of items	**Amazon website**					**Digital camera**				
		Min	Max	M	SD	Cronbach's Alpha	Min	Max	Mean	SD	Cronbach's Alpha
Competence	5	1	5	3.62	0.9	.85	1	5	3.2	1	.90
Autonomy	4	1	5	3.62	0.9	.74	1.25	5	3.34	.95	.79
Relatedness	4	1	3.75	2.12	0.95	.82	1	5	2.67	1	.83
Pleasure	4	1	5	2.81	1.05	.83	1	5	2.71	1.06	.87
Security	5	2	5	3.93	0.7	.70	1.2	5	3.44	.92	.83
Influence	4	1	5	2.38	1.04	.86	1	4.25	2.04	.95	.86
Self-Actualizing	4	1	5	2.62	0.94	.79	1	4.75	2.51	.87	.70

We observed no significant gender differences with regards to Amazon's ability to fulfil UX needs and only one gender effect related to the camera's ability to fulfil the need for competence, *diff* = .55, $t(68)=2.37$, $p = .021$, men feeling more competent ($M = 3.47$) than women ($M = 2.92$) while using the camera. Age is not correlated with the fulfilment of needs, except for a moderate negative correlation between age and the fulfilment of self-actualizing in the case of Amazon ($r = -.26$, $p = .027$).

Familiarity with technology is not correlated with any need fulfilment in the case of Amazon, and only correlated with the fulfilment of the need for competence in the case of the camera ($r = .27$, $p = .021$): the more users felt at ease with technology, the more competent they felt using the camera. Neither level of familiarity with Amazon nor opinions about Amazon are correlated to the fulfilment of UX needs. Level of familiarity with digital cameras is correlated to the fulfilment of the need for security while using the camera ($r = .25$, $p = .040$). We found an effect of language on the fulfilment of the need for autonomy in both use cases: French native speakers feel more autonomous than non-native speakers while using Amazon (M = 3.84 vs. M = 3.34 respectively), diff = -0.5, t(68) = -2.39, p = .019. The same phenomenon was observed for the camera (M = 3.58 for French native speakers vs. M = 3.05 for non native speakers), diff = -0.53, t(68) = -2.40, p = .019.

Beyond the completion of the UX needs scale, we also provided users with a definition of each need and asked them to rate on a 5-points Likert scale from 1 (not at all) to 5 (completely) how present this feeling is during the interaction. As these seven items summarize the seven dimensions explored through the UX needs scale, we expect the definition items to be strongly linked to the average score of each dimension of the UX needs scale. As expected, each definition item is significantly correlated to its relative UX needs subscale, ranging from $r = .26$ ($p = .029$) for Amazon autonomy to $r = .68$ ($p < .001$) for Amazon competence; and from $r = .33$ ($p = .006$) for camera relatedness to $r = .82$ ($p < .001$) for camera competence. The definition of each need therefore seems compliant with the items composing the needs subscales.

We also tested the differences in means between the definition items and their relative subscale through paired-samples t-test (Table 7). With regards to Amazon, all differences were significant at $p < .05$ levels, except for the relatedness subscale where the difference was not significant. In the case of the camera, differences were significant at $p < .05$ levels for the needs of relatedness, competence, pleasure and self-actualizing. Overall, the fulfilment of needs reported through the seven single definition items is lower than the one reported through the multiple items of the UX needs subscale. We observe two exceptions in both use cases: the need for pleasure and the need for competence collected higher ratings through the definition evaluation than through their respective need subscale.

Table 7. Paired-sample t-tests between need fulfilment assessed through need definition and UX needs scale (N = 70)

UX needs		Amazon					Digital camera				
		M	SD	t (ddl 69)	Sig	Pearson's correlation	M	SD	t (ddl 69)	Sig	Pearson's correlation
Autonomy	Definition	3,20	1,04	-2.92	.005	.26 p =.029	3,29	1,24	-.43	.671	.50 p = .000
	Subscale	3,61	0,90				3,34	,95			
Relatedness	Definition	2,06	1,24	-.53	.594	.56 p =.000	2,17	1,27	-3.14	.002	.33 p =.006
	Subscale	2,12	0,95				2,67	1,01			
Competence	Definition	3,87	1,08	2.64	.010	.68 p =.000	3,49	1,25	3.34	.001	.82 p =.000
	Subscale	3,62	0,90				3,20	1,00			
Security	Definition	3,69	1,07	-2.47	.016	.63 p =.000	3,29	1,12	-1.46	.149	.64 p =.000
	Subscale	3,93	0,70				3,44	,92			
Pleasure	Definition	3,19	1,24	2.78	.007	.54 p =.000	3,04	1,17	3.04	.003	.66 p =.000
	Subscale	2,81	1,05				2,71	1,06			
Influence	Definition	2,00	1,08	-3.2	.002	.55 p =.000	1,97	1,03	-.63	.528	.55 p =.000
	Subscale	2,38	1,04				2,04	,95			
Self-Actualizing	Definition	1,94	1,13	-5.65	.000	.54 p =.000	2,16	1,11	-3.02	.004	.54 p =.000
	Subscale	2,62	0,94				2,51	,87			

Finally, we also asked users to rate how important the fulfilment of each need is, in the context of an interaction with Amazon or in the context of an interaction with a digital camera. According to the participants, the most important needs to be fulfilled when interacting with a website such as Amazon are the needs for security (M = 4.39, SD = 0.95), competence (M = 4.10, SD = 0.98) and autonomy (M = 3.83, SD = 1.03). Regarding the camera, the most important needs to be fulfilled are competence (M = 4.31, SD = 0.75), pleasure (M = 4, SD = 0.99) and security (M = 3.83, SD = 1.1). In both cases, we notice that the needs for security and competence are in the top-3 important needs to be fulfilled.

Figures 3 and 4 (section 6) present the differences in the ratings for need fulfilment using the UX needs scale, need fulfilment using the definition items and importance of needs fulfilment. As we can see in the case of the camera the needs are always rated as more important based on the importance scale as compared to the definition scale. This could suggest that users' expectations about the UX of the camera were not satisfied. Huge differences between importance and presence are observed for the needs of competence, pleasure, self-actualizing or relatedness. The results are different in the case of Amazon with a better balance between needs fulfilment and needs importance; some needs being assessed as equal or even more present than important.

6.3.4. INTERVIEW

We collected participants' feelings during a debriefing interview and asked them to describe their experience with each of the two assessed systems using a single word. Amongst 70 words collected (one per participant) to describe the UX of Amazon, 83 % had a positive meaning, 13 % were neutral and 4 % had a negative meaning. Most cited words were "practical" (11 occurrences), "effective" (8 occurrences) or "good" (8 occurrences). Opinions regarding the digital camera were more heterogeneous with 47 % of positive words, 20 % of neutral words and 33 % of negative words. Most cited words were "satisfying" (6 occurrences), "novel" (4 occurrences) or "banal" (3 occurrences).

Table 8. Single-word UX description for each of the two use cases (N = 70)

Single-word UX description	Amazon		Camera	
	Frequency	Valid percentage	Frequency	Valid percentage
Negative	3	4,3	23	32,9
Neutral	9	12,9	14	20
Positive	58	82,9	33	47,1
Total	70	100	70	100

We also asked users which elements influenced their UX positively or negatively. For Amazon, they mainly pointed to the content (31 %, 138 citations), the usability (22 %, 97 citations) and the service experience (16 %, 71 citations). The elements influencing their UX while interacting with the camera were mainly the usability (37 %, 128 citations), the design (26 %, 91 citations) and the features (26 %, 88 citations). The ratio between positive and negative elements is much more positive in the case of Amazon (64 % positive vs. 36 % negative) than in the case of the camera (52 % positive vs. 48 % negative), which follows the UX scores reported through the questionnaires.

Regarding the testing session, a majority of participants reported difficulties to assess some of the UX needs due to the testing situation. Relatedness or influence needs items were highlighted as problematic because of the absence of people in the lab, especially people that are important to the user. For instance, a participant said, "this camera would probably contribute to the fulfilment of the need for relatedness, if I were at home or on holidays taking pictures of my wife and kids. But here alone in the lab, I truly don't feel that way, so I assessed it as not fulfilled at all" (participant #10).

Similarly, the testing situation impacted the need for autonomy, which was perceived as ambiguous. Even if one might feel autonomous when surfing on Amazon or when using a camera, the controlled testing situation places individuals in a context where freedom is inherently limited. During questionnaire completion, several participants reported (by thinking aloud) that their feeling of autonomy was reduced by the situation, even if they could imagine that they would feel autonomous with the systems.

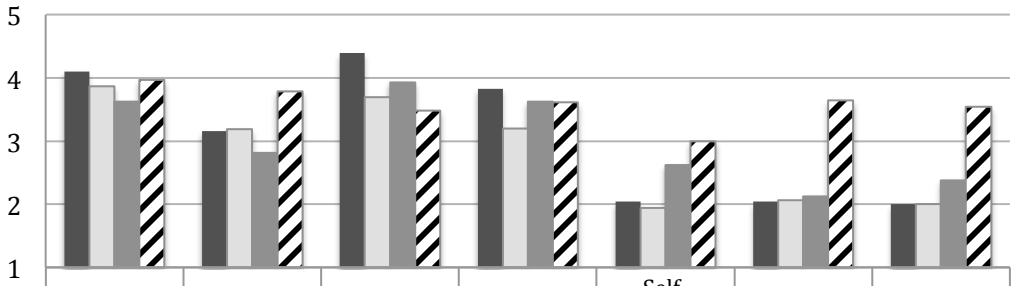
Furthermore, several participants also reported on the fact that, beyond the feeling of autonomy, their experience was impacted by the testing situation in other ways. Many of them mentioned the fact that they performed actions through the testing scenarios that they would not have performed at home, because they usually are not using these kinds of systems this way. Sometimes the scenarios would lead to positive experiences; this was for instance the case for a participant who discovered nice shoes on Amazon though she generally would only look for books or computer material. But more often in this experiment, this led to frustration and negative experiences; for instance when users had to modify settings on the camera (e.g., picture size, filter effects) and reported that they would only have used the Auto mode at home and would probably have been very satisfied with it. So despite the fact that we tried to keep scenarios easy and we instructed the participants that they could skip any of the scenarios if they want to, the laboratory setup indeed modified the felt experience.

In some cases, we also observed major differences in the evaluation made by participants using the questionnaires and the feelings reported during the debriefing interview. For instance participants could rate their experience with the camera as quite negative, because they had experienced several issues while performing the test, but then they might report the same experience as satisfying during the interview because they somehow felt that the device was interesting and could be enjoyable after a short learning period (participant #21). Any time we felt that there was an inconsistency between users' ratings and their experience report, we asked the users to explain why in order to understand the rationale behind their UX evaluation.

7. COMPARISON BETWEEN UX EXPERT EVALUATION AND LABORATORY EXPERIMENT

For ease of comparability, we limited our analysis to the reported levels of needs fulfilment and the categories of elements on which experts' and users' evaluations were respectively based.

Fig. 3 compares users' needs fulfilment ratings (e.g., importance of needs, needs fulfilment scale, and needs fulfilment definitions) to the global rating made by experts (overall evaluation using Likert scales at the end of the expert evaluation session). Independent samples t-tests were conducted to see in which cases experts' assessments were close to users' real experiences. Significant differences between experts' and users' ratings therefore mean that the expert evaluation does not precisely reflect the experience as felt by users.

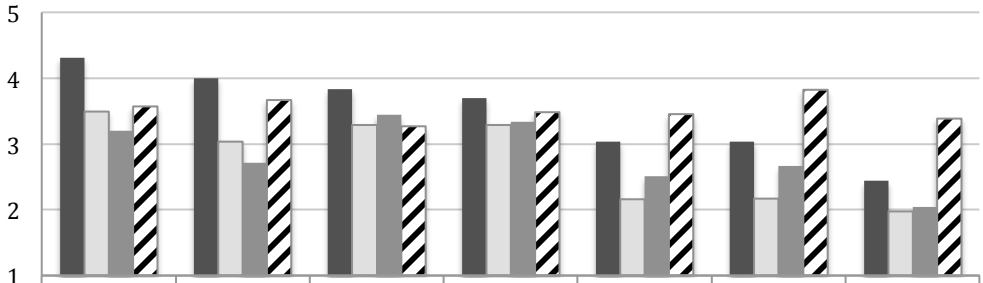


	Competence	Pleasure	Security	Autonomy	Self-actualizing	Relatedness	Influence
Importance	4,1	3,16	4,39	3,83	2,04	2,04	2
Need definition	3,87	3,19	3,69	3,2	1,94	2,06	2
UX needs scale	3,62	2,81	3,93	3,62	2,62	2,12	2,38
Expert evaluation	3,97	3,79	3,48	3,61	3	3,64	3,54

Fig. 3. Comparison between needs fulfilment ratings across users and experts – Amazon

Regarding Amazon, differences between experts' ratings and users' evaluation using the UX needs scale were significant for every need, except for the need for autonomy. Average differences show that experts tend to overestimate the fulfilment of five needs (out of six significant differences) but not the need for security. Average differences are especially high for the needs of relatedness (*diff* = 1.51), influence (*diff* = 1.16) and pleasure (*diff* =

0.97), suggesting that experts largely overestimated the fulfilment of these needs. Differences regarding competence, security and self-actualizing are significant yet smaller (ranging from 0.35 for competence to 0.44 for security). When comparing experts' ratings to the rating made by users through the needs definition items, results show non-significant differences for the needs of competence, autonomy and security. Highest significant differences are related to the needs of relatedness (*diff* = 1.58), influence (*diff* = 1.54) and self-actualizing (*diff* = 1.05).



	Competence	Pleasure	Security	Autonomy	Self-actualizing	Relatedness	Influence
Importance	4,31	4	3,83	3,7	3,04	3,04	2,44
Need definition	3,49	3,04	3,29	3,29	2,16	2,17	1,97
UX needs scale	3,2	2,71	3,44	3,34	2,51	2,67	2,04
Expert evaluation	3,57	3,67	3,27	3,48	3,45	3,82	3,39

Fig. 4. Comparison between needs fulfilment ratings across users and experts – Camera

With regard to the camera use case (Figure 4), differences between experts' ratings and users' evaluation using the UX needs scale were found for the needs of relatedness, pleasure, influence and self-actualizing. Average differences show that experts tend to overestimate the fulfilment of these four needs, the highest average differences being found for the needs of influence (diff = 1.35) and relatedness (diff = 1.14). Noteworthy is that all significant differences are quite high for this use case (the smallest difference being .94 for self-actualizing). Inversely, non-significant differences suggest that experts accurately assessed the needs for competence, autonomy and security. When comparing experts' ratings to the rating made by users through the needs definition items, results show a similar pattern with highest average differences observed for relatedness, influence and self-actualizing and non-significant differences for the needs of competence, autonomy and security.

All in all, we observed that experts tend to overestimate the fulfilment of UX needs, and this holds particularly true for the needs of relatedness, influence and pleasure. Their assessment more accurately reflects users' evaluations when it comes to the fulfilment of autonomy (with non significant differences in both use cases), security, or competence needs. Results also show that experts tend to perform better in the camera use case and that their evaluation is closer to the users' rating using the definition items than to users' rating using the UX needs scale. As users mentioned difficulties feeling related or influent in a laboratory setup, one might

assume that their rating for the fulfilment of these needs would have been higher in a natural setting. One might also wonder whether experts more accurately rated the needs for autonomy, security or competence because these needs are closer to usability factors and heuristics (e.g., user control, feedback, ease of use, learnability) that they are used to assessing. The difference in self-actualizing might be explained by the phenomenon called false alarms (i.e, reported problems not verified by user tests) in usability expert evaluation (Hvannberg, Law, & Larusdottir, 2007): during their evaluation task, experts mentioned difficulties in identifying elements related to the need of self-actualizing. This is not surprising considering that this is a very high level need, not easily fulfilled through the use of interactive systems (Hassenzahl, 2010). Nevertheless, experts frequently felt compelled to use the self-actualizing card and therefore tended to expressively search for some elements fulfilling this need. We hypothesize that these elements might have acted as anchors (Tversky & Kahneman, 1974), leading experts to assess the fulfilment of self-actualizing more positively than they would have done without having previously identified elements related to this need. Finally, one should mention that experts were unsurprisingly more familiar than users with both Amazon (M = 4.07 for the experts vs. M = 3.74 for the users) and digital cameras (M = 4.09 for the experts vs. M = 3.41 for the users), which could also have an influence on their rating.

We also compared the categories of elements used for the assessment of UX by experts and users respectively (Table 9).

Table 9. Categories of identified elements by experts (N= 33) and users (N= 70)

		Marketing / brand	Concept content	Design	Usability	Features	Inter-operability	Service experience	Adverts
Amazon	Experts	6 %	31 %	11 %	14 %	28 %	2 %	6 %	2 %
	Users	4 %	31 %	6 %	22 %	14 %	1 %	16 %	6 %
Camera	Experts	3 %	13 %	36 %	28 %	17 %	1 %	2 %	0 %
	Users	1 %	10 %	26 %	37 %	26 %	0 %	0 %	0 %

Bivariate Pearson's correlations have been used to investigate how the categories of identified elements are related to the UX needs being evaluated as positive or negative by experts. In the case of Amazon, significant correlations were found between the total number of needs cited as positive and the number of elements in the categories features (r = .40, p = .021), interoperability (r = .49, p = .004), service support (r = .39, p = .026) and concept / content (r = .30, p = .042 one tail). Inversely, significant correlations were found between the total number of needs cited as negative and the number of elements in the design category (r = .32, p = .033 one tail). These results suggest that experts assessed Amazon's design as rather negative, whereas they assessed its features, interoperability and service support as positive.

In the case of the camera, significant correlations were found between the total number of needs cited as positive and the number of elements in the categories marketing / brand (r = .40, p = .020), concept/content (r = .46, p = .006) and features (r = .40, p = .021). Inversely, significant correlations were found between the total number of needs cited as negative and the

number of elements in the categories design ($r = .64$, $p < .001$) and usability ($r = .51$, $p = .003$). These results suggest that experts assessed the camera's design and usability as rather negative, whereas they assessed its concept/content and features as positive.
Amongst users, Amazon's concept/content, service UX, features and usability were mainly cited as positive, whereas the brand and adverts were cited as negative. Interoperability was only cited twice.
The camera's usability was mainly reported as negative, whereas its design or features were majorly cited as positive. Furthermore, while nineteen users reported the absence of a design element or feature on the camera as negatively impacting their experience, only two experts thought about identifying an "absence" as an issue possibly impacting UX. Similarly, two users also complained about the location of the buttons on the right side of the camera, which is not adapted for left-handers. None of the experts addressed this kind of issue in their report.

8. DISCUSSION

The results of the present study shed light on issues and challenges to be addressed when evaluating UX using established HCI methods such as expert evaluation or laboratory user testing. In the discussion section, we will show that some issues are not novel and were already recognized as problematic for the evaluation of usability. However, the extended scope of UX along with its subjective, temporal and situated nature has brought additional challenges to tackle. We will first discuss our two experiments separately before addressing UX evaluation at a global level by reviewing alternative evaluation methodologies.

8.1. LIMITATIONS, BENEFITS AND CHALLENGES OF UX EXPERT EVALUATIONS

As expert evaluation is seen by practitioners as a cheap and effective method to assess usability (Nielsen, 1993), we expect UX practitioners to apply this method to UX evaluation as well. However, assessing something as complex and inherently subjective as UX without involving users is even more challenging than assessing the usability of a system. Despite the fact that expert evaluation is often used in combination with other evaluation methods involving final users, it is necessary to reflect on the suitability of expert evaluation in the context of UX evaluation. The results of our first experiment involving UX experts allow us to better understand whether expert evaluation remains relevant in the context of UX and to suggest ways of adapting expert practices.

First, we saw that experts encountered no blocking issue during the task and felt able to conduct an expert evaluation of UX based on a psychological needs-driven approach. Interestingly, a majority of experts was not aware of the psychological needs-driven approach linking basic human needs to UX. However our participants all showed a strong interest for this topic and agreed on the fact that this seems a promising approach to assess UX. As shown in the results section, experts tend to link elements to positive needs rather than to negative ones. This highlights a tendency to consider UX more from a positive perspective than from a negative one. This is a first

major difference between conducting an expert evaluation of UX as compared to usability, where the main focus is on the identification of design flaws. A UX expert evaluation is thus not primarily meant to "debug" a system – even if some issues and negative points are also reported - but to support assessing the system in its ability to fulfil human needs, ultimately leading to a positive experience. From this perspective, UX expert evaluation might constitute a complementary and relevant contribution to the design process. The comparison between expert- and laboratory evaluations shows that experts performed rather well at assessing the needs for competence, autonomy and security. Other parts of their assessment did not match the user perspective; it is therefore crucial to be aware of the limitations of UX expert evaluation in order to adapt our practices towards a better validity.

Our study contributes to highlighting that limitations that already existed for the evaluation of usability are still present and even exacerbated by the nature of UX.

A major criticism of expert evaluation techniques is that they do not make use of information about users and their use contexts. Hornbæk (2010) criticized studies on usability evaluation methods for considering usability issues as stable, independent from circumstances and users. Obviously, the strong embedment of UX into daily routines, social interactions and physical, technical, or organizational contexts of use challenges attempts at conducting UX expert evaluations to an extreme. In our expert experiment, contexts of use as well as temporal dynamics of UX were neglected by a majority of the participants. Our results show for instance that experts almost never reported the absence of an element as being a potential UX issue, whereas users' assessment in the laboratory experiment attested that is it indeed an important matter of concern. The explanation relates to the situatedness of UX: users mentioned for instance the absence of a rotating screen on the digital camera as problematic to take nice "selfies" (i.e., pictures that one takes of oneself). Awareness about current users' practices and context of use is therefore required to conduct a thorough UX evaluation. Similarly, users also complained about the absence of a tactile screen mainly because this is nowadays a standard feature on this kind of cameras. Expectations and previous technological experiences are also essential factors and UX evaluation should therefore be based on user research data. On this issue, it is also important to note that the lack of contextual clues was inherent to our experimental design. As we wanted to stay at a generic level, we asked the experts to evaluate how each identified element would impact the UX of a "regular user" or of "the majority of users". While research clearly shows that UX is unique to an individual and influenced by several individual and contextual factors (Roto et al., 2011), we cannot consider that a stable number of elements will impact UX needs or that this impact will be the same for every user involved in any context of use. It would be necessary to investigate the issue of subjectivity further, by implementing a research design where UX experts would have access to contextual information (under the form of personas, or scenarios of use) while conducting their evaluation.

Interestingly, we also noticed that experts were less used to assessing subjective and hedonic aspects of UX and therefore mentioned more pragmatic aspects in their reports. Making an informed guess of what users are likely to feel during an experience is undoubtedly harder than estimating

users' likelihood to succeed or fail in performing a task. Vermereen, Cremers, Kort, and Fokker (2008) already highlighted this observation in their study comparing a UX evaluation conducted through a field study, laboratory evaluations and expert reviews. As did our experts, their participants also overestimated the ratings of hedonic UX aspects in their expert reviews. Additionally, our observations and the analyses of the categories of elements they identified show that experts struggle at considering UX as holistic, and at assessing UX beyond the graphical user interface. In the case of Amazon for instance, service experience was mentioned by users in 16 % of the cases as impacting their UX but only constituted 6 % of the elements identified by experts. In some cases, apparently important UX elements were forgotten during the evaluation task (sometimes recalled later on): as an example, some experts did not mention in their report any relatedness elements supported by the digital camera. However, when assessing the overall impact of the digital camera on the need for relatedness, they suddenly remembered that taking pictures of family or friends could have a positive impact on relatedness. The UX cards provided to the experts already encompass an instruction card entitled "UX is holistic", which aims at encouraging experts to think about UX in a global way. We assume that without the presence of this card, the categories of elements identified by the experts would have been even more focused on the interface only than they actually are.
A second recurrent criticism of expert evaluation is related to reliability (Molich et al., 1998): several experts do not report the same set of problems when conducting a usability evaluation. In the case of UX, we saw that the issue of reliability becomes even more challenging as more subjectivity comes into play. The significant correlation between the familiarity level with a system and its overall evaluation seems problematic since it implies that experts' evaluations were somehow biased by their own subjective experiences. Some differences were also observed in the results according to the background of the experts. It seems for example that younger and less experienced experts cited more needs on average and were more focused on positive needs than their more experienced counterparts. Even if these differences were significant in a few cases only, we could expect some experts to perform better than others at accurately evaluating UX.

As pinpointed by surveys on UX practice (Venturi & Troost, 2004; Vredenburg et al., 2002), informal low-cost methods such as expert evaluation are widely used, notwithstanding their inherent limitations and the ever-growing development of alternative methods involving users. Despite the limitations highlighted in the present research, it would thus be erroneous and unrealistic to simply recommend the abandon of expert-based methods for UX evaluation. When adapted to the specificities of UX, expert evaluation might produce relevant and impactful outcomes, as demonstrated by Rantavuo and Roto's study at Nokia (2013). To achieve this goal, it is yet essential to raise awareness on the limitations of the method and to suggest better ways of conducting UX expert evaluations.

First, we want to emphasize that UX expert evaluation should be based on a thorough user research stage. Experts should use contextual data while conducting their evaluation. This only will help them to overcome their own expectations, feelings, opinions, and standards in order to assess the actual user experience as perceived by others. To help experts adopting an

end-user perspective, we therefore suggest combining the use of UX Cards with methods providing contextual information, such as scenarios of use or personas (Pruitt & Adlin, 2010). We also expect that a domain/application expert might be required to assist the UX expert in the case of business-specific systems requiring a deep understanding of users' tasks and objectives.

Based on our observations, we also believe that UX experts willing to conduct a holistic UX evaluation need more guidance during the evaluation. We saw that despite the fact that more than one third of our participants used the term UX in their job title, experts lack understanding of UX holistic scope and related factors. In the case of our UX cards, we propose a step-by-step guidance to support experts in providing a thorough and overall UX assessment, which would not be limited to pragmatic elements or category of elements only. First, experts will be advised to think about the UX of the assessed system at a very generic level (e.g., concept, brand, associations, visual design, service design) before assessing the system from a functional perspective (e.g., features, interoperability, interaction design). Then, the evaluation should focus on detailed user interface elements (e.g., content, information design, accessibility, usability issues). Finally, we will invite the experts to reflect on missing elements (which would be required to provide the desired UX or to satisfy user expectations) and temporal dynamics of the experience. This guidance might be applied to several expert evaluation tools and could help UX experts to improve their evaluation and also to harmonize their practice in case of multiple experts assessing the same system or product. The ability to use UX evaluation outcomes in the design process is also of major importance (Law, Abrahão, Vermeeren, & Hvannberg, 2012). As experts will not only identify issues to be fixed, the outcomes of their evaluation are less obvious to be directly exploited. This raises a challenge in terms of communicating and implementing their findings in a design process; and needs further investigation.

From a research perspective, future studies could overcome the limitations described above (i.e., lack of contextual clues, low guidance), specifically by testing expert evaluation using contextual data or supported by a better guidance. The 15-minutes duration of the evaluation task might also be debatable. We were aware that this is undoubtedly a short time to achieve an expert evaluation task (according to Nielsen, 1994, an evaluation session lasts one or two hours), however it allowed each expert to evaluate four systems within a single 2-hours session. We also intended to reduce the identification of false positive or false negative elements by focusing on the most prominent elements an expert would be able to identify within 15 minutes. In two cases out of four (Angry Birds and the camera) some experts declared having finished the task before the end of the 15-minutes timespan. It therefore seems that systems encompassing few features might be assessed quickly. For more complex systems, one should investigate further how the duration spent on the evaluation impacts the results. Finally, the differences we observed between experts from different backgrounds and levels of expertise raises questions on the importance of expertise selection (i.e., the process of choosing an expert from a list of recommended people) (Yarosh, Matthews, & Zhou, 2012) or on the necessity to use multiple experts to conduct an accurate UX evaluation. The

application of expert evaluation to UX will probably rekindle discussions on the number of evaluators required for a thorough assessment (Nielsen, 1994).

8.2. LIMITATIONS, BENEFITS AND CHALLENGES OF UX LABORATORY EVALUATIONS

Through our laboratory experiment, we were able to identify which aspects of the user testing situation were still suitable for the evaluation of UX and which aspects seemed to be challenged.

As stressed by Rogers et al. (2007), we saw that traditional usability metrics such as task completion time did not inform a lot about the felt experience. While time spent on testing scenarios in both use cases was negatively correlated with the AttrakDiff pragmatic scale, it had no impact on the perceived attractiveness, nor on the fulfilment of UX needs. Results also show that already known potential issues related to laboratory evaluations remain problematic in the case of UX. Sequence biases were observed and impacted both the perceived experience assessed through the AttrakDiff scale and the reported fulfilment of UX needs.

The testing situation and the laboratory setting impacted the felt experience in many ways. First, being in a laboratory hindered the fulfilment of specific needs, such as relatedness or influence, which are so closely embedded into the social, physical and daily context that they are not easily reproducible in a lab. This issue was frequently reported during the debriefing interviews and therefore it is hard to claim for the validity of our results regarding relatedness and influence needs. Although our participants did not explicitly mention this, the same might apply to the pleasure while interacting with Amazon: the pleasure in that case is mainly derived from buying something that one desires. In a laboratory setting, the tasks are somewhat standardized and even if the users were allowed to freely explore each system, the usage situation was not oriented towards the pleasure of the discovery or buying of appealing products. Similarly, the feeling of self-actualizing might arise from a wonderful photo shoot where one feels particularly creative and spontaneous, however this is the kind of situation that we cannot capture in a laboratory because it is too much embedded in a real-life context. The assessment of the needs for security, autonomy and competence seemed at first sight to be less impacted by the testing situation.

Nonetheless, the artificiality of testing scenarios impacted the felt experience by directing users' towards actions that they would probably not have done in a real-life context. First, users reported a direct negative impact of the testing situation on their assessment of the need for autonomy: as they were guided through the process by achieving standardized scenarios, they felt globally less autonomous and this impacted their evaluation of the system's ability to make them feel autonomous. Moreover, the artificial actions triggered either positive or negative feelings and distorted users' experiences, thereby biasing the outcomes of UX evaluation. This holds unfortunately true even for the needs that seem easier to assess in a controlled experiment, such as security or competence.

The scenario where users had to modify settings on the camera for instance negatively impacted some users' feeling of competence: at home, they would only have used the Auto mode and would probably have been very satisfied with it, not feeling frustrated or incompetent. These artificial behaviours and task selection biases (Cordes, 2001) were already identified as problematic within usability studies (Kjeldskov et al., 2004).

To address these issues, some authors have proposed adding contextual features to laboratory setups in order to improve the realism of the laboratory setting (Kjeldskov & Skov, 2003; Kjeldskov et al., 2004; Kjeldskov & Skov, 2007). Kjeldskov et al. (2004) recreated for instance a healthcare context in a laboratory to study the usability of a mobile application. In the case of UX however, recreating a meaningful setting in-situ seems challenging as UX is very often embedded in - and influenced by – daily routines and usual social interactions. Nevertheless, in the case one needs a controlled experimental setting, trying to recreate the context of use seems relevant. One could think about adding specific furniture, triggering specific situations through role-playing (Simsarian, 2003) or involving families or friends to co-discover the system (Jordan, 2000). Augmented reality devices or simulators might also support a more contextual approach. However, these technological approaches are costly and not yet widely used in industry (Alves et al., 2014).

Another limitation of laboratory UX evaluation relates to the dynamics of UX, which is difficult to assess in a single session. We were already able to observe the impact of time on UX, especially by noticing a difference between the momentary evaluation made by users through the questionnaires and the more reflective evaluation they reported during the debriefing interview. However, there is much more than that to account for UX temporality and this reflects in the growing interest for long-term UX evaluation methods, such as longitudinal methods or retrospective UX assessments (Karapanos, Martens, & Hassenzahl, 2012; Kujala et al., 2011). Without adopting a novel method, how could we adapt laboratory evaluations to improve the assessment of the temporal dimension of UX? As laboratory evaluations often entail a combination of evaluation methods, we could add specific tools to better understand UX over time during a testing session. First of all, it seems essential to investigate users' history by inquiring about their expectations, previous experiences and level of familiarity with the system (or similar ones), opinions about the system or even anticipated UX. Then, one might use tools to assess the changes in UX during the session. Mood maps aim at documenting the emotional states of users over time by asking users to frequently report their emotional state during the test. They might be used to better catch momentary frustrations and to match mood with specific parts of the interaction, thereby informing designers about what specifically should be improved. It is also possible to ask users to answer several questionnaires before, during and after the interaction. Other tools, such as retrospective assessment curves (Karapanos et al., 2012; Kujala et al., 2011) might be used to represent the evolutions of UX over time during the session. While UX curves were primarily designed to assess UX over long periods of time, UX curves such as iScale (Karapanos et al., 2012) might also be used on a shorter timeframe. Finally, thinking aloud protocols along with observation and debriefing interviews has proven in our study to be effective at

detecting changes in the UX. If one wants to be more accurate in detecting changes in emotions or behaviours, novel devices might enter the lab and provide psychophysiological measurements (Park, 2009), eye-tracking data or facial expressions assessment (Zaman & Shrimpton-Smith, 2006).

With regards to another issue related to time, one should be aware that the duration of the session itself constraints the experience and resulting evaluation. Several participants mentioned for example that they had not enough time to truly explore and appreciate the features of the camera, or to truly enjoy exploring products they like on Amazon. As shown in Karapanos et al.'s model of temporal aspects in UX (Karapanos, Zimmerman, Forlizzi, & Martens, 2010), the first period of use is characterised by an orientation phase where the user discovers the system. At this stage, strong UX-related factors such as functional dependency or emotional attachment are absent from the interaction. The evaluation of UX in a single short user testing session therefore remains incomplete. For long-term UX assessed in a laboratory, one could think about multiple sessions involving the same participants, however this approach seems too costly. In practice, longitudinal or retrospective methods are more suitable to address that goal.

Finally, assessing UX in a laboratory requires a thorough reflection on data collection and ethical issues. In the present study, we first conducted expert evaluations before selecting two use cases to be assessed in the laboratory experiment. At first sight, Facebook seemed a good candidate to be used in our users' evaluation sessions because of the diversity and intensity of experiences it triggers. However, we were challenged by privacy issues. While privacy issues were already relevant in the context of usability (and whenever we ask users to perform actions that are observed and recorded), additional challenges arise when dealing with UX. Assessing a realistic Facebook experience would have implied users logging on their own personal account (with their own friends and timeline), whereas at the usability level we could probably have tested the system using a fake account. Systems and products are more and more offering a personalized experience to their users and privacy issues will therefore more frequently come to the fore when assessing the UX of a product on the market in the presence of an observer (or a recording device). This doesn't apply of course to early prototypes or new products, but in those cases the challenge will be to simulate a personalized experience in order to assess their potential UX. Amongst our use cases, Amazon allowed us to assess a personalized service with less privacy issues, although participants still had to agree logging in using their passwords on our testing computer and to show the recommended products based on their previous purchases. The digital camera was the less problematic from a privacy perspective but, as we showed it, its evaluation was nevertheless strongly impacted by the testing situation.

8.3. ALTERNATIVE EVALUATION METHODS

Provided that established methods entail several limitations for the evaluation of UX, what alternatives do we have to evaluate UX in a more

naturalistic context, by taking into account all UX related factors such as temporality?

Several researchers argue in favour of more ecological evaluation methods of UX (Rogers, 2011), highlighting that only "in the wild" studies allow for understanding the complexity and richness of experiences (Schneiderman, 2008). Moreover, authors also claim that field studies provide more valuable insights, thereby better serving design purposes. The main drawback of field studies however is the time and cost required to conduct them, more than twice the time of laboratory evaluations (Kjeldskov et al., 2004; Rogers et al., 2007). This issue is even more critical if one wants to use them as a longitudinal method. Real settings also challenge observation and data collection as one should try to observe and record interactions without interfering too much in the situation. Finally, field studies require working prototypes and are therefore not suitable for early UX evaluation.

The diary study seems to be a good candidate of a research methodology having apparently all requirements to capture the experience from the user point of view by taking into account all aforementioned factors. Following Allport (1942) who was encouraging the use of personal documents in psychological science, Bolger, Davis and Rafaeli (2003) claim diary methods to be able to "capture life as it is" by reporting events and experiences in their natural, spontaneous context. The advantages of diary methods for the study of UX are indeed numerous. First, diary methods allow studying and characterizing temporal and contextual dynamics of UX; this constitutes a real added value in comparison to more widespread methods like interviewing or think-aloud protocols. It also provides more accurate data on the observed phenomenon since the likelihood of retrospection is reduced (Bolger et al., 2003). Validity and reliability of the collected data is therefore expected to be higher than those related to a methodology implying retrospection of an event, such as UX curves for instance (Kujala et al., 2011). Diaries can help determine the antecedents, correlates, and consequences of daily experiences and therefore to better understand the experiences in context. However, diary studies also have main disadvantages related to the cost and time associated with the recruitment of users, training or briefing sessions and data analysis. They are bound to the expressive abilities of participants and might therefore not be used with any population type (Allport, 1942). As diaries involve self-reporting data only, they also have the drawback to be an indirect approach to data collection; they do not provide first hand insight into the user experiences.

All in all, it seems hard to conciliate both the capture of the experiential and emotional flow during interaction, and the cumulative and reflective experience. This is a choice to be made according to the objectives of the study and expected outcomes. To address the limitations of single evaluation methods, it is of course possible to adopt a mixed-method approach by combining several methods (Ardito et al., 2008; Schmettow, Bach, & Scapin, 2014). No UX evaluation method is perfect in the sense of a one-size-fits-all solution and one needs to look at the pros and cons of each method before deciding how to evaluate UX. The trade-off between costs and benefits plays a major role in the choice and adoption of an evaluation method (Vredenburg et al., 2002). Consequently, if UX research wants to foster the adoption of more ecological or longitudinal approaches to UX evaluation, we should put more emphasis on their benefits in comparison to established methods, which are less demanding and costly.

9. CONCLUSION

By gaining better insights into how UX alters expert evaluation and laboratory user testing, we showed that established evaluation methods need to be adapted in order to fit the nature and characteristics of UX. As illustrated by Rantavuo and Roto (2013), "if we had applied heuristic evaluation as it was defined in 1994, the attempt would have most likely been a failure. By adjusting the method to cover UX aspects…. the UX123 program in Nokia has been highly successful" (p. 3). Furthermore, we should also be aware that practitioners adapt the methods to fit their needs and match specific project circumstances (Cockton, 2014; Woolrych, Hornbæk, Frøkjær, & Cockton, 2011). This is why it is important to investigate and communicate on the strengths and limitations of UX evaluation methods (also by considering methods as collections of resources, as suggested by Woolrych et al., 2011), thereby supporting UX experts in selecting the most suitable method according to the application domain, project constraints or organizational factors.

While we should pursue to investigate further into the methods and metrics for UX evaluation, a better transfer from research to practice would also support the dissemination of novel evaluation methods, which were specifically designed for the assessment of UX. By raising awareness on the relevance of field studies (for evaluating UX in context) or in longitudinal studies (to evaluate the dynamics of experiences), we could provide UX practitioners with a larger palette of methods. In return, researchers could benefit from practitioners' feedback and data, leading to a win-win situation.

In conclusion, while UX definitely alters established HCI methods, we should see this as an opportunity to adapt and improve our research and evaluation practices. The question at hand is not whether we might still use established HCI methods for the evaluation of UX or not, but rather how to adapt existing UX evaluation methods, develop new ones and, over all, be able to wisely select the most suitable method depending on the objective of our study.

REFERENCES FOR PAPER E

Allport, G. W. (1942). *The use of personal documents in psychological science*. New York: Social Science Research Council.

Alves, R., Valente, P., & Nunes, N. J. (2014). The state of user experience evaluation practice. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction*. New York, NY: ACM Press (pp. 93-102). doi:10.1145/2639189.2641208

Ardito, C., Buono, P., Costabile, M. F., De Angeli, A., & Lanzilotti, R. (2008). Combining quantitative and qualitative data for measuring user experience of an educational game. *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland.

Arhippainen, L. (2013). A Tutorial of Ten User Experience Heuristics. *Proceedings of International Conference on Making Sense of Converging Media* (AcademicMindTrek '13). New York, NY: ACM Press.

Bargas-Avila, J.A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. Proceedings of the *ACM Conference on Human Factors in Computing Systems* (*CHI 2011)*, 2689-2698. doi:10.1145/1978942.1979336

Benedek, J., & Miner, T. (2002). Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting. Proceedings of Usability Professionals' Association Conference (UPA 2002), Orlando, FL.

Bevan, N. (2008). Classifying and selecting UX and usability measures. Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM) Reykjavik, Iceland, June 18th 2008.

Bødker, S. (2006). When second wave HCI meets third wave challenges. In Proceedings of the 4th Nordic Conference on Human-Computer Interaction. New York, NY: ACM Press (pp. 1-8). doi:10.1145/1182475.1182476

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616.

Bradley, M., & Lang, P. (1994). Measuring emotion: The self-Assessment Manikin and the semantic differential. *Journal of Behavioral Therapy & Experimental Psychiatry*, *1*, 49-59.

Cockton, G. (2014). Usability Evaluation. In M. Soegaard & D. Rikke Friis (Eds.). The Encyclopedia of Human-Computer Interaction, 2nd Ed. Aarhus, Denmark: The Interaction Design Foundation. Available online at https://www.interaction-design.org/encyclopedia/usability_evaluation

Colombo, L., & Pasch, M. (2012). 10 Heuristics for an Optimal User Experience. Proceedings of alt.chi 2012, Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press.

Cordes, E.R. (2001). Task-Selection Bias: A Case for User-Defined Tasks, *International Journal of Human-Computer Interaction*, 13 (4), 411-419. doi:10.1207/S15327590IJHC1304_04

Crabtree, A., Chamberlain, A., Grinter, R., Jones, M., Rodden, T. and Rogers, Y. (2013). Special Issue of "The Turn to The Wild", ACM Transactions on Computer-Human Interaction (TOCHI), 20(3). doi:10.1145/2491500.2491501

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227-268. doi:10.1207/S15327965PLI1104_01

Desmet, P.M.A. (2003). Measuring emotion; development and application of an instrument to measure emotional responses to products. In M.A. Blythe, A.F. Monk, K. Overbeeke, & P.C. Wright (Eds.), *Funology: from usability to enjoyment* (pp. 111-123). Dordrecht: Kluwer Academic Publishers.

Dillon, A. (2001). Usability evaluation. In W. Karwowski (ed.*) Encyclopedia of Human Factors and Ergonomics*, London: Taylor and Francis.

Friedman, B., & Hendry, D. (2012) The Envisioning Cards: A Toolkit for Catalyzing Humanistic and Technical Imaginations. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12). New York, NY: ACM, 1145-1148. doi:10.1145/2207676.2208562

Hassenzahl, M. (2013). Experiences Before Things: a Primer for the (Yet) Unconvinced. *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2059-2068). New York, NY: ACM. doi:10.1145/2468356.2468724

Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*, *3*(1), 1-95. doi: 10.2200/S00261ED1V01Y201003HCI008

Hassenzahl, M., Burmester, M., & Koller, F. (2003). *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*. In J. Ziegler, & G. Szwillus (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung* (pp. 187–196). Stuttgart, Germany: B.G. Teubner. doi: 10.1007/978-3-322-80058-9_19

Hassenzahl, M., Eckoldt, K., Diefenbach, S., Laschke, M., Lenz, E., & Kim, J. (2013). Designing moments of meaning and pleasure. Experience design and happiness. *International Journal of Design,* *7*(3), 21-31.

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behavior & Information Technology*, *25*, 91-97. doi: 10.1080/01449290500330331

Hornbæk, K. (2010). Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology*, 29(1), 97-111. doi:10.1080/01449290801939400

Huisman, G. (2009). *LEMtool – Visual Emotion Measurement in Digital Media.* Unpublished bachelor's thesis, University of Twente, Enschede, the Netherlands.

Hvannberg, E. T., Law, E. L.-C., Larusdottir, M. (2007). Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with computers*, *19*, 225-240.

International Organization for Standardization. (1998). ISO 9241-11:1998, *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.* Geneva, Switzerland: International Organization for Standardization.

Jordan, P. W. (2000). *Designing pleasurable products: an introduction to the new human factors*. London: Taylor and Francis.

Karapanos, E., Martens, J.-B., Hassenzahl, M. (2012). Reconstructing Experiences with iScale. *International Journal of Human-Computer Studies*, *70*(11), 849-865. doi:10.1016/j.ijhcs.2012.06.004

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J.-B. (2010). Measuring the dynamics of remembered experience over time. *Interacting with Computers*, 22(5), 238-335. doi:10.1016/j.intcom.2010.04.003

Kim, J., Park, S., Hassenzahl, M., & Eckoldt, K. (2011). The Essence of Enjoyable Experiences: The Human Needs - A Psychological Needs-Driven Experience Design Approach. In A. Marcus (Ed.) *Lecture Notes in Computer Science: Vol. 6769. User Experience, and Usability. Theory, Methods, Tools and Practice* (pp. 77-83). doi: 10.1007/978-3-642-21675-6_9

Kjeldskov, J., & Skov, M. B. (2003). Creating a Realistic Laboratory Setting: A Comparative Study of Three Think-Aloud Usability Evaluations of a Mobile System. In Proceedings of the 9th IFIP TC13 International Conference on Human Computer Interaction (Interact 2003). IOS Press, 663-670.

Kjeldskov, J., & Skov, M. B. (2007). Studying Usability In Sitro: Simulating Real World Phenomena in Controlled Environments, *International Journal of Human-Computer Interaction, 22* (1-2). doi: 10.1080/10447310709336953

Kjeldskov, J., Skov, M. B., Als, B. S., and Høegh, R. T. (2004). Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In Proceedings MobileHCI 2004, Berlin, Germany: Springer-Verlag, 529-535. doi: 10.1007/978-3-540-28637-0_6

Korhonen, H., & Koivisto, E. (2006). Playability heuristics for mobile games. Proceedings of the 8th conference on Human-computer interaction with mobile devices and services (pp. 9-16). New York, NY: ACM. doi:10.1145/1152215.1152218

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers, 23*(5), 473-483. doi:10.1016/j.intcom.2011.06.005

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:10.1016/j.chb.2014.10.048

Lallemand, C., Koenig, V., & Gronier, G. (2014). How Relevant is an Expert Evaluation of User Experience based on a Psychological Needs-Driven Approach? In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction*. New York, NY: ACM Press (pp. 11-20). doi:10.1145/2639189.2639214

Lallemand, C., Koenig, V., Gronier, G., & Martin, R. (in press). Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs.

Lavie, T., & Tractinsky, N. (2004). Assessing Dimensions of Perceived Visual Aesthetics of Web Sites, *International Journal of Human-Computer Studies*, 60(3), 269-298. doi:10.1016/j.ijhcs.2003.09.002

Law, E., Abrahão, S., Vermeeren, A., & Hvannberg, E. (2012). Interplay between User Experience Evaluation and System Development: State of the Art. International Workshop on the Interplay between User Experience (UX) Evaluation and System Development (I-UxSED 2012), Copenhagen, Denmark, 14-17 October 2012.

Law, E., Bevan, N., Christou, G., Springett, M., & Lárusdóttir, M. (2008). *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland.

Law, E., Roto, V., Hassenzahl, M., Vermeeren, A. & Kort, J. (2009). Understanding, scoping and defining UX: a survey approach. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New-York, NY: ACM Press. doi: 10.1145/1518701.1518813

Molich, R. et al. (1998). Comparative Evaluation of Usability Tests. *Proceedings of the Usability Professionals' Association Conference (UPA'98)*. Washington D.C. USA.

Nielsen, J. (1989). Usability engineering at a discount. In G. Salvendy and M.J. Smith (Eds) *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*. Amsterdam: Elsevier Science Publishers, 394-401.

Nielsen, J. (1993). *Usability Engineering*. New York: AP Professional.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'90), 249-156. doi:10.1145/97243.97281

Odom, W., & Lim, K.-Y. (2008). A Practical Framework for Supporting "Third-Wave" Interaction Design. Proceedings of alt.chi 2008, Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press.

Park, B. (2009). Psychophysiology as a Tool for HCI Research: Promises and Pitfalls. In J. A. Jacko (ed.), *Lecture Notes in Computer Science: Vol. 5610. Human-Computer Interaction. New Trends* (pp 141-148). doi: 10.1007/978-3-642-02574-7_16

Pruitt, J., & Adlin, T. (2010). *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. San Francisco: Morgan Kaufmann Publishers.

Rantavuo, H., & Roto, V. (2013). Heuristic Evaluation of User Experience - Case Nokia, Made for Sharing: HCI Stories of Transfer, Triumph and Tragedy. Workshop Paper presented at the *SIGCHI Conference on Human Factors in Computing Systems (CHI'13).*

Rogers, Y. (2011). Interaction design gone wild: striving for wild theory. *Interactions*, *18*(4), 58–62.

Rogers, Y., Connelly, K., Tedesco, L., Hazlewood, W., Kurtz, A., Hall, R.E., Hursey, J., & Toscos, T. (2007). Why it's worth the hassle: the value of in-situ studies when designing Ubicomp. In Proceedings of the 9th international conference on Ubiquitous computing (UbiComp '07). Berlin, Heidelberg: Springer-Verlag, 336-353.

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (2011). User Experience White Paper: Bringing clarity to the concept of user experience. *Result from Dagstuhl Seminar on Demarcating User Experience*, Sept. 15-18, 2010, Finland. http://www.allaboutux.org/uxwhitepaper

Roto, V., Obrist, M., Väänänen-Vainio-Mattila, K. (2009). *User Experience Evaluation Methods in Academic and Industrial Contexts.* In User Experience Evaluation Methods in Product Development (UXEM'09). Workshop in Interact'09, Sweden.

Schmettow, M., Bach, C., & Scapin, D. (2014). Optimizing Usability Studies by Complementary Evaluation Methods. In *Proceedings of the 28th British HCI Conference (BSC-HCI 2014).* Southport, UK.

Schneiderman, B. (2008). Science 2.0. *Science*, *319*, 1349-1350.

Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: A two process model. *Psychological Review*, *118*(4), 552-569. doi: 10.1037/a0024758

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, *89*, 325-339. doi: 10.1037//O022-3514.80.2.325

Sheldon, K. M., Ryan, R. M., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, *22*, 1270-1279. doi: 10.1177/01461672962212007

Silva, P. A., & Dix, A. (2007). Usability – not as we know it! In Proceedings of the 21st BCS HCI Group Annual Conference on People and Computers (BCS-HCI'07). Swinton, UK: British Computer Society, 103-106.

Simsarian, K. T. (2003). Take it to the next stage: the roles of role playing in the design process. Extended Abstracts on Human Factors in Computing Systems *(CHI EA'03)*. New York, NY, USA: ACM Press. doi:10.1145/765891.766123

Sutcliffe, A., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. *Interacting with Computers*, *4*(16), 831-849. doi:10.1016/j.intcom.2004.05.001

Tuch, A., Trusell, R., & Hornbæk, K. (2013). Analyzing users' narratives to understand experience with interactive products. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2079-2088). Paris, France: ACM. doi:10.1145/2470654.2481285

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185* (4157), 1124–1131. doi:10.1126/science.185.4157.1124

Väänänen-Vainio-Mattila, K., Roto, V., & Hassenzahl, M. (2008). Towards practical user experience evaluation methods. *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland.

Väänänen-Vainio-Mattila, K., & Wäljas, M. (2009). Development of Evaluation Heuristics for Web Service User Experience. *Extended Abstracts on Human Factors in Computing Systems (CHI EA'09)*. New York, NY, USA: ACM Press. doi: 10.1145/1520340.1520554

Venturi, G., & Troost, J. (2004). Survey on the UCD integration in the industry. In *Proceedings of the 3rd Nordic Conference on Human-Computer Interaction (NordiCHI'04)*. New York, NY: ACM Press, 449-452. doi:10.1145/1028014.1028092

Vermeeren, A., Cremers, A., Kort, J., & Fokker, J. (2008). Comparing UX measurements, a case study. *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, 18th June 2008, Reykjavik, Iceland.

Vermeeren, A, Law, E., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User Experience Evaluation Methods: Current State and Development Needs. *Proceedings of the 6th Nordic Conference on*

Human-Computer Interaction: Extending Boundaries, New York, NY: ACM Press. doi :10.1145/1868914.1868973

Vredenburg, K., Mao, J.-Y., Smith, P.W., and Carey, T. (2002). A Survey of User-Centered Design Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'02)*. New York, NY, USA: ACM Press, 471-478. doi:10.1145/503376.503460

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen and R. Mack (Eds.) *Usability Inspection Methods*. New York: Wiley, 105-140.

Wilson, C. (2011). Perspective-Based Inspection. 100 User Experience (UX) Design and Evaluation Methods for Your Toolkit. Available on http://dux.typepad.com/ (last verified 15 Dec 2014).

Woolrych, A., Hornbæk, K., Frøkjær, E., & Cockton, G. (2011). Ingredients and meals rather than recipes: a proposal for research that does not treat usability evaluation methods as indivisible wholes. International Journal of Human-Computer Interaction, 27(10), 940-970. doi: 10.1080/10447318.2011.555314

Yarosh, S., Matthews, T., & Zhou, M. (2012). Asking the Right Person: Supporting Expertise Selection in the Enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New-York, USA: ACM. doi:10.1145/2207676.2208382

Zaman, B., & Shrimpton-Smith, T. (2006). The FaceReader: measuring instant fun of use. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction (NordiCHI'06)*. New York, NY: ACM Press.

ENHANCING THE DESIGN PROCESS BY EMBEDDING HCI RESEARCH INTO EXPERIENCE TRIGGERS

F

Carine Lallemand, Kerstin Bongard-Blanchy, & Ioana Ocnarescu

Published 2014 (position paper)
Proceedings of Conférence Internationale Ergonomie et Informatique Avancée Ergo'IA 2014, Bidart-Biarritz, France.

Keywords: User Experience, Design, Experience Triggers, Materiality, HCI Research, Science-Based Design.

ABSTRACT

Over the last decade, User Experience (UX) has become a core concept in the field of Human-Computer Interaction (HCI). Beyond the fact of understanding and assessing the User Experience derived from the use of interactive systems, practitioners and researchers from a wide range of disciplines are now facing the challenges of *designing for User Experience*.

Some authors have pinpointed the existence of a gap between the theoretical knowledge developed in HCI Research and the practical knowledge actually used by designers to create rich experiences with interactive artefacts. A special focus of this paper is to translate theoretical work into experiential objects (or situations) called "Experience Triggers". Through their materiality, these artefacts bring emotions and sensations to the design process and designers can immerge into and understand the theories on experience. As a consequence of this immersion, the final product designed by the team is assumed to be more experiential. Experience Triggers are introduced here as a new tool for science-based UX design.

1. INTRODUCTION

In a world moving from a materialistic view to an ever-growing experiential perspective (Hassenzahl, 2013b; Pine & Gilmore, 1998), designing (for) User Experience (UX) has become a major concern for both researchers and practitioners (Hassenzahl & Tractinsky, 2006). However, the complexity of designing experiences, and even of knowing which kind of experiences are desirable or not in specific contexts of use, is a daily challenge that experience designers have to cope with. Despite the availability of theoretical and empirical findings on the way people interact with and therefore experience the world and its artefacts, only few artefacts designed within the HCI field are actually explicitly rooted on this body of knowledge (Hassenzahl, Heidecker, Eckoldt, Diefenbach, & Hillmann, 2012).

This obviously leads to questioning the collaboration between several disciplines sharing the common goal of designing interactive products or systems, able to stimulate positive user experiences. Design, Ergonomics and HCI are all at the core of interaction design. In recent approaches that are more and more interdisciplinary, these disciplines profit from each other and we see new methodologies and theoretical frameworks, which do not exclusively relate to one of these domains only (Lazar, Feng, & Hochheiser, 2010). It is obvious both researchers and practitioners have started profiting from cross-fertilization between those domains. It remains unclear however how each domain actually contributes to an integrated design process in order to support UX design and to what extent there remain domain-specific approaches (Wright & McCarthy, 2010). Moreover, studies have shown the existence of differences between academia and industry both in the understanding of UX and the underlying UX design practices (Hassenzahl et al., 2012; Lallemand, Gronier, & Koenig, 2013). Researchers in the field of UX are seeking to understand the nature of human experiences and the drivers of positive experiences with technologies. Models and theories of UX are developed and tested. Unfortunately, this ever-growing body of knowledge developed within the HCI and UX research fields seems to be actually underused by designers in practice (Hassenzahl et al., 2012).

To address this issue, the concept of "Experience Triggers" has been proposed as a promising approach (Bongard-Blanchy, Lallemand, Ocnarescu, & Labrune, 2013). Experience Triggers (E.T.) are defined as objects or situations created for the design team to influence the design process by embedding design guidelines and various theories of experience. It is therefore assumed that the use of Experience Triggers within the design process could help designing for an optimal UX of the final product.

In this paper, we first examine how the design process integrates the focus on user experience. We show that design practitioners underuse the existing body of knowledge on UX. As a consequence, we introduce the concept of Experience Triggers as a potential bridge between UX research and UX practice. The benefits of Experience Triggers are presented in relation to the design process and its result. We then propose a first experiment to test

the effectiveness of this approach. Finally, we discuss perspectives and challenges related to Experience Triggers as a UX design tool.

2. DESIGNING (FOR) USER EXPERIENCE

User Experiences are experiences created and shaped through technology (Hassenzahl, 2013). Designing for User Experience is frequently considered as a challenge (Hassenzahl, 2013; McCarthy & Wright, 2004) since it goes beyond the quality or originality of the design and involves a deep understanding of the way technology involves people emotionally, intellectually and sensually (McCarthy & Wright, 2004). As stated by Hassenzahl (2013), "UX is not about good industrial design, multi-touch, or fancy interfaces. It is about transcending the material. It is about creating an experience through a device". During the last decade, several theoretical models have been developed (Forlizzi & Ford, 2000; Hassenzahl, 2010; Mahlke, 2008) to account for the complexity of user experience. In 2007, authors of the User Experience Manifesto (Law, Vermeeren, Hassenzahl, & Blythe, 2007) stated, "developing theoretically sound methodologies for analysing, designing, engineering and evaluating UX should be high in the UX research agenda". Since then, several UX evaluation and design methods have been developed and applied in research. It remains however unclear to what extent these methods have been transferred into daily practice by UX professionals and thus needs to be assessed.

We will start by trying to analyse how practitioners do design to show what could be improved. Noteworthy is that the population of practitioners working as "UX Designers" is highly heterogeneous (Lallemand et al., 2013). A majority of UX Practitioners have been educated in one of those four fields: Design, Psychology/Social Sciences, Technology/Software or Human-Computer Interaction (Lallemand et al., 2013). It is therefore not an easy task to describe design in practice as each field might apply specific design processes. In this paper, we will focus on practitioners educated in the field of Design.

The design process is traditionally constituted of four phases: Exploration, Generation, Evaluation, and Communication (Cross, 2008). During the exploration phase, designers gather information related to the design brief, the user, and sources of inspiration. The generation phase consists in the creation of ideas, mainly through sketches, storyboards, wireframes and mock-ups. The evaluation phase consists of selecting the most appropriate solution(s) among the generated ideas. Finally the chosen solution is communicated to the development team and to clients before entering the product development. Designing for UX is an iterative process with multiple feedback loops between the development and evaluation phases. Each of the design process phases has its specific tools of which a certain number comes from the HCI or Ergonomics domain and others that have been developed by the design community.

First of all, during the exploration phase, the design team seeks to understand who future users are and what their use context looks like. Classical methods used in Ergonomics like field studies, interviews or Focus Groups (Bruseberg & McDonagh, 2001) help gathering explicit or

observable information. Observations can also be done indirectly through diaries and camera journals (Lallemand, 2012; Lazar et al., 2010). These fields-based approaches aimed at understanding users were accompanied by the development of "Day in the life" scenarios (Moll, 2006) or Personas (Pruitt & Adlin, 2006) in order to make the gathered information tangible for practitioners. These methods notably helped field researchers communicate their findings to designers (Weisberg, 1999). Moreover, designers also need information on tacit or latent user desires (Visser, 2009). The design domain has therefore developed own tools like Design Probes (Gaver, Dunne, & Pacenti, 1999; Wallace, McCarthy, Wright, & Olivier, 2013) or Role-Playing (Stappers & Sanders, 2003), which give the future user an active role to play.

Once the user and use context are explored, the identified needs and desires have to be translated into design ideas. So far there are no explicit tools used for this UX generation phase. Most commonly used are creativity sessions expected to stimulate UX idea generation through mind maps or brainstorming (Cross, 2008; Goncalves, Cardoso, & Badke-Schaub, 2014). Another way to bring the User Experience into early design steps is by inviting users to join the generation process, the so-called participatory design. The goal is to initiate a dialogue between the designer and the user (Prahalad & Ramaswamy, 2004; Sanders & Stappers, 2008). However, common design generation tools, like sketch or wireframing, simply rely on the empathic capacity (i.e., to project oneself into the user and use context knowledge while conceiving the interface) of the designer.

Finally designers have to choose the most promising solutions from the set of ideas developed so far. To do so, they often follow their instinct, their project leader or the client's choice while more objective UX evaluation is possible at this stage. Mainly the psychology components in HCI provide UX design with a range of tools to test whether or not a design idea is able to trigger the desired UX. User responses to stimuli come in form of emotions, sensations, accorded meaning, etc. These can be measured on three levels: cognition/language, behavioural events, and physiological events (Bradley & Lang, 2000). Conscious UX is often measured with self-evaluation questionnaires (Desmet, 2002; Lazar et al., 2010). The behavioural dimension of UX has so far been the core of Ergonomics. Task analysis and user testing are classical means to measure effectiveness and efficiency (Tullis & Albert, 2008). Tools like eye-tracking show which properties of the design the user perceives. Last but not least, physiological parameters like body temperature, heart rate, breath rhythm, sweating, etc. (Scherer, 2005) as well as facial and other somatic muscle movements can be indicators for UX design (Bradley & Lang, 2000). They provide data on arousal or valence (i.e., positive vs. negative feelings) evoked by a stimulus.

To summarize, this analysis of design in practice allows us to see that UX Design benefits from numerous tools coming from Ergonomics, Design and HCI. Designers already employ various UX specific tools and methods, especially during the exploration and evaluation phases. However, there seems to be a lack of tools to support practitioners in the design of UX for the generation phase. This is not surprising considering that, even in research, few methods only exist to design for UX. Amongst them, the use of *Experience Patterns* (Hassenzahl et al., 2013) or *Needs-Driven Experience*

Design approaches (Kim, Park Hassenzahl, & Eckoldt, 2011) are rarely known by designers. Another observation that can be made is that practitioners are developing and using many design methods that have never been rigorously tested (Dorst, 2008). Design practice and HCI research could benefit from a closer integration. HCI research can help designers in the rigorous development of new methods, while designers can provide researchers with industrial use cases to test and enhance design methods.

2.1. SCIENTIFIC GROUNDING OF CURRENT PRACTICE IN EXPERIENCE DESIGN

Experience Design requires a deep understanding of people (Hassenzahl, 2010), their cognitive and affective processes (e.g., such as cognition, affects, motivation and volition) and basic needs. A profound theoretical and empirical knowledge on the understanding of human experiences has been accumulated through decades of research in several fields such as psychology, sociology, ethnology, philosophy, etc. Universal human needs have for example been identified and thoroughly investigated (Sheldon, Elliot, Kim, & Kasser, 2001), as well as psycho-cognitive and psycho-social processes, human values (Rokeach, 2000), human emotions (Fridja, 1986; Scherer, 2005) or even optimal experiences at a more generic level (Csikszentmihalyi & Lefevre, 1989). All these considerations are crucial when studying human experiences and especially, within the HCI research field, user experiences with interactive systems and artefacts.

Unfortunately, despite the availability of theoretical and empirical psychological findings, it seems that existing knowledge remains largely underutilized by designers. In a systematic review of 92 publications presenting 143 artefacts from the HCI and Interaction Design domain, Hassenzahl et al. (2012) show that less than half of those make explicit use of external theoretical and empirical psychological knowledge. This might be explained by a commonly shared "bottom-up approach to the analysis of people and contexts [...] (where) designers immerse themselves into the context to build up the empathy necessary for sensible design".
Even if the HCI community is highly interdisciplinary by nature, another reason explaining this phenomenon could be the feeling of incompetence to master concepts from other disciplines, especially those studying the human with regard to his full complexity, such as psychology. Moreover, setting academic and research areas aside, it is easily understandable that designers (as practitioners) may neither have full access to this body of knowledge nor have the time to get acquainted with and use it within their designs (Goncalves et al., 2014). Finally, designers might also fear to constrain their creativity and inspiration if relying on theoretical knowledge instead of listening to their sensitive empathic feelings towards potential future users.

In summary, despite the fact that designers adopt more and more of the tools and methods developed by the research community for the exploration and evaluation phases, there are few tools only specifically developed for User Experience generation. Therefore we propose "Experience Triggers" as one way to transfer HCI knowledge to designers in an attractive way.

3. EXPERIENCE TRIGGERS: EXPERIENCING TO DESIGN BETTER EXPERIENCES

The concept of "Experience Triggers" is based on the assumption that experiencing through materiality will help designing better experiences. Living a specific experience before or during the design process might unconsciously help designers to develop an intuitive and empathic knowledge about the experience(s) an object can evoke. Once the experiential purpose is understood (even intuitively) by the designers, we assume they are more likely to find ways of expressing and designing this specific experience through the interactive object or system. Our idea is to speak the same language as designers, materiality being a potential medium to reach this goal. The ability of artefacts to embody and thus mirror theoretical notions, concepts and empirical findings became therefore the core idea of Experience Triggers as an inspirational tool for UX generation.

3.1. WHAT ARE EXPERIENCE TRIGGERS AND WHAT BENEFITS DO THEY BRING TO THE DESIGN PROCESS?

Experience triggers (E.T.) are objects or situations created by a UX expert, whose goal consists of embedding specific theories of experience within those objects or situations.

E.T.s serve three purposes in the design process:

1. to bridge theory and practice by providing designers with new knowledge in an informal way, that is intuitively integrated and does not constrain creativity;
2. to enable designers to experience the type of experience they seek to create;
3. to unite the design team around a reference experience.

Experience Triggers are introduced within the design process during the early steps of idea generation. They are intended to help designers understanding a specific theory of experience without actually having to read anything formal about it. E.T. will bring new knowledge on human experience that will potentially be intuitively integrated and does not constrain the creativity of designers. Being informal, this embedded knowledge is also less likely to hinder the empathy developed by designers using a bottom-up approach to the analysis of people and contexts. By interacting with E.T. we therefore expect designers to gain new insights on how to design positive experiences, for example how interactive products might support final users in the fulfilment of their primary needs (Hassenzahl, 2010; Sheldon et al., 2001). Moreover, the use of E.T. might also trigger the designers' interest in theories and encourage them to further explore the literature on human experience.

The main assumption behind E.T. is that designers will design better product experiences if they have been previously (i.e., during the design process) immersed in the experience they seek to create. Lived experiences are hard to express and to understand using words and the vocabulary often

fails to transmit with accuracy feelings or emotions. People thus frequently feel that their peers or beloved ones are not able to understand what they feel during a memorable event unless they live the same experience. Placing designers in a situation of intense relatedness or making them feel the optimal experience of flow (Csikszentmihalyi & Lefevre, 1989) might be a good way to help them transfer this concern in the objects or system they are designing. As inspirational objects, E.T. could stimulate designers to feel something and then do something creative by analogy, so that final users will feel the same experience. In design, analogical reasoning plays a double role by supporting creativity and learning simultaneously (Gentner & Toupin, 1986). For now, design by analogy has been focused mostly on visual, textual or functional analogy (Cross, 2008; Goldschmidt & Sever, 2011). In our case, we intend E.T. to act as "experiential analogies". It is the felt experience of the final product that is meant to be designed by analogy. Imagining different sets of E.T. to reflect the numerous existing but unfortunately underused theories of UX would therefore be a good way to enhance design practice.

As experiential objects, beyond the fact of embedding a theory of experience (or some elements of a theory), Experience Triggers are meant to boost the creativity of the design team and stimulate a better group dynamic. Indeed, as it is not that trivial to embed a UX theory in a common object, E.T. are likely to be designed under the form of very peculiar objects or situations. This assumption has been explored during a first workshop (Bongard-Blanchy et al., 2013) where we asked 35 participants (all of them being UX practitioners or researchers) to act as E.T. designers. Working in small teams during about an hour, participants had to study four specific human needs: *security relatedness, pleasure* and *self-esteem* adapted from Sheldon et al. (2001) and to come up with tangible objects embedding this experience. Two of the resulting E.T. objects are presented in Figure 1. Participants from other groups tested the designed E.T.

Despite the very explorative nature of this first experiment, we were able to witness the potential power of E.T. to have an effect on designers by triggering *something* (at this point we are not able to characterize exactly what kind of feelings/emotions or experiences have been triggered) and stimulate a reflection. Research has shown that the more a lived experience is interesting, intense, confusing or impressive, the more humans feel the need to talk about it and to share it with others (Hassenzahl, 2013a). Sharing experiences has a high social value and helps feeling related to others (van Boven & Gilovich, 2003). Therefore, we expect the possibly complex and unusual nature of E.T. to be experiential in itself and hence likely to foster discussions and debates between team members. Several design tools already use materiality to inspire design teams and create a shared experience, like for example the well-known design probes (Gaver et al., 1999) or the open-ended objects imagined by Cruz and Gaudron (2010).

Figure 1. Examples of draft Experience Triggers created during the UX Workshop. Left: 'security' E.T., right: 'relatedness' E.T.

Thanks to their experiential nature, E.T. are also meant to resonate with the personal history of the designer. Each E.T., by triggering a specific experience, inevitably relates to the identity of the designer using it. It will evoke memories of objects or previous experiences and will therefore help designers to rely on past experiences and personal history to get design inspiration. As an UX ideation tool, E.T. can be classified as an intuitive approach, meant to help designers to "break routines and overcome mental blocks" (Goncalves et al., 2014).

3.2. THE DESIGN OF EXPERIENCE TRIGGERS

The design of Experience Triggers basically requires three main elements:

1. a UX expert, who designs a single or a set of E.T. (i.e., objects or situations) assisted by an artist or designer for the creation of the object.
2. a theory of experience to be embedded or partly embedded within this set of E.T.
3. a methodology or guidelines to guarantee the coherence and effectiveness of the process.

Experience Trigger designers are specialists in human experience and could be new actors in the design process. Their role is to embed design guidelines and theories of experience within objects or situations in order to influence the design process. In that sense, Experience Triggers act as tangible translations of a specific body of knowledge on Experience Design. Experience Trigger designers might be considered as "Meta-Designers", since they will not directly be involved in the design of a specific product or system but will influence the whole process by providing the design team with one or several particular E.T. A solution to cope with the fact that experts having both fundamental (i.e., being an expert at a theoretical level) and creative skills (i.e., being able to design an object) are hard - if not impossible - to find, would be a collaboration between a specialist in human experience and an artist or creative designer.

There is no comprehensive list of theories of experience that the E.T. designer might be willing to embed in E.T.; they might be as diverse as the theories of user experiences. Depending on his background and knowledge, the E.T. designer may use theories from HCI, Design and Ergonomics but also from disciplines such as psychology, social sciences, cognitive sciences or even biology.

Last but not least, designing Experience Triggers requires a process to support E.T. designers in their work (see Figure 2). We propose the following methodology as a starting point:

1. *Selection of a theory on user experience*: the E.T. designer selects a theory to be embedded into the E.T.
2. *Extraction of key elements*: the E.T. designer extracts key findings of the theory. If working in collaboration with a designer or an artist, he transcripts these key findings under an easy-to- understand form.
3. *Idea generation*: the E.T. designer (and his collaborators) explores concrete and tangible experiential translations of the theory. These might be physical objects or role-playing situations.
4. *Creation of one or several E.T.*: the E.T. designer and his collaborator choose the most appropriate form for their E.T. and generate one or several E.T.s.
5. *Pre-Evaluation of E.T.*: before providing a design team with the novel set of E.T., a pre- evaluation on a control group is performed to ensure that the E.T. truly triggers the intended experience.
6. *E.T. in use*: the design team interacts with or manipulates the Trigger object during their design activity.
7. *UX evaluation of the final design*: test if the intended experience was translated into a design solution.

Figure 2. Experience Triggers Methodology

Now that we have presented the concept of E.T., the rationale behind the development of this new method and the main expected outcomes, we need to assess the potential of E.T. as a new tool to design for UX.

3.3. ASSESSING THE POTENTIAL OF EXPERIENCE TRIGGERS AS A NEW TOOL TO DESIGN FOR UX

During our preliminary workshop on Experience Design organized in Paris during the FLUPA UX-Day (Bongard-Blanchy et al., 2013), the community of French UX practitioners showed a strong interest for the concept of "Experience Triggers". Feedback gathered during a short post-task questionnaire shows that the Triggers were perceived as a potentially valuable bridge between research and design practice. However, this exploratory experiment does not allow us to claim any benefit provided by Experience Triggers. A more thorough and scientific-based experiment is planned within the next few months.

To assess the potential of Experience Triggers, one should positively answer two main questions:

- Is the final product better than it would have been without Experience Triggers?
- Is the design process experience more memorable (i.e., will be remembered as a particularly positive experience) than it would have been without Experience Triggers?

The goal of the study is to test the potential of the E.T. by comparing three design teams (3-4 members per team) during a design challenge. All teams will work on the same design brief, a short written document focused on the desired results of design. This document also includes basic data on target users and the context of use. The type of object or system to be designed will therefore be determined beforehand. The theory of UX that will be embedded in E.T. for this experiment is derived from the Psychological Needs-Driven Experience Design approach (Kim et al., 2011), which states that technology shapes experiences through fulfilling (or not fulfilling) certain psychological needs. The experiment will be an independent measures design involving three independent groups. A first team will act as a control group and will therefore design a specific object only by using their knowledge and expertise. A second team will design the same kind of object by using UX Cards (which are cards formally describing a theory of UX). These UX Cards (designed by the first author of this paper) constitute an intermediate condition, using a formal source of knowledge about human experience. They will allow us to see how designers welcome formal UX theories and how textual stimuli will influence idea generation. Previous findings seem contradictory on this issue, some studies arguing that text stimuli may have negative effects (Malaga, 2000) while some others (Goldschmidt & Sever, 2011) show a positive influence of word stimuli as compared to no external stimuli. Finally the third team will design the object by using one or several Experience Triggers embedding the same UX theory as the UX Cards provided to the second team. In order to control for biases, a special attention will be paid to the homogeneity of the groups, especially in terms of seniority of team members, educational background, previous collaboration experience between members, etc.

After the design task, the experiential potential of each designed objects or system will be assessed both by UX experts and potential final users. Qualitative and quantitative measures will be used for this assessment. We also assess the experience lived by the members of each design team in order to know whether the presence and use of E.T. contribute to foster discussions and creativity within the design team. This planned experiment should show how effective E.T. are to enhance the design process.

While this preliminary work on the concept of Experience Triggers is very motivating and might lead to promising outcomes for the practice of design, we are aware that a lot of critical questions on this new tool remain unanswered at this stage. We discuss some of these issues in the last part of this paper.

4. DISCUSSION

The rationale behind the idea of E.T. is the existence of a gap between research and practice (Hassenzahl et al., 2012; Lallemand et al., 2013) that we would like to reduce by translating the theoretical body and methodologies of researchers into the language of designers. However, we are aware that this translation is very challenging. Experiential objects are already hard to create without having to embed any specific theoretical knowledge. Since these kinds of objects or situations should trigger rich experiences specific to a future product or service, one could raise further questions: should every company invent its own E.T.s based on theories that seem relevant to its projects? Or should E.T.s be universal and only specific to UX theories? For now it is indeed impossible to say if E.T. will be valuable for every designer, dealing with every possible design problem in any design context. As stated by Dorst (2008), design research should not only focus on design processes, as if they would be universally valid for each design context, but also on a "deep and systematic understanding of the design object, the designer and the design context" (p.6).

Another question raised by E.T. is their potential of stimulating analogical reasoning, i.e. being used as examples of what should be lived and felt by final users through the use of the product. Our assumption is that designers live specific a kind of experience and then, by analogy, intuitively embed this specific experience in the final product, so that final users will live the same experience. In a study on the influence on analogies during idea generation, Casakin (2005) shows that designers are stimulated by visual analogues (i.e., pictures) without any instruction to use analogical reasoning. Participants to our experimental study will therefore not be instructed to use the felt experience to design by analogy. After this first experiment, a considerable amount of work will be needed to explore the conditions required for E.T. to be an effective design tool.

Furthermore, E.T. as tangible objects also carry the risk of subjective interpretation of the intuitive knowledge they are supposed to embed. The question is: do E.T. trigger the same experience for all members of a team design? Based on research findings on inspiration in design (Eckert, Stacey, & Clarkson, 2000; Weisberg, 1999), the answer would be "not really". Each E.T. will resonate differently for each person, depending on her personal history and sensitivity (Weisberg, 1999). We do not fear subjectivity in the interpretation of E.T. as we consider this as a positive outcome that might foster dynamic group discussions and idea generation. However, this dynamic is only one goal out of the three main goals assigned to E.T. Despite its interest, it is probably the easiest goal to achieve and numerous existing tools and design methods succeed in doing so. The biggest goal that might be hindered by this subjective interpretation of E.T. is the one of knowledge embedding. We do not intend the E.T. to trigger the same singular feelings for each person, but we do intend that these feelings relate to the same kind of experience. For example, an object might be considered as aesthetic for one person and anaesthetics for another depending on their personal taste. This subjective assessment is an inevitable process and might be seen as a critical issue for the design of E.T. However, we believe that this is not a problem if the main goal of the object was to embed the notion

of visual pleasure, the feeling of visual displeasure being one possible expression of this global intended experiential notion. No matter if some designers have experienced visual pleasure while others have experienced displeasure as long as all of them have intuitively understood the importance of including visual pleasure as an experiential quality of their final design.
The experiment we intend to conduct within the next few months will be a first step to explore the potential of E.T. as a new UX design tool. We hope to be able to analyse the way E.T. impact the design process and the quality of design outcomes. We also aim at finding ways of improving this method by understanding design mechanisms involved in the use of E.T.

5. CONCLUSION

In the current *experience economy* (Pine & Gilmore, 1998), designing rich or memorable user experiences has become a key goal to achieve when designing interactive products. For more than a decade, research studies are conducted to understand the mechanisms underlying user experiences and to develop UX evaluation and design methods. However, as we have seen, the bridge between research and practice, as well as the effective integration of several disciplines in the design process, is not yet fully successful.

Experience Triggers are introduced here as a new promising tool for the design of UX. By embedding some theoretical knowledge about user experiences within artefacts or situations to be used or lived by designers, we hope to enhance the quality of both the design process and the design outcomes. The concept of E.T. is only in its early stages and numerous challenging questions are raised. An experimental study will be conducted to bring understanding about the benefits, limitations and prerequisite of E.T. design and usage as a UX design tool.

REFERENCES FOR PAPER F

Bongard-Blanchy, K., Lallemand, C., Ocnarescu, I., & Labrune, J.-B. (2013). *Workshop on User Experience Design*, FLUPA UX-Day 2013, Paris.

Bradley, M. M., & Lang, P. J. (2000). Measuring Emotion: Behavior, Feeling, and Physiology. In Cognitive Neuroscience of Emotion (pp. 242–276). New York: Oxford University Press.

Bruseberg, A., & McDonagh, D. (2001). New product development by eliciting user experience and aspirations. *International Journal of Human-Computer Studies*, *55*(4), 435–452.

Casakin, H. (2005). Design aided by visual displays: a cognitive approach. *The Journal of Architectural and Planning Research*, *22*, 250-265.

Cross, N. (2008). Engineering Design Methods. (Wiley-Blackwell, Ed.). Chichester: John Wiley & Sons.

Cruz, V., & Gaudron, N. (2010). Open-ended objects: a tool for brainstorming. Proceedings of the 8th ACM Conference on Designing Interactive Systems.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology. 56* (5), 815–822.

Desmet, P. (2002). *Designing Emotions*. TU Delft.

Dorst, C.H. (2008). 'Design Research: A revolution-waiting-to-happen', *Design Studies*, *29* (1), 4-11.

Eckert, C. M., Stacey, M. K., & Clarkson, P. J. (2000). Algorithms and inspirations: creative reuse of design experience. In Proceedings of the Greenwich 2000 Symposium: Digital creativity (pp. 1-10).

Forlizzi, J., & Ford, S. (2000). The building blocks of experience: an early framework for interaction designers. *Proceedings of the 3rd conference on Designing interactive systems : processes, practices, methods, and techniques (pp. 419-423). New York City, New York, United States:ACM.*

Frijda, N.H. (1986). *The Emotions*. Maison des Sciences de l'Homme and Cambridge University Press.

Gaver, W., Dunne, A., & Pacenti, E, (1999) Design: Cultural probes, *Interactions*, *6* (1), Jan/Feb.
Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10* (3), 277-300.

Goldschmidt, G., & Sever, A. L. (2011). Inspiring design ideas with texts. *Design Studies*, *32* (2), 139- 155.

Goncalves, M., Cardoso, C., & Badke-Schaub, P. (2014). What inspires designers? Preferences on inspirational approaches during idea generation. *Design Studies*, *35*, 29-53.

Hassenzahl, M. (2010) Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*, *3* (1), 1- 95.

Hassenzahl, M. (2013a). User Experience and Experience Design. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "*The Encyclopedia of Human-Computer Interaction*, 2nd Ed.". Aarhus, Denmark: The Interaction Design Foundation.

Hassenzahl, M. (2013b). Experiences Before Things: a Primer for the (Yet) Unconvinced. *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 2059-2068.

Hassenzahl, M., Eckoldt, K., Diefenbach, S., Laschke, M., Lenz, E., & Kim, J. (2013). Designing moments of meaning and pleasure. Experience design and happiness. *International Journal of Design*, *7*(3), 21-31.

Hassenzahl, M., Heidecker, S., Eckoldt, K., Diefenbach, S., & Hillmann, U. (2012). All You Need is Love: Current Strategies of Mediating Intimate Relationships through Technology. *ACM Transactions on Computer-Human Interaction*, *19*, 4.

Hassenzahl, M., Tractinsky, N. (2006). User experience - a research agenda. *Journal of Behaviour & Information Technology*, *25* (2), 91–97.

Kim, J., Park, S., Hassenzahl, M., & Eckoldt, K. (2011). The Essence of Enjoyable Experiences: The Human Needs Design, User Experience, and Usability. Theory, Methods, Tools and Practice. In A. Marcus (Ed.),Vol.6769, pp.77-83, Springer Berlin Heidelberg.

Lallemand, C. (2012). Using diaries to study UX: Explication, Application, *User Experience*, *11* (3).

Lallemand, C., Gronier, G., & Koenig, V. (2013). L'expérience utilisateur : un concept sans consensus ? Enquête sur le point de vue des professionnels, EPIQUE 2013, Bruxelles, Belgique.

Law, E., Vermeeren, A. Hassenzahl, M., & Blythe, M. (2007). Towards a UX Manifesto. *COST294- MAUSE affiliated workshop*, Lancaster, UK.

Lazar J., Feng J., & Hochheiser H. (2010) *Research Methods in Human-Computer Interaction*. London: Wiley.

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.

Malaga, R. S. (2000). The effect of stimulus modes and associative distance in individual creativity support systems. *Design Support Systems*, *29*, 125- 141.

McCarthy, J.C., & Wright, P.C. (2004) *Technology as Experience*. MIT Press.
Moll, B. (2006). Scenarios – Recording a Day in the Life. Pathfinder. Retrieved from http://pathfindersoftware.com/2006/06/scenarios_recor

Nieminen, M., Runonen, M., Nieminen, M., & Tyllinen, M. (2011). Designer experience: exploring ways to design in experience. *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 2449-2452.

Pine, B.J., & Gilmore, J.H. (1998). Welcome to the Experience Economy: work is theatre & every business a stage. *Havard Business Review*, July-August 1998.

Prahalad, C. K., & Ramaswamy, V. (2004). Co-creation experiences: The next practice in value creation. *Journal of Interactive Marketing*, *18*(3), 5– 14.

Pruitt, J., & Adlin, T. (2006). *Persona lifecycle*. San Francisco: Morgan Kaufmann.

Rokeach, M. (2008). *Understanding human values: Individual and societa*l. New York: Free Press.

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.

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.

Simsarian, K. T. (2003). Take it to the next stage: the roles of role-playing in the design process. In *Proceedings of CHI*, pp. 1012-1013, ACM.

Stappers, P., & Sanders, E. (2003). Generative tools for context mapping: tuning the tools. In *Design and Emotion*. Loughborough: Taylor & Francis.

Tullis, T., & Albert, W. (2008). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Burlington, MA: Morgan Kaufmann Publishers.

Van Boven, L., & Gilovich, T. (2003). To do or to have? That is the question. *Journal of Personality and Social Psychology*, *85*, 1193-1202.

Visser, F. S. (2009). Bringing the everyday life of people into design. TU Delft.

Wallace, J., McCarthy, J., Wright, P. C., & Olivier, P. (2013). Making design probes work. In *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI'13* (p. 3441). Paris: ACM Press.

Weisberg, R. (1999). Creativity and knowledge: a challenge to theories. In R. J. Sternberg (Ed.), *Handbook of creativity*. Cambridge: Cambridge University Press.

Wright, P.C., & McCarthy, J.C. (2010). Experience-Centered Design: Designers, Users, and Communities in Dialogue. Synthesis Lectures on Human-Centered Informatics. Morgan & Claypool Publishers.

Zimmerman, J., Forlizzi, J., & Koskinen, I. (2009). Building a Unified Framework for the Practice of Experience Design. CHI'09 Extended Abstracts on Human Factors in Computing Systems.

Part III

GENERAL DISCUSSION

GENERAL DISCUSSION

In this last chapter, we will first present a summary of the main contributions of this thesis to UX research and practice. We will then report on the limitations of the present work and describe perspective for future research. Finally, we will adopt a high-level perspective and provide a critical discussion on the necessity of UX research consolidation.

During the past years, the research community has been developing a wealth of methods, techniques and tools to address the new challenges brought by the topic of UX. Great dynamics have grown and have led to interesting, inspiring, and often creative outcomes. Many methodological issues however have been identified and have to be addressed in order to maximize the effectiveness and strategic impact of UX research. As highlighted in the first chapter of this thesis, major limitations of current UX design and evaluation methods are related to their scientific quality and the scarcity of research consolidation initiatives with regards to UX methods. The contributions of this thesis are oriented towards consolidated methods for UX design and evaluation. We furthermore strive for raising more awareness on consolidation-related issues and encouraging consolidation practices within the UX community.

1. CONTRIBUTIONS OF THIS THESIS

In this thesis, we adopted a high-level perspective on UX design and evaluation methods in order to inform research consolidation strategies.

The contributions of our work encompass:

- Theoretical findings on UX;
- Methodological contributions under the form of UX design and evaluation tools, as well as empirical findings on the relevance of established HCI methods for the evaluation of UX;
- Direct outputs and relevant implications for UX practice;
- A critical discussion on consolidation strategies and their necessity in UX research.

In the following, we present each contribution as a dedicated subsection.

1.1. THEORETICAL CONTRIBUTIONS

While the overall focus of the present work is predominantly methodological, the studies we conducted also make relevant theoretical contributions to UX research.

In our first study (paper A), we analyzed practitioners' perspectives on UX and the evolutions over time in the understanding and practice regarding this concept. Through this replication study, we confirmed most previous theoretical assumptions from literature on the perceived nature of UX as being unique to an individual, influenced by the context and highly

dynamic. We also revealed substantial differences in the understanding of UX between academia and industry and contributed, through a comparative analysis with the original survey, to uncover the trends and changes in UX practice.

The main theoretical contribution of our second study (Paper B) lies in the validation of Hassenzahl's model (2003) on a French-speaking sample. Through the translation of the scale, we were able to confirm in our sample the soundness of the distinction between perceived hedonic and pragmatic qualities, which combine to form an overall judgment of attractiveness. The cross-cultural validation of theoretical models is valuable to explore the generalizability of findings and to better understand the processes underlying UX.

Finally, our third and fifth studies (Paper C and E) provide numerous findings on factors impacting UX. In Paper C, we found strong evidence that contextual factors impact the assessment of UX, which confirms previous findings. But most of all, we gained better insight on how this impact of context on UX works. The fine-grained categorization of contextual dimensions in our UX context scale provided many insights into how the interaction context impacts the felt experience, with regards to both objective and subjective contextual attributes. In Paper E, we gathered in-depth data on psychological needs fulfillment and the links between reported needs fulfillment and UX.

1.2. METHODOLOGICAL CONTRIBUTIONS

The primary objective of this thesis was to contribute to the consolidation of UX design and evaluation methods. We addressed this goal through the development of design and evaluation tools, and through empirical research on UX evaluation.

1.2.1. DESIGN AND EVALUATION TOOLS

First, we have undertaken the translation and validation of a French version of the AttrakDiff 2 scale (Paper B). While UX evaluation is a core concern within the HCI field, there was no valid self-administered UX evaluation tool in French. This absence of a valid research instrument to explore perceived UX qualities of a system was a blocking issue to study UX on French-speaking samples. In terms of measuring the UX, our French version of the AttrakDiff scale turns out to be internally consistent and to support Hassenzahl's model of UX (2003). Even though they did not significantly alter the global quality of the scale, three problematic items were identified and still have to be double-checked through additional investigations. Our study on the Attrakdiff scale provides the French-speaking research community with a reliable quantitative evaluation instrument. This will allow conducting research on UX more efficiently and in a more rigorous way, while at the same time also providing a basis for cross-cultural comparison studies. It is worth mentioning that we also realized during this study that the validation study of the initial German version of the AttrakDiff questionnaire (Hassenzahl et al., 2003; Hassenzahl, 2004) actually presented limitations with regards to statistical

analyses (sample size being too small for conducting a factor analysis). While the German questionnaire has been considered by the research community as a validated tool, this observation confirms the need for more methodological consolidation and better publication standards.

Second, we designed the UX context scale (UXCS), a self-reported questionnaire for the evaluation of the contextual dimension impacting user experience. While the context of use has been highlighted for many years as a key factor impacting UX, there was no tool so far based on self-reported quantitative data that could support an explicit context investigation. The creation of this tool thereby contributes to filling the gaps in UX evaluation questionnaires. With the ever-growing trend for mobile and innovative technologies, knowing how the interaction context impacts end-users' experience is of prime importance. The UXCS supports both a measure of objective contextual factors and a measure of perceived context properties through seven subscales: *Physical Context*, *Social Context*, *Internal Context*, *Perceived Resources*, *Task Context*, *Technical Context* and *Temporal Context*. The UXCS will contribute to ease and promote the integration of the context as a core dimension of UX evaluation. In academia, the UXCS might be used for studying the interrelationships between contextual dimensions and UX. Research could be transferred into practice through recommendations for UX practitioners.

Third, we designed the UX Cards in order to translate the UX needs identified within the literature into tangible supports that could be used for the design and evaluation of UX. Based on the literature on psychological needs, the psychological needs-driven UX approach appears to be a powerful framework for the design of more experiential interactive systems. We thus wanted to support this approach with a pragmatic lightweight card-based method, which can be used for qualitative studies in academia, and might support product development processes in industry as well.

Finally, we also adapted and translated a psychological needs scale developed by Sheldon et al. (2001). The initial questionnaire has been developed in the framework of positive psychology research. Even though this tool has already been used with minor adaptations in several UX studies (Hassenzahl, Diefenbach, & Göritz, 2010; Partala & Kallinen, 2012; Tuch, Trusell, & Hornbæk, 2013), we observed during pilot tests that the wording of several items did not perfectly fit the context of UX. To preserve the scale's face validity and avoid excluding several items from the scale based on a factor analysis (as done in Hassenzahl et al., 2010), we adapted or replaced problematic items by carefully taking the literature on psychological needs into account. We kept 4-5 items in each need subscale, for a total of 30 items. With bilingual experts, we also translated the scale into French, to be used on French-speaking samples. We conducted pilot tests on 20 participants before using the scale during our laboratory study (Paper E). While this study provides first insight into the scientific quality of our adapted needs scale, our sample size (N = 70) did not allow for sound statistical investigations of the scale's structure. The validation of the scale on a large sample is thus a perspective for future work and is not further detailed in the present manuscript.

1.2.2. EMPIRICAL FINDINGS ON UX EVALUATION

Besides the aforementioned direct methodological contributions, this dissertation also presents empirical findings on UX evaluation. In Paper E, we experimentally investigated how UX alters two well-established HCI evaluation methods, namely expert evaluation and user testing in a laboratory setting. These two techniques were initially developed as usability evaluation methods; yet several studies show that they are now used by extension for the evaluation of UX (Alves et al., 2014).
Our study shows that these methods both present significant limitations for the evaluation of UX. Established evaluation methods therefore need to be adapted in order to fit the nature and characteristics of UX or be replaced by alternative methodologies. We also discussed implications of our findings for UX research and practice and reviewed alternative evaluation methods such as in the wild studies and longitudinal evaluations.

1.3. RELEVANCE TO PRACTICE

Through the present work, we attempted to conciliate scientific requirements and relevance to practice. The methodological contributions identified in the preceding section have all in common a scientific grounding and a strong concern for their relevance to practice. Each of the tools we developed (cf. Section 1.2) is intended to be pragmatic enough to be used by practitioners and support industrial UX design and evaluation processes.

The French version of the AttrakDiff 2 scale, which we chose to publish in French to make it more accessible to its target audience (Paper B), is a lightweight and cost-effective way of assessing UX. The German and English versions of the scale, freely accessible online (http://attrakdiff.de/), already support practitioners around the world in their UX evaluation. We organized workshops on UX evaluation in France and Luxembourg, and could witness a great enthusiasm for the questionnaire. By providing practitioners with easy to use evaluation tools, we hope to limit the use of homegrown methods and to enhance the scientific quality of UX evaluations.
Similarly, the UX Context Scale supports a self-reported assessment of contextual properties by final users. It is a flexible method that can be used both during face-to-face or remote studies, and for the evaluation of functional prototypes or products available on the market. The UXCS is especially relevant for highly context-dependent products, for example mobile devices (Korhonen et al., 2010). Moreover, for a faster administration and in the case someone is not interested in the objective contextual elements, the scale can be shortened by excluding those items.
In the case of the UX Cards, the relevance to practice comes both from the card format, popular amongst practitioners, and the fact of synthesizing a major UX theory to encourage its use in design practice and training. For UX design, the card-set is accompanied by instructions cards describing brainstorming techniques aimed at supporting the triggering of ideas and outside-the-box thinking processes. As card tools are primarily pragmatic and generally involve no training to be used, they are easy to integrate into the design process. Finally, they are also powerful collaborative tools that

may help initiating discussions on a topic and foster collaboration between project stakeholders.

In addition to the potential transfer to practice of the three aforementioned tools, we also tried to increase the relevance to practice by specifically investigating the practitioners' perspective. In our first study (Paper A), we surveyed 758 professionals to understand their view on UX. While the issue of a shared UX definition is often discussed from an academic viewpoint, we also explored the need for such a definition in industry. A better understanding gained of practitioners' perspectives is a necessary step toward continual improvement of UX activities. In our fifth study (Paper E), we investigated how UX alters UX expert evaluation and user testing in-situ. These methods are the most widely used UX evaluation methods in industry. By exploring what evaluation practices remain unchanged yet effective and valid, or on the contrary what practices do not meet the requirements of UX evaluation, we derived useful recommendations for practitioners. While our sixth study (paper F) offers more of a perspective rather than a comprehensive research work, it seems worth mentioning. This position paper intends to stimulate a reflection on potential links between research and practice. The idea of using materiality to "speak the language of designers" invites the research community to think about other ways of transferring science to practice.

1.4. CONTRIBUTION TO CONSOLIDATION

We believe that our six studies and their respective outputs concretely contributed to UX research consolidation. In addition, this dissertation also informs about the benefits, challenges and limitations of UX consolidation strategies as derived from our respective studies. Our studies mainly provide insights into the topics of replication (Paper A), translation and cross-cultural validation of a quantitative instrument (Paper B), design of a summated rating scale (Paper C), and analysis of the transferability of established HCI methods to the evaluation of UX (Paper E). These aspects are discussed within each paper.
The combination of multiple methodological studies within a single research project enabled us to reach consolidation goals more efficiently than other research projects could do. A frequent limitation to consolidation in research indeed stems from the fact that most research projects are financed based on novelty and originality of expected outcomes. It is thus difficult to dedicate a multi-annual research project to the objective of consolidating a specific research field. Isolated initiatives tend to have less impact, as they require to be coordinated. This is one of the original contributions of the present research project.

2. LIMITATIONS AND PERSPECTIVES FOR FUTURE WORK

This thesis explored several research consolidation strategies and proposed a number of methods to support UX design and evaluation. Several limitations and challenges may however be noted that might be addressed in future research.

2.1. CONSOLIDATION OF THE PRESENT WORK

First, as we advocate for research consolidation and good scientific practice, it is important to stress that the methods presented in this dissertation also require further validation steps.

With regards to the French AttrakDiff 2 questionnaire, three problematic items were identified and additional investigations are required to know whether this can be attributed to the quality of the translation or to a weakness in the original instrument. In both cases, decisions will have to be taken regarding the possible deletion of these items. Complementary studies also need to be conducted in order to assess the convergent and divergent validity of the translated questionnaire. Ideal candidate instruments to check for convergent validity of the AttrakDiff scale would be the User Experience Questionnaire (UEQ; Laugwitz, Held, & Schrepp, 2008) or the me-CUE (Minge & Riedel, 2013). Unfortunately no valid French translations of these instruments are currently available, and the ways of checking for convergent and divergent validity are thus additional challenges to tackle.

Similarly, the UX Context Scale should be tested and validated on additional samples to confirm its factorial structure and the quality of its items. Instead of simply disseminating the scale online, it would also be interesting to experimentally induce deteriorated contextual elements, in order to see if the underlying contextual factors behave according to our assumptions. For example, we could compare the UX of a system across two samples; one being placed in a deteriorated physical or social context while the other group would interact in a more positive setting; in compliance with strict ethical requirements of course. Thanks to the UXCS, we also intend to better explore what Korhonen, Arrasvuori and Väänänen-Vainio-Mattila (2010) call the "triggering context", i.e. the most important factor that affects UX in a specific situation.

Finally, further investigations are also required on the UX Cards. In upcoming studies, we intend to experimentally assess the use of the UX cards for UX design by analyzing how the cards impact the emergence of design ideas, from both a quantitative and a qualitative viewpoint. Similarly, we intend to explore their potential benefits for UX training: how well are the cards understood and used by students? Is the card-set more effective in providing learning outcomes than a more traditional course about psychological needs? These additional methodological studies would allow us to better assess the relevance of the UX cards to practice.

2.2. ASSESSING RELEVANCE TO PRACTICE

Within Section 1.3, we claimed our methods and work to be relevant to practice. While the concern of addressing practitioners' needs has been an additional strong motivation to our research activities, only the UX Cards have so far been exposed to practitioners for research purposes (Paper D). This was an excellent opportunity for us to understand how methods might fit into industrial settings and a great source of inspiration for further

refinements of the method. However, relevance to practice cannot be simply assumed or taken for granted, even if the developed tools seem lightweight and cost-effective. More systematic investigations are needed to see how our tools are used in practice and what benefits they bring to product development processes. Field studies on real use cases should be conducted to know under which conditions a method is effective. We are currently releasing the tools developed in this thesis to the community in order to gather feedback on their use in practice. In a second step, we envision systematic investigations of their effectiveness in practice.

2.3. DOCUMENTATION OF OUR STUDIES

In the introductory chapter, we stressed a number of times the issues caused by poor documentation of research findings: impossibility for the community to check for the validity of methods used or the appropriateness of statistical inferences made, difficulty to compare results across studies or to replicate previous findings.

Just as other members of the research community, we sometimes felt constrained by the limitation in the number of pages or words imposed by publication formats. This constraint even prevented us in some cases to submit our work to a particular journal whose aims and scope perfectly matched our findings; the limitation on the number of words was too strict for us to be able to comprehensively describe the method used and discuss our results. While we tried as much as possible to thoroughly document our studies and to provide all necessary details and material, our work is not exempt from flaws. Four out of the six papers included in this manuscript are still in the process of peer-review and the exchanges with reviewers greatly contribute to enhance the quality of the final publications and to minimize potential flaws, omissions or vagueness. In this manuscript, we provide all experimental material as Appendices. As we would like to encourage the community to do so, we also intend to publish raw data as supplementary material to the papers or to upload the files on online open access platforms such as ResearchGate (http://www.researchgate.net) or Academia (http://www.academia.edu).

3. ON THE NECESSITY OF CONSOLIDATION IN UX RESEARCH

Through our six studies and their respective outputs, we concretely contributed to UX research consolidation. There is obviously much more to achieve and a dissertation can only be one part in this overall objective. Each step of the present project indeed revealed complementary studies that would deserve (if not demand) to be conducted. The more we progressed with our project, the more methodological consolidation appeared as necessary to inform UX studies in general. We therefore argue for the necessity of consolidation in UX research and discussed ways through which the research community might address this ambitious objective. We provide recommendations for future research to better support consolidation.

3.1. TOWARDS MORE SCIENTIFIC RIGOR IN UX STUDIES

While recent years have witnessed the flourishing development of novel UX methods and tools (Vermeeren et al., 2010), these initiatives have been unequally successful (Fallman & Stolterman, 2010). As shown in our introductory chapter, UX research on evaluation and design methods suffers from several major limitations. We believe that one way for the research community to overcome these challenges is to foster and strengthen research consolidation practices and to establish higher standards related to scientific rigor.

Our point here is not to argue that every study should focus on the consolidation of previous research instead of proposing new findings or methods. We acknowledge the importance of innovation in UX research: in the future, the development of novel UX design and evaluation methods should focus on the categories of methods that are currently lacking research: early-stage methods to study the imagined experience before usage, methods to assess the UX of user groups, and long-term methodologies able to grasp the highly dynamic nature of experience (Vermeeren et al., 2010). Innovative methods relying on sensorial approaches might also be explored as alternatives to the almost exclusive verbal or visual instruments currently under use (Isbister, Höök, Laaksolahti, & Sharp, 2007, Tscheligi et al., 2014).
However, innovation and consolidation are no opposites. Every study should be based on good scientific practice, be well documented, and allow for replication. While this statement seems to be common sense, counterexamples of good scientific practice are easy to find in the literature and are also illustrated in the present work; this confirms how important it is to bear this major concern in mind. As stated by Dix (2010), "mere acceptance of knowledge by a group is not sufficient; we need some assurance of the truth and validity of our knowledge". Novel scientific tools and methods require careful in-depth examination and several validation studies. The psychometric properties of quantitative tools should be assessed with appropriate statistical analyses and researchers using these tools in their projects should be trained to apply relevant analyses, as well as to publish the results of these analyses (Cairns, 2007). Qualitative studies should be no exception to good scientific practice. Guidance can be found in publications in social sciences or research methods handbooks (Bordens & Abbott, 2011). Whenever possible and when no data confidentiality or ethical issues arise, researchers should make their datasets available to the community, which will encourage building on previous findings and replicating previous work (Wilson et al., 2013).

Through the present research, we also stressed the importance of applying several forms of consolidation to UX research. As emphasized in the present work, a first key step in UX research with regards to design and evaluation methods is to know more about the scientific quality of existing UX methods. Validity of methods should be checked for instead of being taken for granted and it is the responsibility of both authors and editors to set high publication standards on this issue. More rigorous translations of tools should be conducted and the resulting cross-cultural validation studies should be published so as to benefit to the whole community.

While a majority of studies on UX currently focuses on leisure and consumer products, investigations on other application domains should be an objective (Bargas-Avila & Hornbæk, 2011), and should allow challenging the assumptions underlying UX models and theories developed so far. Similarly, cross-cultural studies should also inform UX models by assessing to what extent they are generalizable across different cultures.

The dynamics of the research community will be an essential factor in the success of consolidation initiatives.

3.2. ENCOURAGING COMMUNITY DYNAMICS

As already emphasized in the introductory chapter, we believe that the research community has the power to create and maintain a dynamic around the topic of research consolidation. The promotion of UX research consolidation under all forms should be pushed further by raising awareness on the importance of replication studies, translation and cross-cultural validation of UX methods, or thematic meta-analyses.

Workshops and researchers thematic seminars are great ways to push UX research further and to setup quality standards for UX methods. Similarly, the regular publication of research agendas (i.e., Hassenzahl & Tractinsky, 2006; Kuutti & Bannon, 2014; Law & Van Schaik, 2010) should be encouraged, as they constitute reference points and stimulate further research in the field by articulating issues that researchers should address.
Provocative reseach papers, which tend to raise methodological debates, are also a way to influence and foster follow-up studies, as exemplified by the impact of Kjeldskov et al.'s (2004) paper on field evaluations "Is it worth the hassle?". Ten years and after many debates in the community, the authors conclude, "we believe that the last ten years of empirical work and research discussions of lab and field evaluations have been highly valuable for the mobile HCI research field, and therefore also that engaging with this topic of research has been *worth the hassle*." (p. 50). With regards to UX evaluation, Law (2011) stated, "the tension between the two camps (i.e. qualitative design-based and quantitative model-based) stimulates scientific discussions to bring the field forward." Forums for "controversial, boundary pushing" presentations such as alt.chi also provides opportunities for raising debate and stimulating research. Baumer et al.'s (2014) speculative research visions for CHI'2039 or Hassenzahl's paper (2013) on "Experiences before things" illustrate this alternative thought-provoking approach.

The role of standardization in the consolidation of research might also be discussed, and some authors call for a wider use of international standards as starting points for design education, training, and practice (Bevan, 2009). HCI standards aim at supporting good practice in design and might also support legal requirements (e.g., specifying in a contract that the design and development process is required to comply with ISO 9241-210, 2010) (Bevan, 2001; Stewart, 2010). Standards also have limitations (Bevan, 2001): (1) they often describe principles, not useful solutions (2) they are developed over several years and can quickly become out of date (3) they

might stifle innovation by imposing unnecessary constraints on design. The current ISO 9241-210 (2010) standard is described as "a powerful tool to assure a human centered design process" (Bevan, 2009) or a "manifesto for the field of user experience" (Travis, 2011). As ISO 9241-210 is under periodical review and related new standards are currently under development (e.g., ISO 9241-220), it is good to remind that everyone might get involved as an expert member in the development of standards. In that sense, standardization provides a further mean of integrating UX research and practice.

3.3. MAKING THE MOST OUT OF INTERDISCIPLINARITY

When raising debates on methodological topics, one should nevertheless be careful not to polarize the field into separate camps of thoughts. While provocative papers put fundamental discussions in the research agenda, UX gains its richness through interdisciplinarity. We therefore believe that UX research should continue to strive for a harmonious integration of the variegated perspectives and disciplines the field is drawing on. Each of these disciplines might contribute to the advancements of the field and the UX community should try to maximize the contributions of each related discipline by relying on their respective strengths (Dix, 2010).

One of the challenges highlighted in the introduction of this thesis is the fact that UX research is not always based on a sound theoretical knowledge or model. Experience design and evaluation indeed require a deep understanding of people (Hassenzahl, 2010), their cognitive and affective processes and basic needs. This is a challenge that the disciplines UX is drawing on might contribute to overcome. A profound theoretical and empirical knowledge on the understanding of human experiences has been accumulated through decades of research in fields such as psychology, sociology, ethnology, or philosophy. Universal human needs have for example been identified and thoroughly investigated (Sheldon, Elliot, Kim, & Kasser, 2001), as well as psycho-cognitive and psycho-social processes, human values (Rokeach, 2008), human motivation (Ryan, 2012) human emotions (Fridja, 1986; Scherer, 2005) or even optimal experiences at a more generic level (Csikszentmihalyi, & LeFevre, 1989).

In recent approaches that are more and more interdisciplinary, we also see new methodologies and theoretical frameworks, which are not domain domain-specific anymore (Lazar, Feng, & Hochheiser, 2010). It appears that both researchers and practitioners have started profiting from cross-fertilization between the domains UX is drawing on. It remains unclear however how each domain actually contributes to an integrated design process in order to support UX design and to what extent there remain domain-specific approaches (Wright & McCarthy, 2010). The harmonious and effective integration of varied perspectives is therefore another ambitious objective that UX as a community has to address.

As the present dissertation falls within the scope of psychology and HCI, we were able to show how psychology might support UX research. Psychology does not only provide theories and models of human experiences (e.g., the psychological needs-driven approach used in Papers

D-F); it also provides support to scientific practices through well-established standards (e.g., APA publication manual) and methodological guidance (e.g., cross-cultural research guidelines, summated rated scale construction methodology). As stated by Carroll almost twenty years ago:

> « There is unprecedented potential for interdisciplinary synergy here. Social science has always borne the vision of what human society might become, but it has typically lacked the means to be constructive. Computer science—quite the converse—cannot avoid causing substantial social restructuring. An integrated and effective HCI can be a turning point in both disciplines and, perhaps, in human history. » (Carroll, 1997)

Research on UX and the overall aim of designing for positive UX might well represent this turning point.

3.4. BEING RELEVANT TO PRACTICE BY BEING RELEVANT TO RESEARCH FIRST

Relevance to practice is often expressed as a major concern in studies on UX methods. As UX is an applied field, many researchers indeed feel that the work they conduct and the tools they develop should straightforwardly support design practice. To address the needs of UX practitioners, academics attempt to develop methods that are flexible, applicable, cost-effective, and lightweight (Karapanos et al., 2012; Roto, Ketola, & Huotari, 2008). However by doing so, they sometimes have to reach a challenging compromise between the scientific quality of a method and its practicability (Vermeeren et al., 2010). We already illustrated this issue through the example of the Sentence Completion method (Kujala et al, 2013). The authors explained that their goal was to develop a practical technique rather than a strict measurement tool and they therefore wonder whether scientific quality criteria such as validity or reliability were applicable to their work.

While the close links between research and practice constitute a main strength of our field, Dix (2010) emphasizes the dangers of confusing research and practice. The methods and tools used by both researchers and practitioners are often similar (e.g., user testing), yet the goals are different. In academia methods should serve the primary goal of providing support to understand phenomena, whereas in industry methods are only means to achieve the primary goal of designing successful systems and products (Roto et al, 2013). In that sense, practical usefulness should not be the main focus on HCI and UX research and especially not be an excuse for a lack of scientific quality, otherwise "non-rigorous research may be impossible to separate from consultancy work, journalism, or simply matters of opinion" (Fallman & Stolterman, 2010).

> "If we focus only on practical usefulness and exclude explanation and interpretation, we do serious harm to our very nature as researchers. The purpose of research – at least the research done in universities – is to develop better understanding of the world around us." (Kuutti, 2010)

Instead of developing techniques and measurement tools to directly support UX practice, several researchers (Dix, 2010; Tractinsky, in Roto et al., 2010) claim that UX research should develop measurement instruments to test

and improve our UX theories. It is through the development and consolidation of a sound body of knowledge on UX that UX research will then eventually support UX practice in designing interactive systems. This stance is well summarized by Dix (2010): "An excessive utility focus tends to mean that research runs behind technology. Work on the newest thing is too late for it, and looking for the next big thing is almost bound to fail; the big win is in using the new, the old and ideas for the next as means of uncovering more general knowledge." (p. 8).

Of course, this does not prevent academia from maintaining strong links with practitioners, and from trying to understand the nature of design (Stolterman, 2008) in order to best address the needs of UX practice (Roedl & Stolterman, 2013). To be relevant to practice, research might address problems and topics important to professionals and make academic findings as well as scientifically grounded methods and tools accessible to practitioners (Fallman & Stolterman, 2010).

CONCLUSION

During the past years, the HCI research community has spent considerable research efforts developing a wealth of methods, techniques and tools to address the new challenges brought by the topic of UX. Great dynamics have grown and have led to stimulating outcomes and future perspectives. We have however been able to highlight through the present work that UX research on evaluation and design methods still suffers from several limitations, despite the aforementioned efforts; we have shown that the remaining limitations relate above all to the scientific quality of methods and the scarcity of research consolidation initiatives. This might also explain why the standing of UX has not entirely overcome yet the status of a buzzword or a trend in some parts of the HCI community.

We argue that one way for the research community to overcome these challenges is to promote and foster research consolidation practices. The promotion of UX research consolidation should systematically raise awareness on the importance of replication studies, translation and cross-cultural validation of UX methods, methodological research and meta-analyses. As highlighted by Dix (2010), the research community has to *think methodologically*, which means to encourage an on-going methodological critique of the methods we use and how we use them.

As reflected by the title of this thesis, the consolidation of UX research is not to be thought of as a state that could easily be reached. It is rather a process, a continuous effort "towards" consolidated research, a virtuous circle aimed at strengthening HCI and UX as scientific fields of study. Through the present work, we hope to play a dynamic role in raising awareness on the need for research consolidation and hope our research contributes to taking UX consolidation to the next step.

REFERENCES FOR PART III

Alves, R., Valente, P., & Nunes, N. J. (2014). The state of user experience evaluation practice. *In Proceedings of the 8th Nordic Conference on Human-Computer Interaction NordiCHI'14*. New York, NY: ACM Press (pp. 93-102). doi:10.1145/2639189.2641208

American Psychological Association. (2010). *Publication manual of the American Psychological Association. - 6th ed.* Washington, DC: American Psychological Association.

Bargas-Avila, J.A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. In Proceedings of the ACM conference on human factors in computing systems (CHI 2011) (pp. 2689–2698). New York, NY: ACM Press. doi: 10.1145/1978942.1979336.

Baumer, E., June, A., Bie, M., Bonsignore, E., Börütecene, A., …, & Yip, J. (2014). CHI 2039: Speculative Research Visions. Proceedings of the International Conference on Human Factors in Computing Systems Alt.Chi, pp. 761-770. http://hdl.handle.net/1942/16378

Bevan, N. (2001). International standards for HCI and usability. *International Journal of Human-Computer Studies, 55*(4), 533-552. doi: 10.1006/ijhc.2001.0483

Bevan, N. (2009). International Standards for Usability Should Be More Widely Used. Journal of Usability Studies, 4(3), 106-113.

Bordens, K. S., & Abbott, B. B. (2011). Research Design and Methods. A process approach (8th edition). New-York, NY, USA: McGraw-Hill.

Cairns, P. (2007). HCI... Not As It Should Be: Inferential Statistics in HCI Research. *Proceedings of the 21st British HCI Group Annual Conference on People and Computers BCS-HCI '07*. Swinton, UK: British Computer Society.

Carroll, J. (1997). Human-computer interaction: Psychology as a science of design. *Annual Review of Psychology, 48*, 61-83.

Csikszentmihalyi, M., LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*. 56, 5, 815–822.

Dix, A. (2010). Human-Computer Interaction: a stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers, 22*(1), 13-27.

Fallman, D., & Stolterman, E. (2010). Establishing Criteria of Rigor and Relevance in Interaction Design Research. In *Proceedings of the 2010 international conference on The Interaction Design Create'10*. Swinton, UK: British Computer Society.

Frijda, N.H. (1986). *The Emotions*. Maison des Sciences de l'Homme and Cambridge University Press.

Hassenzahl, M. (2003). The thing and I: understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A.F. Monk, & P.C. Wright (Eds.), *Funology: From Usability to Enjoyment* (pp. 31-42). Kluwer: Dordrecht.

Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction, 19*, 319-349. doi: 10.1207/s15327051hci1904_2

Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*, *3*(1), 1-95. doi: 10.2200/S00261ED1V01Y201003HCI008

Hassenzahl, M. (2013). Experiences Before Things: a Primer for the (Yet) Unconvinced. *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2059-2068). New York, NY: ACM. doi:10.1145/2468356.2468724

Hassenzahl, M., Burmester, M., & Koller, F. (2003). *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*. In J. Ziegler, & G. Szwillus (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung* (pp. 187–196). Stuttgart, Germany: B.G. Teubner. doi: 10.1007/978-3-322-80058-9_19

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers*, 22(5), 353-362.

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behavior & Information Technology*, *25*, 91-97. doi: 10.1080/01449290500330331

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

International Organization for Standardization. ISO/CD 9241-220, *Ergonomics of human-computer interaction -- Part 220: Processes for enabling, executing and assessing human-centred design within organizations*. Geneva, Switzerland: International Organization for Standardization.

Isbister, K., Höök, K., Laaksolahti, J., and Sharp, M. (2007). The sensual evaluation instrument: Developing a trans-cultural self-report measure of affect. *International Journal of Human-Computer Studies, 65*(4), 315–328.

Karapanos, E., Martens, J.-B., Hassenzahl, M. (2012). Reconstructing Experiences with iScale. *International Journal of Human-Computer Studies*, *70*(11), 849-865. doi:10.1016/j.ijhcs.2012.06.004

Kjeldskov, J., Skov, M. B., Als, B. S., and Høegh, R. T. (2004). Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In *Proceedings MobileHCI 2004*, Berlin, Germany: Springer-Verlag, 529-535. doi: 10.1007/978-3-540-28637-0_6

Korhonen, H., Arrasvuori, J., & Väänänen-Vainio-Mattila, K. (2010). Analysing user experience of personal mobile products through contextual factors. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia* - MUM '10, 1–10.

Kujala, S., Walsh, T., Nurkka, P., & Crisan, M. (2013). Sentence Completion for Understanding Users and Evaluating User Experience. *Interacting with Computers*, *26*(3), 238-255. doi:10.1093/iwc/iwt036

Kuutti, K., & Bannon, L. J. (2014). The turn to practice in HCI: towards a research agenda. In Proceedings of the ACM conference on human factors in computing systems CHI'14 (pp. 3543–3552). New York, NY: ACM Press.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.): HCI and Usability for Education and Work USAB 2008, LNCS 5298, pp. 63-76. Berlin/Heidelberg: Springer.

Law, E. C. (2011). The measurability and predictability of user experience. *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems EICS'11*. New York, NY: ACM Press.

Law, E., & Van Schaik, P. (2010). Modelling User Experience – An agenda for Research and Practice. *Interacting with computers, 22 (5),* 313-323. doi: 10.1016/j.intcom.2010.04.006

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in Human-Computer Interaction*, London, UK: John Wiley & Sons.

Minge, M. & Riedel, L. (2013). meCUE – Ein modularer Frage- bogen zur Erfassung des Nutzungserlebens. In: S. Boll, S. Maaß & R. Malaka (Hrsg.): Mensch und Computer 2013: Interaktive Vielfalt (S. 89-98). München, Oldenbourg Verlag.

Partala, T., & Kallinen, A. (2012). Understanding the most satisfying and unsatisfying user experiences: Emotions, psychological needs, and context. *Interacting with Computers, 24*(1), 25-34. doi:10.1016/j.intcom.2011.10.001

Roedl, D. J., & Stolterman, E. (2013). Design research at CHI and its applicability to design practice. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'13*, New York, NY: ACM Press. doi: 10.1145/2470654.2466257

Rokeach, M. (2008). *Understanding human values: Individual and societa*l. New York: Free Press

Roto, V., Ketola, P., & Huotari, S. (2008). User Experience Evaluation in Nokia. Now Let's Do It in Practice. *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. New York, NY: ACM Press. doi: 10.1145/1358628.1358967

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (2010). *Demarcating User Experience Dagstuhl Seminar*, 10373 Abstracts Collection. Retrieved from http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=10373

Ryan, R. M. (Ed.) (2012). The Oxford Handbook on Human Motivation. Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780195399820.001.0001

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729.

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.

Stewart, T. (2010). Exciting Times for Human Centred Design [Blog post] Retrieved from http://www.system-concepts.com/articles/usability-articles/2010/exciting-times-for-human-centred-design-standards

Stolterman, E. (2008). The nature of design practice and implications for interaction design research. *International Journal of Design, 2*(1), 55-65.

Travis, D. (2011, June 6). ISO 13407 is dead. Long live ISO 9241-210! [Blog post] Retrieved from www.userfocus.co.uk/articles/iso-13407-is-dead

Tscheligi, M., Isbister, K., Höök, K., Obrist, M., Busch, M., & Hochleitner, C. (2014). "Touch me" – Workshop on Tactile User Experience Evaluation Methods. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, New York, NY: ACM Press. doi: 10.1145/2559206.2560471

Tuch, A., Trusell, R., & Hornbæk, K. (2013). Analyzing users' narratives to understand experience with interactive products. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2079-2088). Paris, France: ACM. doi:10.1145/2470654.2481285

Vermeeren, A, Law, E., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User Experience Evaluation Methods: Current State and Development Needs. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, New York, NY: ACM Press. doi:10.1145/1868914.1868973

Wilson, M.L.L., Resnick, P., Coyle, D., & Chi, E.H. (2013). RepliCHI: The Workshop. *CHI '13 Extended Abstracts on Human Factors in Computing Systems (*pp. 3159–3162), New York, NY, USA: ACM Press.

Wright, P.C., & McCarthy, J.C. (2010). Experience-Centered Design: Designers, Users, and Communities in Dialogue. Synthesis Lectures on Human-Centered Informatics. Morgan & Claypool Publishers.

APPENDICES

APPENDIX 1 – UX SURVEY (PAPER A)

General instructions

Have you ever heard of the concept of User Experience (also called UX)? Does its definition seem clear to you? At present, despite its popularity among professionals, there is little scientific evidence to help understand what the User experience actually is. We are trying to address this problematic through the present study. Regardless of your profile and background (professional, researcher, student...) get involved and help bringing this concept to maturity by sharing your views on User experience.
Thank you for responding as seriously as possible to the questionnaire, which will take you approximately 20 minutes (with the possibility to save your answers and come back to the survey later if needed). Your answers will be recorded and processed anonymously. Thank you for your interest and participation!

This survey is adapted from the following study: Law, E., Roto, V., Hassenzahl, M., Vermeeren, A. & Kort, J. (2009) Understanding, scoping and defining UX: a survey approach. In *Proceedings of CHI 2009*, Boston, USA.

Your Background

These first questions will focus on your professional background and the place of User experience (UX) in your career.

1. I work in:	O Industry O Academia O Both or between
2. My primary role is:	O Researcher O Consultant O Manager O Practitioner O Student O Other _____
3. I was originally educated in the field of:	O Arts O Design O Marketing O Business O Quality / Processes O Psychology / Social Sciences O Technology / Software O Human-Computer Interaction O Other _____
4. What is your current job position?	_____
5. Which of the following statements applies the best to your primary interest in UX? *I'm interested in understanding the nature of UX:*	O per se O to design better products O to better sell products O to make people happier O Other _____
6. In the moment, how central is your UX to your professional work?	O very central O central O less central O UX is not central at all to my professional work
7. In the case UX is not central at all to your professional work, do you collaborate with people working in that field ?	O yes O no
8. Indicate on a scale from 1 to 10 your level of familiarity with the concept of UX. (where 1 is the lowest familiarity level and 10 the highest)	*Scale from 1 to 10* + O I've never heard about UX
9. My age is (in years):	_____
10. For how many years have you been working in the field of UX?	_____
11. For how many years have you been working in the field of User-Centred Design (UCD)?	_____
12. My country of residence is:	*List of countries*
13. My gender is:	O female O male

UX Statements

Here is a list of statements about User experience (UX).
Could you state how much you agree or disagree with each of these statements?
Indicate "I don't understand" if you don't understand one of the statements.

Please note that this study focuses on understanding the nature of UX and delineate its scope. It is therefore possible that the statements seem too complicated or vague. This is normal and meets the purpose of the present study. After the statements you will be able to comment them freely and give your precise point of view on UX.

Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*	*I don't understand*

- Fleeting and more stable aspects of a person's internal state (e.g., needs, motivations) affect a person's experience of something
- UX occurs in. and is dependent on the context in which the artefact is experienced
- Designing (for) UX must be grounded in User-Centred Design
- Prior exposure to an artefact shapes subsequent UX
- Usability is a necessary precondition for good UX
- Measuring UX implies determination of merits, values, and significance of an artefact in relation to a person's goals and needs
- UX should be assessed while interacting with an artefact
- We cannot design UX, but we can design for UX
- UX can change even after a person has stopped interacting with the artefact
- UX is highly dynamic - it changes constantly while interacting with a product
- UX refers to affective states, i.e., any combination of valence (good - bad. pleasant – unpleasant) and physiological arousal (calm – excited)
- Imagined use of a product can result in real experiences
- There is a definite need for a standardized definition of the term UX
- UX must be approached qualitatively
- UX can be quantified and thus compared across similar (or competitive) artefacts
- UX is based on how a person perceives the characteristics of an artefact but not on the characteristics per se
- UX is not about people's performance (ability to understand and use) in their relation with an artefact, but about the person's perception of that performance
- UX should be assessed after interacting with an artefact
- UX is not new, it is already covered by existing engineering approaches
- Only an individual person can have an experience. An experience is something personal. Something 'within' a person.
- People will never have comparable UX - each and every interaction with a product results in a unique experience
- UX is equal to emotional attachment
- UX is best viewed in terms of marketing

Do you have any additional comments on these statements? Please elaborate

_____

UX Definitions

Definition 1

All aspects of the end-user's interaction with the company. Its services and its products. The first requirement for an exemplary user experience is to meet the exact needs of the customer without fuss or bother. Next comes simplicity and elegance that produce products that are a joy to own, a joy to use. True user experience goes far beyond giving customers what they say they want, or providing checklist features. [Nielsen & Norman Group, nngroup.com]

What do you think of this definition? (give your answer in the box below)

Definition 2

A consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.) the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.) [Hassenzahl & Tractinsky, 2006]

What do you think of this definition? (give your answer in the box below)

Definition 3

The entire set of affects that is elicited by the interaction between a user and a product including the degree to which all our senses are gratified (aesthetic experience) the meanings we attach to the product (experience of meaning) and the feelings and emotions that are elicited (emotional experience). [Desmet & Hekkert, 2007]

What do you think of this definition? (give your answer in the box below)

Definition 4

The value derived from the interaction(s) [or anticipated interaction(s)] with a product or service and the supporting cast in the context of use (e.g. time, location, and user disposition). [Sward & MacArthur, 2007]

What do you think of this definition? (give your answer in the box below)

Definition 5

The quality of experience a person has when interacting with a specific design. This can range from a specific artefact such as a cup toy or website up to larger integrated experiences such as a museum or an airport. [UXnet.org]

What do you think of this definition? (give your answer in the box below)

If you had to pick one of these UX definitions, which one would it be?

- O Definition 1
- O Definition 2
- O Definition 3
- O Definition 4
- O Definition 5

Could you comment the reasoning for your choice?

APPENDIX 2 – FRENCH ATTRAKDIFF SCALE VALIDATION STUDY (PAPER B)

Evaluation de l'expérience utilisateur Facebook

Dans le cadre d'un projet de recherche sur l'expérience utilisateur des systèmes interactifs, nous souhaiterions évaluer votre expérience du site web Facebook.
Ce questionnaire se présente sous forme de paires de mots pour vous assister dans l'évaluation du système. Chaque paire représente des contrastes. Les échelons entre les deux extrémités vous permettent de décrire l'intensité de la qualité choisie.
Ne pensez pas trop aux paires de mots et essayez simplement de donner une réponse spontanée. Vous pourrez avoir l'impression que certains termes ne décrivent pas correctement le système. Dans ce cas, assurez-vous de donner tout de même une réponse. Gardez à l'esprit qu'il n'y a pas de bonne ou mauvaise réponse. Seule votre opinion compte ! Vos réponses seront enregistrées et traitées de manière anonyme.

Evaluation 1/3

Humain	□	□	□	□	□	□	□	Technique
M'isole	□	□	□	□	□	□	□	Me sociabilise
Plaisant	□	□	□	□	□	□	□	Déplaisant
Original	□	□	□	□	□	□	□	Conventionnel
Simple	□	□	□	□	□	□	□	Compliqué
Professionnel	□	□	□	□	□	□	□	Amateur
Laid	□	□	□	□	□	□	□	Beau
Pratique	□	□	□	□	□	□	□	Pas pratique
Agréable	□	□	□	□	□	□	□	Désagréable
Fastidieux	□	□	□	□	□	□	□	Efficace

Evaluation 2/3

De bon goût	□	□	□	□	□	□	□	De mauvais goût
Prévisible	□	□	□	□	□	□	□	Imprévisible
Bas de gamme	□	□	□	□	□	□	□	Haut de gamme
M'exclut	□	□	□	□	□	□	□	M'intègre
Me rapproche des autres	□	□	□	□	□	□	□	Me sépare des autres
Non présentable	□	□	□	□	□	□	□	Présentable
Rebutant	□	□	□	□	□	□	□	Attirant
Sans imagination	□	□	□	□	□	□	□	Créatif
Bon	□	□	□	□	□	□	□	Mauvais

Evaluation 3/3

Confus	□	□	□	□	□	□	□	Clair
Repoussant	□	□	□	□	□	□	□	Attrayant
Audacieux	□	□	□	□	□	□	□	Prudent
Novateur	□	□	□	□	□	□	□	Conservateur
Ennuyeux	□	□	□	□	□	□	□	Captivant
Peu exigeant	□	□	□	□	□	□	□	Challenging
Motivant	□	□	□	□	□	□	□	Décourageant
Nouveau	□	□	□	□	□	□	□	Commun
Incontrôlable	□	□	□	□	□	□	□	Maîtrisable

Evaluation globale

Globalement, sur une échelle de 1 à 100, quel score donneriez-vous pour évaluer à la qualité de votre expérience avec ce site ? (1 désigne le moins bon score et 100 désignant le meilleur score) : _____

Votre profil

Vous êtes :	o Un homme

	○ Une femme
Quel âge avez-vous ?	_____
Quelle est votre langue maternelle ?	○ Français ○ Autre
Globalement, à quel point vous sentez-vous à l'aise avec l'utilisation des technologies ?	*Echelle en 7 points de "pas du tout à l'aise" à "tout à fait à l'aise"*

Votre utilisation de Facebook

Depuis combien de temps êtes-vous inscrit sur le site Facebook ?	○ C'est ma première visite sur le site ○ Je suis inscrit(e) depuis quelques semaines ○ Je suis inscrit(e) depuis plusieurs mois ○ Je suis inscrit(e) depuis plusieurs années
A quelle fréquence visitez-vous le site de Facebook ?	○ C'est la première fois ○ Une fois par mois ○ Plusieurs fois par mois ○ Plusieurs fois par semaine ○ Chaque jour ○ Plusieurs fois par jour
Vous utilisez le site de Facebook pour :	□ Consulter le fil d'actualité de vos contacts □ Envoyer des messages ou chatter avec vos contacts □ Publier des actus sur votre fil d'actualité □ Jouer □ Autre : _____

APPENDIX 3 – UX CONTEXT SCALE VALIDATION STUDY (PAPER C)

User Experience Questionnaire - LinkedIn

We welcome your participation in the evaluation of your experience of LinkedIn.
This survey is part of a research project focused on the User Experience of interactive systems. Thank you for responding as seriously as possible to the questionnaire, which will take you approximately 10 minutes. Your answers will be recorded and processed anonymously. Thank you for your interest and participation!

Evaluation of the system

Following, are pairs of words to assist you in your evaluation of LinkedIn. Each pair represents extreme contrasts. The possibilities between the extremes enable you to describe the intensity of the quality you choose. Do not spend time thinking about the word-pairs. Try to give a spontaneous response. Keep in mind that there is no right or wrong answer. Your personal opinion is what counts!

I found the system (LinkedIn):

Confusing	□	□	□	□	□	□	□	Clearly structured
Impractical	□	□	□	□	□	□	□	Practical
Unpredictable	□	□	□	□	□	□	□	Predictable
Complicated	□	□	□	□	□	□	□	Simple
Dull	□	□	□	□	□	□	□	Captivating
Tacky	□	□	□	□	□	□	□	Stylish
Unimaginative	□	□	□	□	□	□	□	Creative
Bad	□	□	□	□	□	□	□	Good
Ugly	□	□	□	□	□	□	□	Attractive

General context assessment

Following are pairs of words/sentences that describe the context of interaction. Each pair represents contrasted ideas about the context in which the interaction occurred. Your mission is to rate each item on a 7-points scale. This is not an evaluation of the quality of the system but an attempt to assess your environment and feelings **<u>the last time you used the system.</u>**

□ □ □ □ □ □ □

Each pair of words is presented along with a 7-points Likert scale

Physical context

While using the system:

In an unfamiliar place	In a familiar place
In an unpleasant location	In a pleasant location
In a noisy place	In a quiet place
In a moving/vibrating environment	In a steady environment
Unpleased with the temperature	Pleased with the temperature
Unpleased with the lighting	Pleased with the lighting
Moving (mobile usage)	Remaining still

Social context

While using the system:

In a public space	In a private space
Not interacting with people	Interacting with people
Feeling alone	Feeling related to other people
Feeling unsupported	Feeling supported
*Bothered by others	Unbothered by others*

Internal context

Before using the system:

*I was in a bad mood	I was in a good mood*
I had no expectations about the system	I had high expectations about the system
I had no specific tasks to achieve	I had specific tasks to achieve
I had no previous experience with the system	I already had experience in using the system
I had no information about it	I had information about it
I had a bad opinion about it	I had a good opinion about it
I felt pressed to use it	I felt free to use it

While using the system:

Demotivated	Motivated
Not interested	Interested
In a bad mood	In a good mood
Unsatisfied	Satisfied
I had insufficient skills to use it	I had sufficient skills to use it
I did not have enough time to spend on it	I had enough time to spend on it
I had the need to be helped or trained	I had no need to be helped or trained
Powerless over my environment	In control of my environment

After using the system:

I am in a bad mood	I am in a good mood
I feel unsatisfied	I feel satisfied
I am not willing to use it again	I am willing to use it again
I feel repelled by the system	I feel attracted by the system
*My expectations are not satisfied	My expectations are satisfied*

Technical context

While using the system:

Technical problems were encountered	No technical problems were encountered

Task context

While using the system:

I was doing several things simultaneously	I was focusing on the task
I was often interrupted	I was never interrupted
I focused on the product	I focused on attaining my goals
*I felt serious	I felt playful*

Temporal context

While using the system:

I used the system only once	I am using the system regularly
I spent a short time using the system	I spent a long time using the system
I was interacting at an uncomfortable pace	I was interacting at a comfortable pace
It was not the right moment to use the system	It was the right moment to use the system

About you

Your gender is:	○ female ○ male
What year were you born in?	_____
Do you feel at ease with the use of interactive systems?	*7-points Likert scale from "not at ease at all" to "completely at ease"*
My country of residence is:	*List of countries*
My mother tongue (native language) is:	○ English ○ French ○ German ○ Spanish ○ Others: _____
Your level of proficiency in English:	*7-points Likert scale from "not proficient at all" to "fully proficient"*

APPENDIX 4 – EXPERT EVALUATION (PAPER E)

INFORMED CONSENT FOR PARTICIPANT OF A RESEARCH PROJECT

Title of the project: GUIDE - Understanding User Experience to promote its integration into interactive systems

Principal Investigators: Ms Carine Lallemand (University of Luxembourg)
Dr Vincent Koenig (University of Luxembourg)
Dr Guillaume Gronier (CRP Henri Tudor)
Prof. Romain Martin (University of Luxembourg)

The purpose of this research project
This doctoral research, combining theoretical and empirical findings from psychology and Human-Computer Interaction, is focused on the concept of User Experience (UX). In order to promote its integration into interactive systems, we created a set of UX Cards based on psychological theories of human needs.

Procedures
You will be asked to perform a set of tasks using the UX Cards. These tasks consist of identifying UX elements on five interactive systems. We are not evaluating you or your performance in any way; you are helping us to evaluate our methodology. All information that you help us attain will remain anonymous. Your actions will be noted and you will be asked to describe verbally your classification process. You may be asked questions during and after the evaluation, in order to clarify our understanding of your evaluation. You will also have to fill in several questionnaires before, during and after the completion of the main task. The session will last about 2 hours. The tasks are not very tiring, but you are welcome to take rest breaks as needed.

Risks
There are no risks associated with this study other than those encountered from using a computer and a web-browser in everyday activities.

Benefits of this Project
Your participation in this project will provide information that may be used to improve the UX Cards. No promise or guarantee of benefits has been made to encourage you to participate. If you would like to receive a synopsis or summary of this research when it is completed, please notify Carine Lallemand.

Extent of Anonymity and Confidentially
The results of this study will be kept strictly confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research.

Compensation
You will receive a 50€ FNAC Voucher for the 2 hours spent on the task.

Freedom to Withdraw
You are free to withdraw from this study at any time for any reason.

Approval of Research
This research has been approved by the University of Luxembourg, the Fonds National de la Recherche (Luxembourg) and the Public Research Centre Henri Tudor (Luxembourg).

Participant's Responsibilities
I voluntarily agree to participate in this study. Due to intellectual property issues, I have the following responsibilities:

I understand that the content of the material (i.e. UX Cards, instructions, surveys) used in this study is completely confidential.

I agree not to use, divulge, publish, or otherwise make known to unauthorized persons or to the public any information obtained in the course of this research project.

Participant's Permission
I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty.

Signatures

Should you have any questions about this research or its conduct, you may contact:
Ms. Carine Lallemand

GOALS OF THIS STUDY

During my PhD, I developed a set of User Experience Cards in order to support the design and the evaluation of User Experience. The UX Cards are based on a psychological approach to need-driven experience design. The fulfillment of human psychological needs is thought to be a main trigger of positive experiences with technologies. This means that, in order to design experiential products, designers should consider interactive systems as means to fulfill needs and not only to achieve task-oriented goals. The seven main basic human needs are described here:
Effectance – **M**eaning – **P**leasure– **I**nfluence– **R**elatedness – **I**ndependance - **C**ontrol

Each card provides a definition of the need, linked terms (or synonyms) and some examples of feelings or activities, triggered by using a specific item that would point to this need. Scientific references are also included but you will not have to consider them during the exercise.

PRELIMINARY SURVEY

About you

Your gender is:	O male O female
What year were you born in?	_____
Your country of residence is:	*List of countries*
My mother tongue (native language) is:	O English O French O German O Spanish O Others: _____
Your level of proficiency in English:	*7-points Likert scale from "not proficient at all" to "fully proficient"*

Your background

The following questions will focus on your professional background.

I work in:	O Industry O Academia O Both or between
My primary role is:	O Researcher O Consultant O Manager O Practitioner O Student O Other _____
I was originally educated in the field of:	O Arts O Design O Marketing O Business O Quality / Processes O Psychology / Social Sciences O Technology / Software O Human-Computer Interaction O Other _____

What is your current job position? (optional)	_____
For how many years have you been working in HCI?	_____
Please indicate on a scale from 1 to 7 your level of familiarity with the expert (also called heuristic) evaluation of interactive systems?	*Scale from 1 to 7 from "novice" to "expert"*
How many expert/heuristic evaluations of interactive systems have you performed (approximately)?	○ Less than five ○ 5-15 ○ 15-50 ○ More than 50
Please indicate what sets of heuristics you have already used to perform heuristic evaluations:	_____
Please indicate on a scale from 1 to 7 your level of familiarity with the concept of UX (at a theoretical level)	*Scale from 1 to 7 from "not familiar at all" to "very familiar*
Please indicate on a scale from 1 to 7 your level of familiarity with psychological needs theories	*Scale from 1 to 7 from "not familiar at all" to "very familiar*

YOUR MISSION

Your mission will be to review 4 interactive systems and to freely identify within these systems some elements that could contribute to the fulfilment of the needs described in the cards. Because it might seems hard to relate a psychological need to concrete elements of the system, use the examples provided on the cards. They describe way of operationalizing the need in real life. They are not focus on technology yet, in order not to bias your evaluation. As User Experience is holistic, identified elements might not be reduced to interface or graphical elements. Everything that might influence the UX can be included in your analysis. Identified elements might relate to:

- Graphical or visual elements
- Features/Functionality
- Content
- Information Design
- Interaction Design
- Interface Design
- Service Experience
- Interoperability / Platform
- Technical experience
- Usability / Accessibility
- Context of use
- Brand / Marketing

In summary, do not hesitate to mention every single element that you believe might contribute to the fulfilment of one or several UX Needs.

PRELIMINARY STEP: READING AND UNDERSTANDING THE UX CARDS

Please read carefully all UX Cards. Do not hesitate to make comments or ask questions. For each card, please immediately rate its level of understandability.

How easy is it to understand the card (name of the card)? *

□ □ □ □ □ □ □

Each question was presented along with a 7-points Likert scale from "hard" to "easy"

Using the card, how easy is it to think about HCI elements related to the need of (name of the card)?

□ □ □ □ □ □ □

Each question was presented along with a 7-points Likert scale from "hard" to "easy"

USING THE UX CARDS

We will present you four interactive systems one by one: Facebook, Angry Bird, an Olympus compact digital camera, Amazon.com. For each system, you will have 15 minutes to freely interact with it. During this 15-min interaction time, you will have to use the cards to identify some elements able to support the fulfillment of each of the seven needs described. You might be familiar with one or several systems you will use today. This is not an issue and you only have to report your level of familiarity with the system beforehand.

Please rate your level of familiarity with [Facebook] - [Angry Birds] - [Amazon] - [Camera]

□ □ □ □ □ □ □

Each question was presented along with a 7-points Likert scale from "very unfamiliar" to "very familiar"

Please use the following table to describe precisely the elements you will identify.

Identified elements	UX need(s) fulfilled thanks to this element	UX need(s) hindered by this element
Ex: camera visual design	*Pleasure (aesthetics)* *Security (doesn't look fragile)*	
…		

After finishing the identification task, please fill in this table for each system.

Overall, what is the impact of (name of the system) on the fulfillment of the need for:			
Relatedness	Negative	○○○○○○○	Positive
Security	Negative	○○○○○○○	Positive
Pleasure	Negative	○○○○○○○	Positive
Influence	Negative	○○○○○○○	Positive
Competence	Negative	○○○○○○○	Positive
Autonomy	Negative	○○○○○○○	Positive
Meaning	Negative	○○○○○○○	Positive

POST-TASK SURVEY

After using the UX Cards: (7-points Likert scale)

I am in a bad mood	I am in a good mood
I feel unsatisfied	I feel satisfied
I am not willing to use them again	I am willing to use them again
I feel repelled by the cards	I feel attracted by the cards

Overall, how difficult or easy did you find the use of UX Cards?
7-points Likert scale from "very difficult" to "very easy"

I found the UX Cards:

Confusing	□	□	□	□	□	□	□	Clearly structured
Impractical	□	□	□	□	□	□	□	Practical
Unpredictable	□	□	□	□	□	□	□	Predictable
Complicated	□	□	□	□	□	□	□	Simple
Dull	□	□	□	□	□	□	□	Captivating
Tacky	□	□	□	□	□	□	□	Stylish
Unimaginative	□	□	□	□	□	□	□	Creative
Bad	□	□	□	□	□	□	□	Good
Ugly	□	□	□	□	□	□	□	Attractive

I would find the UX Cards useful…

… to practitioners	Unlikely	○○○○○○○	Likely
… to researchers	Unlikely	○○○○○○○	Likely
… for the design of interactive systems	Unlikely	○○○○○○○	Likely
… for the evaluation of interactive systems	Unlikely	○○○○○○○	Likely

Do you have any comment on this study or suggestion to enhance the UX Cards?

SEMI-STRUCTURED INTERVIEW

1. Please cite the cards that were the easiest to use and the cards that were the most difficult to use (3 cards max. for each). Comment your choices.
2. Which elemnts on the card were the most useful during the evaluation task and which ones were the least useful?
3. Feedback on the cards and how you would improve them.

APPENDIX 5 – LABORATORY EVALUATION (PAPER E)

TEST UTILISATEUR – FORMULAIRE DE CONSENTEMENT

Consentement de participation à un projet de recherche

Objectifs du projet et procédure
Cette expérience, d'une durée de 90 min. environ, vise à étudier le ressenti des utilisateurs durant l'interaction avec deux systèmes. Vous serez amenés à utiliser les systèmes en réalisant des tâches prédéfinies et à compléter plusieurs questionnaires visant à évaluer votre expérience. Nous n'évaluerons en aucun cas votre performance, mais bien l'expérience et le ressenti que vous inspireront chacun des systèmes.

Anonymat et confidentialité
Vous serez observé durant votre test et la session complète de test sera enregistrée sous forme vidéo. Le test auquel vous participez servira uniquement à recueillir des informations essentielles à ce projet de recherche sur l'expérience utilisateur et ne pourra en aucun cas servir à d'autres fins. Toutes les données recueillies sont stockées de manière anonyme et confidentielle. Les enregistrements effectués durant votre passation seront exclusivement exploités par Mlle Carine Lallemand (psychologue, Université du Luxembourg) et seront détruits définitivement à la fin de ce projet.

Risques potentiels
La participation à cette étude n'implique aucun risque connu.

Liberté de retrait
Votre participation à cette étude est volontaire. Vous êtes libre de vous retirer de cette étude à tout moment et sans justification.

Compensation
Pour vous remercier de votre participation, vous recevrez la somme de 30€ en liquide, remis à la fin de l'expérience.

Je déclare avoir pris connaissance de ce formulaire de consentement et avoir compris les conditions de ma participation à cette étude. J'ai eu l'occasion de poser des questions et j'ai obtenu toutes les réponses souhaitées, le cas échéant.

Fait à _____, le_____ en 2 exemplaires

Signatures

Pour toute question au sujet de ce projet de recherche, merci de contacter Mlle Carine Lallemand.

TEST UTILISATEUR – INSTRUCTIONS

Cette étude s'intéresse à la qualité de **l'expérience utilisateur**, c'est à dire le ressenti qu'un utilisateur va avoir durant l'interaction avec une technologie. Nous allons vous demander d'interagir avec 2 systèmes technologiques : le site web e-commerce **Amazon** et un **appareil photo numérique**. Après chaque interaction, vous devrez remplir un questionnaire pour évaluer votre ressenti.Afin de faciliter l'analyse des données, nous aimerions pouvoir récolter vos impressions pendant vos interactions avec les produits. Aussi nous vous demandons **d'exprimer à voix haute ce que vous pensez pendant toutes les tâches que vous réaliserez**. N'hésitez pas à exprimer vos difficultés et vos impressions, qu'elles soient positives ou négatives, ou toute autre réflexion qui vous passerait par la tête.
Essayez de **réaliser tous les scénarios proposés**. Si toutefois vous restez bloqué sur un scénario après plusieurs minutes, vous pouvez l'abandonner. **Les feuillets d'instructions vont vous guider tout au long de ce test. Suivez-les étape par étape en lisant bien les consignes.** Merci pour votre participation !

Quand vous aurez terminé l'ensemble des scénarios ainsi que les questionnaires sur l'ordinateur, nous reviendrons pour discuter avec vous de vos impressions.

QUESTIONNAIRE PRELIMINAIRE

Cette étude s'intéresse à la qualité de **l'expérience utilisateur**, c'est à dire le ressenti qu'un utilisateur va avoir durant l'interaction avec une technologie. Merci de compléter ce questionnaire préliminaire avant de passer aux scénarios d'usage. *Nous vous rappellons que toutes vos réponses sont anonymes et confidentielles. Merci de votre participation !*

Numéro d'anonymat : _____
Entrez ici le numéro d'anonymat qui vous a été attribué pour cette expérience.

A propos de vous :	
Vous êtes:	O un homme O une femme
Quel est votre âge ?	_____
Quelle est votre langue maternelle ?	O français O autre : _____
Quelle est votre catégorie socio-professionnelle ?	O Etudiant O Ouvrier O Employé O Cadre O Artisan, chef d'entreprise O Retraité O Sans activité professionnelle O Autres : _____
Globalement, à quel point vous sentez-vous à l'aise avec l'utilisation des technologies ?	*Echelle de Likert en 7 points de "pas du tout à l'aise" à "tout à fait à l'aise"*

Amazon	
Quel est votre niveau de familiarité avec le site e-commerce Amazon ?	*Echelle de Likert en 5 points de "pas du tout familier" à "tout à fait familier"*
Depuis combien de temps êtes-vous inscrit sur le site Amazon ?	O Depuis moins d'un mois O De 1 à 6 mois O De 6 mois à 1 an O Depuis plus d'un an
En moyenne, à quelle fréquence visitez-vous le site Amazon?	O Moins d'une fois par mois O Une fois par mois O Plusieurs fois par mois O Plusieurs fois par semaine O Chaque jour
Globalement, votre opinion au sujet d'Amazon est-elle plutôt positive ou négative ?	*Echelle de Likert en 5 points de "négative" à "positive"*

Appareil photo	
Quel est votre niveau de familiarité avec les appareils photo numériques en général ?	*Echelle de Likert en 5 points de "pas du tout familier" à "tout à fait familier"*
Possédez-vous un appareil photo numérique ?	O oui O non
Quel type d'appareil photo possédez-vous ? (choix multiple)	☐ Un appareil de type Compact ☐ Un appareil de type Bridge ☐ Un appareil de type Reflex ☐ Je ne sais pas
A quelle fréquence utilisez-vous votre appareil photo numérique ?	O Moins d'une fois par mois O Une fois par mois O Plusieurs fois par mois O Plusieurs fois par semaine O Chaque jour

Merci pour vos réponses ! **Vous pouvez à présent passer au premier cas d'étude et réaliser les tâches demandées dans le scénario.**

TEST UTILISATEUR – INSTRUCTIONS AMAZON

CAS D'USAGE N° ___ (rotation de l'ordre) - SITE E-COMMERCE AMAZON

 Loguez-vous sur votre compte Amazon en entrant votre adresse mail et votre mot de passe dans l'onglet « Mon compte ».
Cette étape vous permet d'accéder au site tel qu'il est personnalisé pour vous. La sécurité de vos données personnelles est une priorité et vous vous déconnecterez de votre session d'ici quelques minutes après la réalisation du test.

 Prenez quelques minutes pour explorer le site Amazon librement avant de réaliser les scénarios décrits à la page suivante.

N'oubliez pas svp de donner vos impressions et pensées à voix haute

TEST UTILISATEUR – SCENARIOS AMAZON

Scenario 1
Commencez ce scénario par la page d'accueil du site Amazon (si vous n'y êtes plus, retournez sur la page d'accueil). Consultez les recommandations personnelles en bas de la page d'accueil pour voir les produits qu'Amazon vous conseille. Ajoutez l'article de votre choix à votre liste d'envie. Si vous n'en avez pas encore, vous pouvez créer une liste.

Scenario 2
A partir du moteur de recherche, cherchez le livre « Le Petit Prince » d'Antoine de Saint-Exupéry (livres en français). Dans la page de résultats, cliquez sur le titre du premier lien pour consulter la page descriptive du livre. Une fois sur le descriptif du livre, feuilletez quelques pages du roman

Scenario 3
Consultez les commentaires des autres clients pour voir ce qu'ils pensent du livre « Le petit prince ». Regardez brièvement les pires commentaires et les meilleurs. Déclarez sur le site que vous trouvez l'un des commentaires utiles.

Scenario 4
Ajoutez le livre à votre panier d'achat et consultez le contenu de votre panier. Attention comme vous êtes connecté à votre compte, ne passez pas la commande et ne cliquez pas sur le bouton commande en 1 click !

Scenario 5
Retournez sur la page d'accueil. Parcourez les boutiques, rendez-vous dans la boutique « Vêtements et chaussures ». Consultez la page descriptive des chaussures de votre choix. N'oubliez pas de vous déconnecter de votre compte Amazon

Une fois les scénarios terminés, faites le nous savoir en disant à voix haute « j'ai fini les scénarios d'Amazon ». **Nous allons à présent vous fournir le questionnaire à remplir pour ce scénario.**

TEST UTILISATEUR – INSTRUCTIONS CAMERA

CAS D'USAGE N° ___ (rotation de l'ordre) – APPAREIL PHOTO NUMERIQUE

Tout d'abord, prenez quelques minutes pour explorer l'appareil photo librement avant de réaliser les tâches décrites ci-dessous. N'oubliez pas de donner vos impressions et pensées à voix haute.

TEST UTILISATEUR – SCENARIOS CAMERA

Scenario 1
Sortez l'appareil photo de sa housse et allumez-le.
Votre première tâche consiste à réaliser une photo en mode automatique « iAuto » (vous pouvez photographier ce que vous souhaitez autour de vous)

Scenario 2
Réalisez une vidéo de quelques secondes.

Scenario 3
Consultez les contenus (photo et vidéo) que vous venez de réaliser

Scenario 4
Effacez votre vidéo (mais conservez la photo). Vérifiez que la vidéo n'est plus présente.

Scenario 5
Prenez à présent une photo en mode « MAGIC » et « Punk ».
Une fois la photo prise, remettez l'appareil en mode iAuto.

Scenario 6
Naviguez dans le menu pour modifier la taille de l'image. Réglez-la sur 8M pixels.
Prenez une photo en 8M. Retournez dans le menu et remettez la taille de l'image à 14M.

Scenario 7
Explorez à présent le menu d'aide pour trouver quelles solutions s'offrent à vous si vos photos sont floues. Dès que vous avez vu que des solutions existent, il n'est pas nécessaire de les explorer.

Une fois les scénarios terminés, faites-le nous savoir en disant à voix haute « j'ai fini les scénarios de l'appareil photo ». **Nous allons à présent vous fournir le questionnaire à remplir pour ce scénario.**

TEST UTILISATEUR – QUESTIONNAIRE POST-SCENARIOS

Vous venez de réaliser les scénarios pour le cas d'étude [Amazon] ou [Appareil photo]. Merci de compléter ce questionnaire centré sur votre expérience et votre ressenti durant l'interaction.
A l'aide des paires de mots, merci d'indiquer ce que vous considérez être la description la plus appropriée pour [le site web d'Amazon] [l'appareil photo numérique]. Merci de cliquer sur votre choix pour chaque ligne !

(1/3) Je trouve [Amazon] ou [l'appareil photo]

Humain	□	□	□	□	□	□	□	Technique
M'isole	□	□	□	□	□	□	□	Me sociabilise
Plaisant	□	□	□	□	□	□	□	Déplaisant
Original	□	□	□	□	□	□	□	Conventionnel
Simple	□	□	□	□	□	□	□	Compliqué
Professionnel	□	□	□	□	□	□	□	Amateur
Laid	□	□	□	□	□	□	□	Beau
Pratique	□	□	□	□	□	□	□	Pas pratique
Agréable	□	□	□	□	□	□	□	Désagréable
Fastidieux	□	□	□	□	□	□	□	Efficace

2/3 Je trouve [Amazon] ou [l'appareil photo]

De bon goût	□	□	□	□	□	□	□	De mauvais goût
Prévisible	□	□	□	□	□	□	□	Imprévisible
Bas de gamme	□	□	□	□	□	□	□	Haut de gamme
M'exclut	□	□	□	□	□	□	□	M'intègre
Me rapproche des autres	□	□	□	□	□	□	□	Me sépare des autres
Non présentable	□	□	□	□	□	□	□	Présentable
Rebutant	□	□	□	□	□	□	□	Attirant
Sans imagination	□	□	□	□	□	□	□	Créatif
Bon	□	□	□	□	□	□	□	Mauvais

3/3 Je trouve [Amazon] ou [l'appareil photo]

Confus	□	□	□	□	□	□	□	Clair
Repoussant	□	□	□	□	□	□	□	Attrayant
Audacieux	□	□	□	□	□	□	□	Prudent
Novateur	□	□	□	□	□	□	□	Conservateur
Ennuyeux	□	□	□	□	□	□	□	Captivant
Peu exigeant	□	□	□	□	□	□	□	Challenging
Motivant	□	□	□	□	□	□	□	Décourageant
Nouveau	□	□	□	□	□	□	□	Commun
Incontrôlable	□	□	□	□	□	□	□	Maîtrisable

Globalement, sur une échelle de 1 à 5, comment évaluez-vous la qualité de votre expérience avec le site web Amazon ? _____
(1 désigne le moins bon score et 5 désignant le meilleur score)

QUESTIONNAIRE DE BESOINS UX (en cours de validation)

Nous allons à présent vous demander d'évaluer un ensemble de sentiments complexes au sujet de votre interaction avec [Amazon ; l'appareil photo]. Merci d'évaluer chacune des phrases suivantes sur une échelle allant de 1 (pas du tout) à 5 (totalement). Essayez d'être aussi précis que possible en nuançant vos réponses.

Quand j'utilise [Amazon; camera], j'ai le sentiment…	**Pas du tout**				**totalement**
…de réaliser des actions basées sur mes intérêts	□	□	□	□	□
…d'être libre de faire les choses à ma façon	□	□	□	□	□
…d'être libre de toute pression ou influence	□	□	□	□	□
…d'avoir des choix significatifs	□	□	□	□	□
…d'accomplir des tâches avec succès	□	□	□	□	□
…de maîtriser des situations complexes	□	□	□	□	□
…d'être très compétent dans ce que je faisais	□	□	□	□	□
…de pouvoir atteindre mes objectifs	□	□	□	□	□
…d'être performant	□	□	□	□	□
…d'un contact avec les gens en général	□	□	□	□	□
…d'être proche et connecté à des personnes importantes pour moi	□	□	□	□	□
…de compter pour les autres	□	□	□	□	□
…d'être conscient des émotions, activités et humeurs des autres	□	□	□	□	□
…de vivre de nouvelles activités	□	□	□	□	□
…de vivre d'agréables sensations	□	□	□	□	□
…d'avoir du plaisir (physique ou émotionnel)	□	□	□	□	□
…de découvrir de nouvelles sources et types de stimulation	□	□	□	□	□
…que les choses sont structurées et prévisibles	□	□	□	□	□
…de pouvoir souvent appliquer mes routines et habitudes	□	□	□	□	□
…d'agir en toute sécurité	□	□	□	□	□
…de comprendre comment les choses fonctionnent	□	□	□	□	□
…d'avoir le contrôle des évènements	□	□	□	□	□
…d'être une personne dont l'opinion compte pour les autres	□	□	□	□	□
…d'influencer les autres	□	□	□	□	□
…d'être quelqu'un que les autres prennent pour modèle	□	□	□	□	□
…d'être une personne appréciable	□	□	□	□	□
…que mes actions ont un but profond	□	□	□	□	□
…que mes actions sont conformes à mes valeurs	□	□	□	□	□
…de m'épanouir personnellement	□	□	□	□	□
…d'être une personne de valeur	□	□	□	□	□

A quel point les dimensions suivantes sont-elles importantes pour vous **quand vous utilisez [un site tel qu'Amazon] [un appareil photo numérique]** ?

Ne considérez pas leur importance dans votre vie en général, mais bien dans le contexte particulier d'utilisation d'Amazon. Les phrases suivantes proposent chacune deux idées qui peuvent paraître distinctes et vous faire hésiter. Si c'est le cas, basez votre jugement sur le sentiment qui vous correspond le plus.

Quand j'utilise [Amazon; camera], à quel point est-il important pour moi de :	**Pas du tout important**				**Très important**
Etre la cause de mes propres actions ou ne pas subir d'influence externe	□	□	□	□	□
Avoir des contacts avec les personnes importantes pour moi ou avoir le sentiment d'appartenir à une communauté	□	□	□	□	□
Etre compétent ou efficace dans mes actions	□	□	□	□	□
Etre en sécurité ou en contrôle de la situation	□	□	□	□	□
Prendre du plaisir ou être stimulé par des choses nouvelles	□	□	□	□	□
Etre aimé, respecté ou avoir une influence sur les autres	□	□	□	□	□
Développer mon meilleur potentiel ou donner du sens à ma vie	□	□	□	□	□

Quand j'utilise [Amazon; camera], j'ai le sentiment de :	**Pas du tout important**				**Très important**
Etre la cause de mes propres actions ou ne pas subir d'influence externe	□	□	□	□	□
Avoir des contacts avec les personnes importantes pour moi ou avoir le sentiment d'appartenir à une communauté	□	□	□	□	□
Etre compétent ou efficace dans mes actions	□	□	□	□	□
Etre en sécurité ou en contrôle de la situation	□	□	□	□	□
Prendre du plaisir ou être stimulé par des choses nouvelles	□	□	□	□	□
Etre aimé, respecté ou avoir une influence sur les autres	□	□	□	□	□
Développer mon meilleur potentiel ou donner du sens à ma vie	□	□	□	□	□

ENTRETIEN SEMI-DIRECTIF

Globalement, pouvez-vous me décrire votre expérience / ressenti ?
Estimez-vous que vous avez rencontré des difficultés pour réaliser les scénarios ?
Qu'est ce qui vous a le plus plu ?
Qu'est ce qui vous a le plus déplu ? ou frustré ?
Si vous deviez décrire votre expérience en 1 mot, vous diriez _____.
Selon vous, quels éléments de l'appareil photo contribuent à une expérience positive ?
Quels éléments contribuent à une expérience négative ?
Globalement, pensez-vous que (l'appareil photo / Amazon) est simple à utiliser ?
Est ce que vous seriez prêts à utiliser ce type d'appareil chez vous (ou en acheter) ?

This PhD research has been supported by the Fonds National de la Recherche, Luxembourg (AFR grant n° 1205972).