

How Relevant is an Expert Evaluation of User Experience based on a Psychological Needs-Driven Approach?

Carine Lallemand

Public Research Centre Henri
Tudor
29, avenue John F. Kennedy
L-1855 Luxembourg
carine.lallemand@tudor.lu

Vincent Koenig

University of Luxembourg
ECCS Research Unit
Route de Diekirch
L-7220 Walferdange
vincent.koenig@uni.lu

Guillaume Gronier

Public Research Centre Henri
Tudor
29, avenue John F. Kennedy
L-1855 Luxembourg
guillaume.gronier@tudor.lu

ABSTRACT

Many methods and tools have been proposed to assess the User Experience (UX) of interactive systems. However, while researchers have empirically studied the relevance and validity of several UX evaluation methods, few studies only have explored expert-based evaluation methods for the assessment of UX. If experts are able to assess something as complex and inherently subjective as UX, how they conduct such an evaluation and what criteria they rely on, thus remain open questions. In the present paper we report on 33 UX experts performing a UX evaluation on 4 interactive systems. We provided the experts with UX Cards, a tool based on a psychological-needs driven approach, developed to support UX Design and Evaluation. Results are encouraging and show that UX experts encountered no major issues to conduct a UX evaluation. However, significant differences exist between individual elements that experts have reported on and the overall assessment they made of the systems.

Author Keywords

User experience evaluation; psychological needs; expert evaluation; UX cards.

ACM Classification Keywords

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

INTRODUCTION

User experience (UX) is commonly described in the literature as the overall quality of the interaction between a user and an interactive system. This concept, described as a “truly distinct and extended perspective on the quality of interactive products” [13], has been growing in popularity

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. *NordiCHI '14*, October 26 - 30 2014, Helsinki, Finland
Copyright 2014 ACM 978-1-4503-2542-4/14/10...\$15.00.
<http://dx.doi.org/10.1145/2639189.2639214>

for more than a decade. In a post-materialistic society [10], where the “experience economy” [26] takes a prominent place, issues related to the design and evaluation of user experience become crucial.

Many methods and tools have been proposed to assess UX. Nearly eighty of them have been identified and categorized by researchers [25, 35] according to:

- *the method type* (field studies, lab studies, online studies, questionnaires/scales)
- *the development phase* (scenarios/sketches, early prototypes, functional prototypes, products on market)
- *the studied period of experience* (before usage, during interaction, long-term UX)
- *the evaluator / information provider* (UX experts, single user, group of users)

This paper focuses on UX evaluation, conducted by expert evaluators. We will first introduce expert evaluation techniques in general before focusing on previous attempts made to apply this method to UX evaluation. We will also introduce the psychological-needs driven approach for UX. In the second part of this paper, we will present the methodology we deployed to explore UX expert evaluation challenges and issues. This paper focuses on two research questions and presents partial results of our global study. Those results will be presented and discussed in the last section of this paper.

EXPERT-BASED EVALUATION METHODS IN HCI

Among Human-Computer Interactions (HCI) evaluation methods, inspection methods involve the inspection of the interface by an evaluator [22]. Developed in the 1990’s as discount usability engineering methods [23], inspection methods are described as cheap, fast and easy to use [22]. Unlike user-based methods where the evaluation relies on the observation of users performing a set of tasks while actually interacting with a system, the evaluation of a system through experts-based methods relies solely on the expertise and judgment of the evaluator [7]. Heuristic evaluation [24] and cognitive walkthrough [36] are the most common usability inspection methods and have been extensively used by HCI practitioners for more than three decades [7].

Heuristic evaluation (HE) involves having a small set of evaluators examining the interface and judging its compliance with recognized usability principles (called “usability heuristics” [22]). Research work done on HE has shown however that a considerable amount of issues identified by an expert evaluator are actually based on his expertise or on judgment and not on the set of heuristics used to assess the system [3]. The term “expert review” [21] therefore designates a less formal evaluation where an experienced expert does not use a single set of heuristics but rather bases his report on his knowledge of users’ tasks, HCI guidelines and standards and on his personal experience. The main limitations of HE result from expert variability (different evaluators find different sets of problems for the same interface) [14, 16] and overestimation of the true number of problems, also called False Alarms [16]. Since HE has been criticized for its low validity and limited reliability, it is recommended to use it in combination with user-based methods like user testing [16].

Following the evolutions of the HCI field, several heuristic sets have been developed to take into account new concerns beyond usability. We can mention for example playability heuristics [19], heuristics for human-environment interaction in virtual worlds [2], for learning experience [30] or social activity [20]. Regarding UX, Väänänen-Vainio-Mattila & Wäljas [33] 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”) deals with hedonic or subjective aspects of the experience. In the revised version of these heuristics, H6 is replaced by a service usability heuristic and a trust and safety heuristic, thus restraining the scope of the evaluation to mainly pragmatic issues.

The aforementioned sets of heuristics have in common that they are specifically relevant for certain types of systems or situations. Recently, generic UX heuristics were proposed for the design and evaluation of systems providing users with a positive experience [4, 1]. Colombo & Pasch [4] derived their ten heuristics for optimal UX from the flow theory [5]. Arhipainen [1] 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, she proposed the 10 UX Heuristics, based on empirical UX studies.

Apart from heuristics, other solely UX expert-based evaluation methods are scarce. Out of their collection of 96 UX evaluation methods, Vermeeren et al. [35] identified only 13 expert methods of which 6 require users or groups of users in addition to the expert. Seven methods are described as purely expert-based. In 2003, Jordan [17] proposed the concept of immersion where the investigator herself uses the system in real contexts and evaluates it. More recently, Wilson [37] 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*”. While interesting, this approach has not yet been studied empirically.

PSYCHOLOGICAL NEEDS-DRIVEN UX EXPERT EVALUATION

An extensive amount of studies conducted within the field of positive psychology have demonstrated that psychological needs are particular qualities of experience that all people require to thrive [6, 28, 29]. Sheldon [27, 28] highlighted the importance of psychosociological needs by showing that these are both necessary inputs and driving motives. The transfer of this assumption to the field of HCI has led to psychological needs-driven UX approaches [12, 18, 31]. The fulfilment of human psychological needs is thought to be a main trigger of positive experiences with technologies [12]. In order to design experiential products, designers should consider interactive systems as means to fulfil needs (“be-goals”) and not only means to achieve task oriented “do-goals” [12]. Needs provide categories of experiences, such as “competence experiences” or “relatedness experiences” [10] that UX practitioners should seek to design.

The concern for basic human needs in UX expert-based evaluation has already been included as one of the ten heuristics for optimal UX developed by Colombo & Pasch [4]. Their 9th heuristic, entitled “Know the user’s motivations” states that “the system should help users to fulfill the motivations behind its use and to satisfy basic psychological needs”. However, no details are given regarding the relevant needs to fulfil or their definition.

RESEARCH QUESTIONS

Combining UX expert-based evaluation with a psychological needs-driven approach, the present study has two purposes. First, we aimed at exploring the adequacy of expert evaluation in the context of UX evaluation. As expert evaluation is a technique initially developed to assess the usability of a system, our objective is to study whether this method is transferable to the evaluation of UX. We know that practitioners considered expert evaluation as a cheap and effective method to assess the usability of the system, even described as “*discount usability*” [22, 23]. It allows them to fix basic problems and spend less money for user testing. Now that the focus has switched from usability to UX, we assume that practitioners may apply expert evaluation to UX in the same way they were using it to assess usability. It is therefore necessary to explore the adequacy of expert evaluation in the context of UX assessment and the conditions under which this method might be valid. We therefore will try answering the following questions: (1) Are experts actually able to

conduct a UX expert evaluation? (2) How do they proceed and (3) on which elements is their assessment based?

In addition, we also wanted to explore the relevance of a psychological-needs driven approach for the assessment of UX. Seven basic human needs (out of the ten basic needs summarized by Sheldon et al. [28]) were selected and represented under the form of UX Cards to be used as an evaluation tool. The design of the UX cards and their use will be presented in detail in the next section. Our research question sounds as follows: do the basic human needs (under the form of UX Cards) constitute a relevant framework to evaluate UX? Do the experts find the UX Cards easy to use and useful for the assessment of UX?

METHODOLOGY

Design of UX Cards

Based on the literature on fundamental human needs [27, 28] and on studies linking these needs to the UX of interactive systems [11, 31, 9], we selected 7 candidate needs supposed to be relevant in the context of UX design and evaluation (Table 1). Following Hassenzahl [12], the needs for luxury, self-esteem and physical thriving were excluded from our selection either because of their incapacity to emerge as a primary need (luxury), their weak connection with interactive technology (physical thriving) or their ambivalence (self-esteem might be an outcome of other needs fulfilment).

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

Table 1. The seven needs represented on UX Cards and their definition [28, 11]

Seven UX cards were designed (Figure 1). Each card is composed of: a title, a definition of the need (adapted from [28]), linked terms (synonyms and keywords), real-life examples of elements or situations able to trigger the fulfillment of the need and finally main scientific references related to this need. Five pictures representing each need were also included on each UX card to enhance visual design and attractiveness.

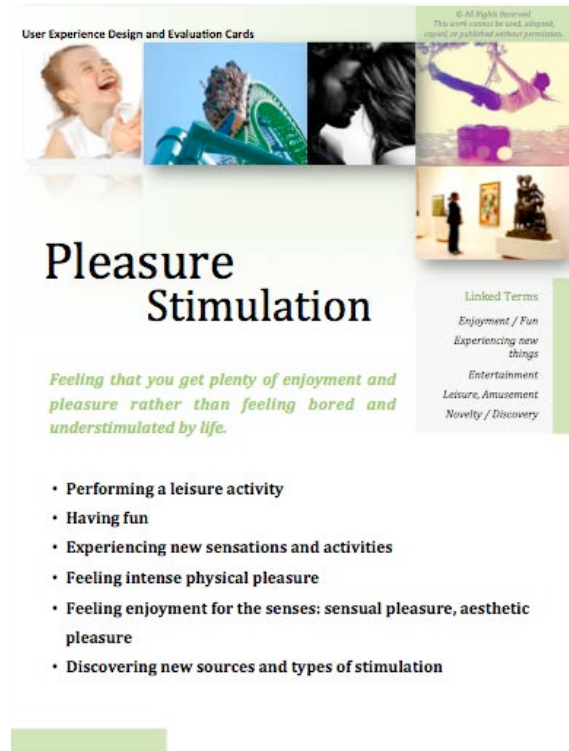


Figure 1. Example of the UX Card “Pleasure / Stimulation”

Even if the UX Cards are designed to be used as an expert evaluation technique, they might not be considered as heuristics. Our 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 fulfillment of human needs, leading to a positive experience.

In the present study we have provided a selection of interactive products already on the market as use cases for the UX Cards. However, it should be emphasized that the UX Cards might be used to conduct a UX expert 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 [8].

Expert Evaluation Participants

Thirty-three UX experts (16 women et 17 men) participated in this study. They were recruited either by personal contact or through a request on social networks. All of them were able to read and understand English material. 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 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). Note that Human-Computer Education degrees in France are often Masters' degrees achieved after a degree course in Psychology.

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 %
Technology / Software	4	12.1 %
HCI	2	6.1 %

Table 2. General profiles of the experts

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).

Procedure

The study took place at several locations in France and Luxembourg, most of the time at the workplace of each participant. Each individual session lasted approximately 2 hours. Participants received a 50€ (about 68 US\$) shopping voucher in compensation for their time spent.

Interactive systems: Four interactive systems were inspected during the experiment: a) the social network Facebook ® (FB) b) the online e-commerce website Amazon ® (AMZ) c) the game Angry Birds ® (AB) on iPhone ® 5S and finally d) an Olympus digital camera. We made the choice of 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 (Table 3).

System	Min	Max	Mean	SD
Facebook	1	7	6.09	1.72
Camera	2	7	5.73	1.35
Amazon	1	7	5.70	1.49
Angry Birds	1	7	4.45	2.12

Table 3. Familiarity level with the use cases

General instructions and preliminary survey: After having welcomed each participant, we first explained the main goals of the study and the theory underlying the UX cards. We then invited each expert to fill in a preliminary survey, indicating gender, age, country of residence. Participants also indicated data about their professional background: domain, role, educational background, job title, experience in the HCI field, level of expertise with expert evaluation, level of familiarity with the concept of UX and level of familiarity with psychological needs theories.

Understandability of the UX Cards: Then, to familiarize the expert with the UX cards, 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.

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.

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 fulfillment of the need for (name of the need)?". The scales ranged from "negative" to "positive". (Table 4).

Overall, what is the impact of (name of the assessed 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

Table 4. Overall assessment of UX performed after each evaluation task

Reporting tool: We chose to provide participants with a paper-based grid to report UX elements during the identification step. This choice has been made to avoid a continuous shift between the tested interactive system and the reporting grid. 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.

After task completion, we collected experts' opinions during a debriefing interview.

RESULTS

The present paper focuses on the results related to the understandability of the UX Cards, the UX Evaluation task and the perceived usefulness of the UX Cards. More detailed results regarding the link between needs and UX experience will be presented in a future paper.

Understandability of the UX Cards

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). Understandability ratings for each card are presented in Table 5.

Understandability	Min	Max	Mean	SD
Security / Control	5	7	6.64	.60
Influence / Popularity	5	7	6.58	.61
Relatedness / Belongingness	5	7	6.42	.61
Autonomy / Independence	5	7	6.39	.70
Pleasure / Stimulation	3	7	6.39	.93
Self-Actualizing / Meaning	3	7	6.03	.92
Competence / Effectiveness	3	7	5.97	1.1

Table 5. Understandability of the UX Cards, ranked in descending order

Anticipated operationalizability of the cards (i.e. imagined ease of using the cards in the context of a UX evaluation) was assessed as satisfactory as well (M=5.78, SD=0.65) (Table 6). However, the need for Self-Actualizing was assessed as harder to operationalize than the others (M=3.82, SD=1.8).

Operationalizability	Min	Max	Mean	SD
Influence / Popularity	3	7	6.36	.93
Security / Control	5	7	6.36	.78
Pleasure / Stimulation	4	7	6.15	.94
Relatedness / Belongingness	4	7	6.15	.83
Autonomy / Independence	3	7	6.09	1
Competence / Effectiveness	2	7	5.55	1.6
Self-Actualizing / Meaning	1	7	3.82	1.8

Table 6. Operationalizability of the UX Cards, ranked in descending order

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 UX Cards.

UX Evaluation Task: Identification of Elements and Related UX Needs

Overall, the experts identified 1794 elements, which corresponds to an average of 54.4 elements per expert and 13.6 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) do 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. 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 2). It thus seems that the appropriation time required for the UX Cards is relatively short (about 15 minutes).

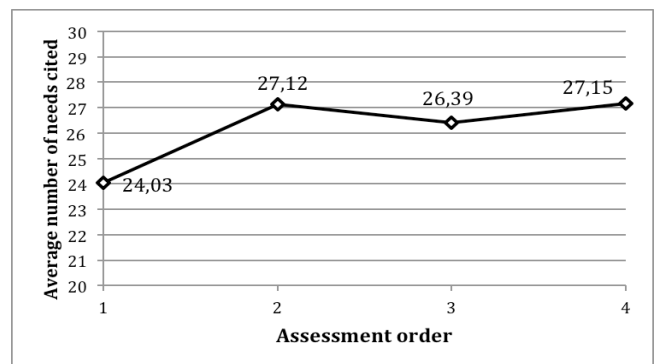


Figure 2. Average number of cited needs according to the order of the evaluation

Most cited needs (Figure 3) 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 use, as not many interactive systems succeed in fulfilling such a need.

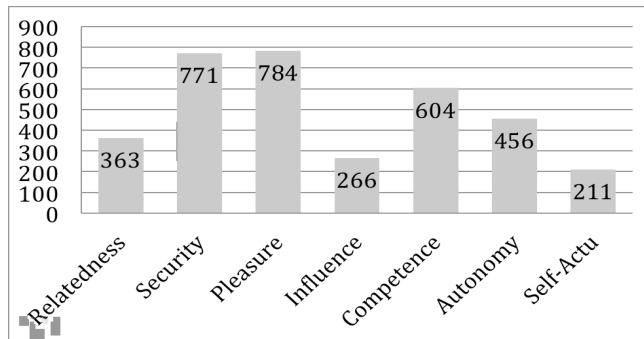


Figure 3. Total number of cited needs (considering both positive and negative) during the UX evaluation task

Before starting the evaluations, experts were varyingly experienced with our four use cases (Table 3). This level of familiarity significantly correlates with some evaluation variables, especially in the case of FB: The experts' level of familiarity with FB positively correlates with the number of elements identified ($r=.361, p<.05$) and with the total number of cited needs ($r=.495, p<.05$). Interestingly, it is also related to the number of needs cited as positive ($r=.478, p<.05$), but not to the number of needs cited as negative. In other words, the more familiar experts are with FB, 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 ($r=-.363, p<.05$), especially positive needs ($r=-.415, p<.05$). As Age and Seniority are of course strongly related ($r=.896, p<.01$), results also show that senior experts tend to cite significantly less positive needs than less experienced practitioners ($r=-.358, p<.05$). Self-reported familiarity with UX, familiarity with UX needs theories and the level of expertise with HE do not significantly impact the total number of elements or needs cited by the experts. The same holds for Gender, Domain, Role or Education.

Overall UX Evaluation of the Interactive systems

In order to understand how UX experts actually assess the UX of interactive systems, we compared the evaluation made through the identification task (number of cited elements and needs) to the overall evaluation (7-points Likert scales) conducted after each evaluation task (Table 4). 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 4 presents the overall UX evaluations of the four interactive systems. An overall UX evaluation score has been computed by averaging the scores of the 7 individual needs.

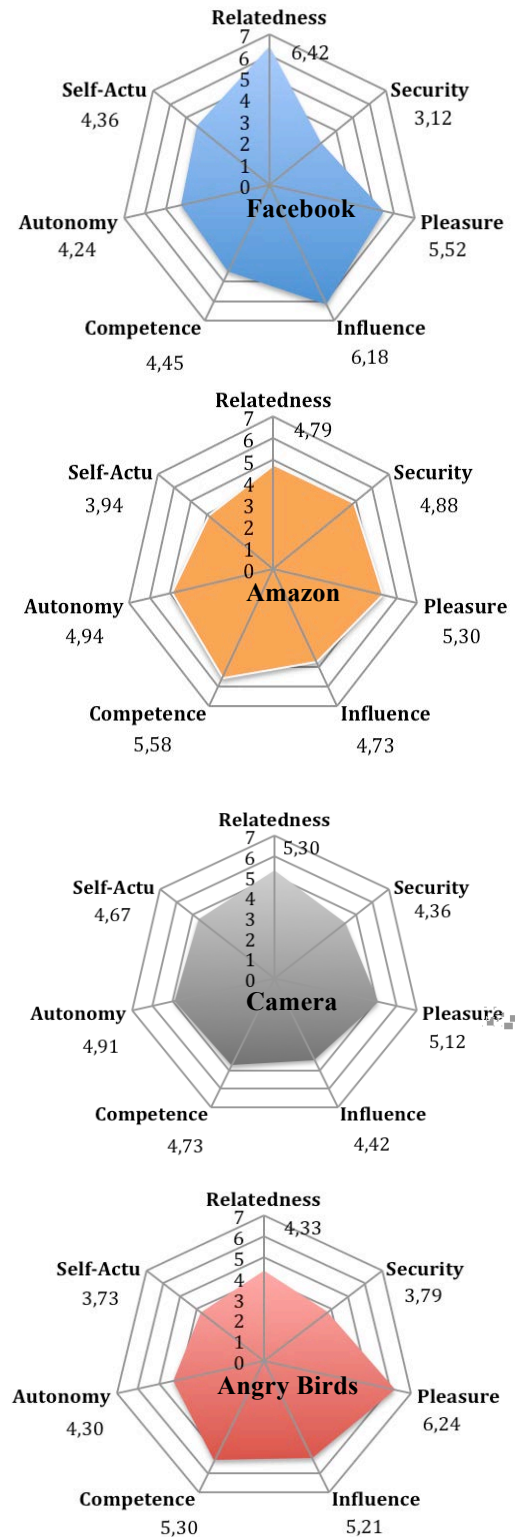


Figure 4. Overall UX Evaluation of the interactive systems. Average ratings (n=33) are presented under each need.

Background variables such as Age, Seniority, Familiarity with UX, Familiarity with needs theories or number of HE performed are not significantly correlated to the overall assessment of our use cases, except for the Camera. In this case, Age and Seniority are negatively correlated to the overall UX assessment ($r=-.670$ and $r=-.527$ respectively, $p<.05$). The assessment of the Camera also differs significantly according to the level of expertise with HE ($r=-.474$, $p<.05$). 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 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 ($r=.465$, $p<.01$). The same holds for Amazon ($r=.446$, $p<.01$) and Angry Birds ($r=.366$, $p<.05$). 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 UX 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 ($r=.389$, $p<.05$). This suggests that the overall UX a posteriori assessment is globally not influenced by the number of elements an expert has identified. The 15-minutes duration of the evaluation task might explain this phenomenon.

Systematic significant links were however observed between the overall UX assessment and the number of needs cited positively or negatively. In all cases, the overall UX assessment is positively correlated to the number of positive needs cited and negatively correlated to the number of negative needs cited. Table 7 presents the correlation coefficients for each use case.

Overall UX assessment	Number of needs cited as positive	Number of needs cited as negative
Facebook	.553**	-.409*
Amazon	.625**	-.406*
Angry Birds	.570**	-.554**
Camera	.708**	-.562**

Table 7. Correlations between overall UX assessment and number of needs cited positively or negatively
 (* significant at $p<.05$ level; ** significant at $p<.01$ level)

This observation suggests that the overall assessment is mainly based on the evaluation task, i.e. 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 may demonstrate this deduction is not systematically valid.

In the case of FB (other use cases will not be detailed in the present paper due to space constraints), the overall UX assessment is significantly correlated only to the number of positive and negative needs for Security ($r=.540$ and $r=-.296$ respectively, $p<.05$), Autonomy ($r=.511$ and $r=-.409$ respectively, $p<.05$) and Self-Actualizing ($r=.465$ and $r=-.323$ respectively, $p<.05$). Moreover, the links between specific a posteriori UX assessment (for each need) and the number of times the related needs were really cited as positive and negative, are not systematically significant. In the case of FB, the evaluation of Relatedness, Pleasure and Competence did not rely on the evaluation 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. Age and Seniority are only positively correlated to the assessment of the need for Relatedness ($r=.333$ and $r=.386$ respectively, $p<.05$). Similarly, Role and Domain are also linked to Relatedness. Experts from Academia ($M=5.6$, $SD=1.14$) assessed FB relatedness as less positive than experts from Industry ($M=6.65$, $SD=.49$) ($F(2,32)=3.39$, $p<.05$). Practitioners assessed FB relatedness more positively than researchers ($F(2,32)=3.46$, $p<.05$). No significant differences were observed according to gender, level of expertise with HE, the familiarity with UX or with needs theories. Finally, the level of familiarity with FB is positively correlated to the specific assessment of the need for Security ($r=.425$, $p<.05$), Autonomy ($r=.313$, $p<.05$) and Self-Actualizing ($r=.467$, $p<.05$). These three needs are exactly the same significantly correlated with the overall UX assessment.

Perceived Usefulness of the UX Cards

Perceived usefulness of the UX Cards was assessed at the end of the evaluation using four 7-points Likert scales. 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 believe 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.

DISCUSSION

As expert evaluation is seen by practitioners as a cheap and effective method to assess usability [22], 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 might be 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. Our results show that UX experts encountered no blocking issues in conducting an UX expert evaluation by using the UX

Cards. High levels of perceived usefulness and high amounts of cited elements and needs during the evaluation task indicate the effectiveness of this approach from an expert's perspective.

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. UX expert evaluation using the UX Cards will therefore differ from a usability evaluation, where the main focus is on the identification of flaws and issues. As we mentioned before, the goal of the UX Cards is not to "debug" a system but to support assessing the system in its ability to fulfill human needs, ultimately leading to a positive experience.

Regarding the relevance of the needs-driven approach for UX evaluation, it is worth noting that 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. The preliminary self-assessment of the UX Cards shows that experts were able to understand the cards easily. Similarly, the number of elements identified and needs cited during the identification task provides evidence for the usability of the need driven approach.

Some differences were 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. The second part of the current study, which is ongoing and will be presented in a second paper, explores what differentiates UX experts in their ability to predict UX by using the UX Cards. To address this research question, we will compare the results of our UX expert evaluations to the results of user tests performed on the same four interactive systems. This will give us valuable insights on the importance of expertise selection (i.e. the process of choosing an expert from a list of recommended people) [38] or on the necessity to use multiple experts to conduct an accurate UX evaluation. In this latter case, the need to find enough experts with the required expertise could impede on the practicability of an expert based method for UX evaluation [35].

Experts were also less used to assess 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. The significant correlation between the familiarity level with a system and its overall evaluation seems problematic since it implies that

subjectivity comes into play. On this point, experts also mentioned that the use of the first person singular in the UX Cards (e.g. pleasure / stimulation: "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. 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. Hornbaek [15] 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. 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. Finally, 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.

This study has also shown some limitations. First, we noticed that some experts felt a bit lost because of their complete freedom / lack of constraints for the evaluation. Most of them felt somewhat uncomfortable deciding what kind of elements they should identify. 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) before (2) assessing the system from a functional perspective (e.g. features, interoperability, interaction design). Then, (3) the evaluation should focus on detailed user interface elements (e.g. content, information design, usability issues). Finally, (4) we will invite the experts to reflect on missing elements, i.e. 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 of their evaluation and also to harmonize their practice in case of multiple experts assessing the same system or product.

Finally, the 15-minutes duration of the evaluation task might be debatable. We were aware that this is undoubtedly a short time to achieve an expert evaluation task, however it allowed each expert to evaluate four different systems without spending several days on the task. We also hoped 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 Digital 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. Finally, results suggest that the overall UX a posteriori assessment is globally not influenced by the number of elements an expert has identified. This is unsurprising considering the fact that the 15-minutes timespan did not allow for a comprehensive evaluation.

CONCLUSION

In the present study, 33 experts performed a UX evaluation on 4 interactive systems. We provided the experts with UX Cards, a tool developed to support UX Design and Evaluation, based on a psychological needs-driven approach. Very few studies have explored the relevance of expert evaluation as a UX evaluation method, and to the best extent of our knowledge, none of them followed a psychological needs-driven approach (i.e. basic human needs are considered as the drivers of UX). Our results show that UX experts encountered no issues in conducting an UX expert evaluation by using the UX Cards. 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.

Expert evaluation as a method has been criticized by previous research on usability. The main limitations highlighted are related to the inadequacy between the expert evaluation and the problems reported by users. Moreover, an expert evaluation of UX inherently encompasses additional issues and challenges to overcome. In both our observations and the few research works done so far on UX expert evaluation [32, 33, 1], the difficulty for an expert to adopt the perspective of the user was identified as a crucial issue. Many questions remain unanswered and need further empirical research: Does a UX evaluation performed by experts accurately reflect the experience felt by users? Or does the primarily subjective nature of UX counter-indicate the use of expert-based evaluation? Are some UX experts better than others in assessing UX? User tests are currently conducted in order to compare expert evaluations to empirically measured user experiences. Are expert evaluations accurate enough to be considered as a valid method in the context of UX? Or are expert evaluators

unable to predict UX without fully involving the user in the evaluation? We expect the results of this on-going study to support Industry in the choice and use of relevant UX evaluation methods and to provide the UX research field with valuable insights regarding the links between needs fulfilment and UX outcomes.

In addition, complementary research work is currently conducted on the UX Cards, especially regarding their use during the design phase as a UX design method.

ACKNOWLEDGMENTS

The present project is supported by the National Research Fund, Luxembourg. The authors would like to thank all the experts who kindly accepted to participate in this study and shared their thoughts and ideas with great enthusiasm.

REFERENCES

1. Arhippainen, L. (2013). A Tutorial of Ten User Experience Heuristics. In Proc. AcademicMindTrek '13, Tampere, Finland.
2. Bach, C. (2004). Elaboration et validation de critères ergonomiques pour les interactions homme-environnements virtuels. Thèse de Doctorat. Université Paul Verlaine, Metz.
3. Cockton, G. and Woolrych, A. (2001). Understanding inspection methods: lessons from an assessment of heuristic evaluation. In: A. Blandford, J. Vanderdonckt, and P.D. Gray, eds. Proceedings of people and computers XV: joint proceedings of HCI 2001 and IHM 2001. Berlin: Springer-Verlag, 171–192
4. Colombo, L. and Pasch, M. (2012). 10 Heuristics for an Optimal User Experience. Proc. CHI2012 Altchi. ACM Press.
5. Csikszentmihalyi, M. (1990). Flow – The Psychology of Optimal Experience. Steps Toward Enhancing the Quality of Life. New York, Harper Perennial.
6. 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.
7. Dillon, A. (2001) Usability evaluation. In W. Karwowski (ed.) *Encyclopedia of Human Factors and Ergonomics*, London: Taylor and Francis.
8. Dix, A., Finlay, J., Abowd, G. & Beale, R. (2004). *Human-Computer Interaction*. Prentice Hall.
9. 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
10. Hassenzahl, M. (2013). Experiences Before Things: a Primer for the (Yet) Unconvinced. Proc. Of CHI '13 Extended Abstracts, 2059-2068.

11. Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers*, 22(5), 353-362.
12. Hassenzahl, M. (2010) Experience Design: Technology for All the Right Reasons. Synthesis Lectures on Human-Centered Informatics, 3, 1, 1-95.
13. Hassenzahl, M. (2008) User Experience (UX): Towards an experiential perspective on product quality. Proceedings of IHM'08, Metz, France.
14. Hertzum, M., & Jacobsen, N.E. (2001). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human Computer Interaction*, 13(4), 421-443.
15. Hornbæk, K. (2010). Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology*, 29: 1, 97 – 111
16. Hvannberg, E.B., Law, E. & Larusdottir, M.K. (2007). Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers*, 19, 225–240.
17. Jordan, P. (2000). Designing pleasurable products: An introduction to the new human factors. CRC Press.
18. 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 Proc. DUXU'2011.
19. Korhonen, H. and Koivisto, E. (2006). Playability heuristics for mobile games. Proc. MobileHCI'06, ACM Press, NY, USA, 9-16.
20. Malinen, S. & Ojala, J. (2011). Applying the Heuristic Evaluation Method in the Evaluation of Social Aspects of an Exercise Community. Proceedings of DPPI 2011.
21. Molich, R. & Jeffries, R. (2003). Comparative experts reviews. Proc. ACM CHI'EA 2003 Conference.
22. Nielsen, J. (1993). *Usability Engineering*. New York: AP Professional.
23. Nielsen, J. (1989). Usability engineering at a discount. In Salvendy, G., and Smith, M.J. (Eds), *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*. Elsevier Science Publishers, Amsterdam, 394-401.
24. Nielsen, J. & Molich, R. (1990). Heuristic evaluation of user interfaces. Proc. ACM CHI'90 Conf., 249-156.
25. Roto, V., Obrist, M., Väänänen-Vainio-Mattila, K. *User Experience Evaluation Methods in Academic and Industrial Contexts*. In User Experience Evaluation Methods in Product Development (UXEM'09). Workshop in Interact'09, Sweden.
26. 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.
27. Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: A two process model. *Psychological Review*, 118(4), 552-569.
28. 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.
29. 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
30. Squires, D. & Preece, J. (1999). Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, 11(5), 467-483
31. Tuch, A. N., Trusell, R. & Hornbæk, K. (2013). Analyzing Users' Narratives to Understand Experience with Interactive Products. In *Proceedings CHI '13*.
32. Väänänen-Vainio-Mattila, K., Wäljas, M. (2009). *Developing an Expert Evaluation Method for User eXperience of Cross-Platform Web Services*. Proceedings of Mindtrek'09, ACM
33. Väänänen-Vainio-Mattila, K., Wäljas, M. (2009). Development of Evaluation Heuristics for Web Service User Experience. A Work-in-Progress paper in Proceedings (Extended Abstracts) of CHI'09, ACM
34. Väänänen-Vainio-Mattila, K., Roto, V., Hassenzahl, M. Towards Practical User Experience Evaluation Methods. Proceedings of the 5th COST294-MAUSE Open Workshop on Meaningful Measures: VUUM 2008, Reykjavik, Iceland.
35. Vermeeren, A, Lai-Chong Law, E., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K. User Experience Evaluation Methods: Current State and Development Needs. In Proc. of NordiCHI'10, ACM.
36. Wharton, C. Rieman, J., Lewis, C. and 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.
37. 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 19 April 2013).
38. Yarosh, S., Matthews, T., and Zhou, M. (2012). Asking the Right Person: Supporting Expertise Selection in the Enterprise. Proc. Of ACM CHI 2012.