Negative Emotions, Positive Experience: What Are We Doing Wrong When Evaluating the UX?

Walter T. Nakamura
Federal University of Amazonas
Manaus, AM, Brazil
walter@icomp.ufam.edu.br

Elaine H. T. de Oliveira
Federal University of Amazonas
Manaus, AM, Brazil
elaine@icomp.ufam.edu.br

Tayana Conte
Federal University of Amazonas
Manaus, AM, Brazil
tayana@icomp.ufam.edu.br

ABSTRACT

Since its idealization more than 20 years ago, much research has been carried out on the User eXperience (UX) field, with several evaluation methods being proposed. However, studies have been pointing out conflicting results when evaluating the UX. Users frequently evaluate their UX as positive, even when experiencing many negative emotions when interacting with a product. Moreover, variables such as the peak-end effect and the memory-experience gap may also have been influencing the results, leading to misinterpretations about the product's quality. In this context, this paper presents our work-in-progress research on the following research question: "What are we doing wrong when evaluating the UX?". We discuss about different variables that may have been influencing users' perceptions about their experiences in previous studies and highlight research opportunities. With

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313000

KEYWORDS
User eXperience; Evaluation methods; Human-Computer Interaction.

CCS CONCEPTS
Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods

this work, we expect to shed light on and bring reflection to current practices on UX evaluations in order to progress the research in the UX field.

INTRODUCTION

User eXperience (UX) is a concept that was widely and quickly disseminated and accepted by the HCI community [10]. It was conceived when researchers realized the limitations of traditional usability framework, which focuses primarily on user cognition and performance [10], making it too narrow to represent a holistic vision of human-computer interactions [9]. In this context, researchers felt the need for a broader concept that considers non-utilitarian concepts, such as fun, pleasure, hedonic and ludic, shifting from material to experiential [7], giving rise to the UX field.

Today, more than 20 years after its popularization, much research has been carried out on the topic, with various methods designed to evaluate the UX of interactive products in different situations and interaction periods [19]. However, despite the efforts [9–11], the HCI community still did not reach a consensus about the concept of UX, making it a buzzword that is increasingly being used, even without a clear definition or scope [9]. As a consequence, many authors have been developing UX evaluation methods considering different aspects to be evaluated, usually without a clear definition about them or without verifying their relation to established constructs [2]. Moreover, divergences between the results from different methods, as well as the positive evaluations from users who faced many negative emotions, raise the question about whether we are really measuring the UX.

This paper presents a work-in-progress on the topic, discussing the current challenges and gaps on UX evaluation research. We expect to foster discussion about current practices on UX evaluation by exposing aspects that researchers might have not been considering during the evaluation process and may be biasing the results.

CURRENT CHALLENGES AND GAPS ON UX EVALUATION RESEARCH

The concept of UX and its implications in the design of evaluation methods

Despite the efforts to scope and define its concept [9–11], many papers still rarely link their studies to UX theory, providing only a vague definition of UX [17]. Moreover, its association with a broad range of fuzzy and dynamic concepts, including emotional, affective, experiential, hedonic, and aesthetic variables makes it difficult to get a universal definition [10]. As a result, multiple definitions and interpretations have been proposed, fragmenting the UX research in diverse theoretical models that focus in different aspects (e.g. pragmatic, experience, pleasure) [10]. ISO 9241-210 [1], for instance, defines UX as "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service", which includes "all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and

after use". Such generic definition leaves room for a broad range of attributes that researchers may consider when evaluating the UX, making it difficult to identify which dimensions of experience to assess. It may also lead to the arbitrary inclusion or exclusion of particular variables according to the author's background and interest [10]. Literature reviews on UX evaluation methods [2, 17], for instance, identified various evaluated dimensions beyond the three core UX dimensions (emotions and affect, enjoyment and aesthetics), such as enchantment, engagement, trust and relevance. Although this variety of dimensions inspires new ways of thinking about experience, the relation between them and the stablished constructs are not usually made clear, which may lead to the creation of an endless number of different dimensions evaluating the same phenomenon [2]. Finally, the vague link between UX evaluations and theory makes it difficult to formulate improvements [17].

Vision for future research: we are aware of the difficulty in defining a general and well-accepted concept of UX. However, given the implications for current and future research pointed above, we encourage the HCI community to keep their efforts towards a better scoping and definition of UX. Researchers, for instance, may carry out studies to map which dimensions and attributes are being considered by academia and industry when evaluating the UX, as well as promote discussions on the topic in conference workshops. A better scope and definition of UX may facilitate researchers to identify which constructs to choose, thus making it possible to design more focused UX evaluations.

Constructs may not be measuring what is intended to be measured

Conflicting results from different studies have been raising uncertainties about their effectiveness in capturing the UX. Variables that are not being considered, as well as inadequate constructs may have been affecting the accuracy of the evaluations. Studies, for instance, have been pointing out that users may make evaluations based on their perceptions, as opposed to the reality. A study from Metatla et al. [14] to evaluate the effects of cross modal perceptions in an audio-visual interface revealed that users found the audio-visual condition easier than the visual-only condition, when, actually, the data showed no improvements in their score. Users may also evaluate based on their conceptions about an attribute, which can be different from those comprehended by researchers. A work from MacDonald and Atwood [12], which aimed to verify what aspects make a system useful, pointed-out that users had a different interpretation of "fun" and "enjoyable", associating them with leisure activities like playing video-games or spending time with friends. This led them to tend to provide lower scores for the system regarding these two attributes. As a consequence, the results indicated that participants who found systems useful did not find them enjoyable. Such results raise questions about whether the constructs are adequate for evaluating the UX and whether they are really measuring what is intended to be measured. This issue is even more worrying when we consider that there is a lack of validation of new proposed methods [2] and their link to UX theory [17], raising uncertainties on the validity of the results. Additionally, little research on the comparison of different UX evaluation methods have

also been carried out [2], making it difficult to gather their strengths and weaknesses, to identify which ones produce better results when combined, and to provide improvement suggestions.

Vision for future research: researchers usually develop their own UX evaluation instruments and validates them, for instance, through confirmatory factor analysis when scale-based. However, it is not very common to evaluate them by gathering feedback from users through interviews and open-ended questions. We encourage researchers to get this feedback and perform qualitative analysis to get a deeper understanding about users' perceptions on the method employed, such as: i) the ease for being used and understood; ii) the capacity of allowing users to fully convey their UX; iii) its adequacy to capture all aspects that influenced users' perceptions about the UX; iv) suggestions of improvements from the users' point of view. We also encourage the HCI community to carry out more comparative studies, which the results would make it possible to suggest improvements and to develop new methods, as well as help researchers to decide which method(s) to choose.

Many difficulties and negative emotions, but positive UX

Different studies [6, 13, 16] have been shown diverging results, where users that demonstrated many negative emotions or faced many difficulties during his/her interaction with a product still considered their UX as positive. Such phenomenon may indicate, for instance, that other factors possibly had a stronger influence in their perception about the UX rather than those evaluated by the methods employed. Some studies [8, 18] have indicated that users tend to super estimate their experiences when evaluating retrospectively. This phenomenon is known as memory-experience gap, which is defined as the "discrepancy between the average of experienced emotions and the overall evaluation of the experience, which is usually more intense than averaged emotions" [15]. Although it is argued that this phenomenon tends to make negative sentiments more prominent in retrospective evaluations [15], studies have been pointing out that users might evaluate positively the UX of a product even when they face many usability problems [13, 16] and many negative sentiments are awakened throughout the interaction [6]. Indeed, a study [3] revealed that the memory-experience gap is more prominent when users interact with an application that has usability problems, leading them to even give higher ratings than users that interacted with a usability free application. There was also evidence that users' preferences and perceptions are affected by the polarity of the most intense event added to the final interaction moment [5], a phenomenon known as peak-end effect. Such research raises indications that the results from the current practices and methods to evaluate the UX might have been influenced by several factors that are not being considered, leading to biased results.

Vision for future research: these issues raise the need to foster discussions on current practices in UX evaluation. Indeed, it is not possible, nor viable, to treat all the variables that influence users' perceptions about their experiences. However, it is important that researchers consider them when developing new methods or carrying out evaluations, given that the results might be being biased. By

ACKNOWLEDGMENTS
We would like to thank the financial support granted by FAPEAM through Edital 009/2017, CNPq through processes 423149/2016-4 and 311494/2017-0, and CAPES through process number 175956/2013.

identifying these variables, it would be possible, for instance, to develop approaches and methods that are less prone to the effect of these variables and, consequently, carry out more precise UX evaluations. Researchers may study, for example, the influence of culture and context in the evaluations in order to identify which aspects have a stronger influence on UX. By doing this, researchers may develop personalized instruments according to the target population. A work in line with this approach was carried out by Burnett et al. [4], who adapted personas into a Cognitive Walkthrough process to evaluate problem-solving software from a gender-inclusiveness perspective. Another possibility is to employ machine learning techniques, for instance, to perform data mining on user reviews from app stores in order to identify which attributes have a stronger correlation with higher and lower ratings.

CONCLUSION AND FUTURE WORK

This paper aimed to shed light on and bring reflections on current practices in UX evaluations. Although much research has been carried out, the challenges and gaps we presented highlights that there is still a long road ahead. It is important that the HCI community reflects on these issues, keeping the efforts to provide more comprehensive theories about UX, to identify which aspects influence on users' perceptions about their experience and to review the current practices in UX evaluations. By doing this, it would be possible to better comprehend about UX and develop better evaluation approaches. We are currently using machine learning to perform sentiment analysis and to classify user reviews from mobile app stores according to the dimensions of UX to identify which have a stronger influence on UX ratings. We will also carry out empirical studies to verify which dimensions and attributes of UX are affected by the peak-end effect when ordering the tasks in different sequences according to their level of complexity and usability. The outcomes may help practitioners to design focusing on aspects that really impacts on UX, as well as to provide insights for researchers to design methods that reduce the influence of variables that affect UX evaluations, obtaining less biased results.

REFERENCES

[1] ISO 9241-210. 2010. *Ergonomics of Human-system Interaction: Part 210: Human-centred Design for Interactive Systems.*
[2] Javier A Bargas-Avila and Kasper Hornb. 2011. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 10.
[3] Anders Bruun and Simon Ahm. 2015. Mind the Gap! Comparing Retrospective and Concurrent Ratings of Emotion in User Experience Evaluation. In *Human-Computer Interaction – INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Vol. 9296. Springer International Publishing, Cham, 237–254. https://doi.org/10.1007/978-3-319-22701-6_17
[4] Margaret Burnett, Simone Stumpf, Jamie Macbeth, Stephann Makri, Laura Beckwith, Irwin Kwan, Anicia Peters, and William Jernigan. 2016. GenderMag: A Method for Evaluating Software's Gender Inclusiveness. *Interacting with Computers* 28, 6 (Nov. 2016), 760–787. https://doi.org/10.1093/iwc/iwv046
[5] Andy Cockburn, Philip Quinn, and Carl Gutwin. 2015. Examining the Peak-End Effects of Subjective Experience. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul,

Republic of Korea, 357–366. https://doi.org/10.1145/2702123.2702139

[6] Giulia de Andrade Cardieri and Luciana Martinez Zaina. 2018. Analyzing User Experience in Mobile Web, Native and Progressive Web Applications: A User and HCI Specialist Perspectives. In *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems - IHC 2018*. ACM Press, Belém, Brazil, 1–11. https://doi.org/10.1145/3274192.3274201

[7] Marc Hassenzahl. 2018. The Thing and I (Summer of '17 Remix). In *Funology 2*, Mark Blythe and Andrew Monk (Eds.). Springer International Publishing, Cham, 17–31. https://doi.org/10.1007/978-3-319-68213-6_2

[8] Sari Kujala and Talya Miron-Shatz. 2015. The evolving role of expectations in long-term user experience. In *Proceedings of the 19th International Academic Mindtrek Conference on - AcademicMindTrek '15*. ACM Press, Tampere, Finland, 167–174. https://doi.org/10.1145/2818187.2818271

[9] Carine Lallemand, Guillaume Gronier, and Vincent Koenig. 2015. User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior* 43 (Feb. 2015), 35–48. https://doi.org/10.1016/j.chb.2014.10.048

[10] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold POS Vermeeren, and Joke Kort. 2009. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 719–728.

[11] Effie Lai-Chong Law, Arnold POS Vermeeren, Marc Hassenzahl, and Mark Blythe. 2007. Towards a UX manifesto. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2*. BCS Learning & Development Ltd., 205–206.

[12] Craig M. MacDonald and Michael E. Atwood. 2014. What does it mean for a system to be useful?: an exploratory study of usefulness. In *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*. ACM Press, Vancouver, BC, Canada, 885–894. https://doi.org/10.1145/2598510.2598600

[13] Leonardo Marques, Walter Nakamura, Natasha Valentim, Luis Rivero, and Tayana Conte. 2018. Do Scale Type Techniques Identify Problems that Affect User eXperience? User Experience Evaluation of a Mobile Application (S). In *The 30th International Conference on Software Engineering and Knowledge Engineering*. 451–501. https://doi.org/10.18293/SEKE2018-161

[14] Oussama Metatla, Nuno N. Correia, Fiore Martin, Nick Bryan-Kinns, and Tony Stockman. 2016. Tap the ShapeTones: Exploring the Effects of Crossmodal Congruence in an Audio-Visual Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 1055–1066. https://doi.org/10.1145/2858036.2858456

[15] Talya Miron-Shatz, Arthur Stone, and Daniel Kahneman. 2009. Memories of yesterday's emotions: Does the valence of experience affect the memory-experience gap? *Emotion* 9, 6 (2009), 885–891. https://doi.org/10.1037/a0017823

[16] Walter Nakamura, Leonardo Marques, Luis Rivero, Elaine Oliveira, and Tayana Conte. 2017. Are Generic UX Evaluation Techniques Enough? A study on the UX Evaluation of the Edmodo Learning Management System. In *Brazilian Symposium on Computers in Education*. Recife, Pernambuco, Brasil, 1007. https://doi.org/10.5753/cbie.sbie.2017.1007

[17] Ingrid Pettersson, Florian Lachner, Anna-Katharina Frison, Andreas Riener, and Andreas Butz. 2018. A Bermuda Triangle?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–16. https://doi.org/10.1145/3173574.3174035

[18] Samaneh Soleimani and Effie Lai-Chong Law. 2017. What Can Self-Reports and Acoustic Data Analyses on Emotions Tell Us?. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*. ACM Press, Edinburgh, United Kingdom, 489–501. https://doi.org/10.1145/3064663.3064770

[19] Arnold POS Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. ACM, 521–530.