

“How Could an Intranet be Like a Friend to Me?” - Why Standardized UX Scales Don’t Always Fit

Carine Lallemand
University of Luxembourg
COSA Institute
Esch-sur-Alzette, Luxembourg
carine.lallemand@uni.lu

Vincent Koenig
University of Luxembourg
COSA Institute
Esch-sur-Alzette, Luxembourg
vincent.koenig@uni.lu

ABSTRACT

*“I hope that this survey is a joke because it made me laugh so much”. This quote is just one example of many negative respondents’ reactions gathered during a large-scale user experience (UX) study. Unfortunately, the survey was no joke, rather a well-constructed and validated standardized UX scale. This paper critically reflects on the use and relevance of standardized UX scales for the evaluation of UX in business contexts. We report on a real-world use case where the meCUE questionnaire has been used to assess employees’ experience (N=263) with their organization’s intranet. Strong users’ reactions to the survey’s items and statistical analyses both suggest that the scale is unsuitable for the evaluation of business-oriented systems. Drawing on the description of this inadequacy, we discuss the quality of academic UX tools, calling into question the relevance for practice of academic methods.

CCS CONCEPTS

• **Human-centered computing** → **Human Computer Interaction (HCI)** → HCI design and evaluation methods, User studies

KEYWORDS

Standardized UX scales; meCUE questionnaire; workplace UX assessment, intranet.

ACM Reference format:

C. Lallemand and V. Koenig. 2017. “How Could an Intranet be Like a Friend to Me?” - Why Standardized UX Scales Don’t Always Fit. In *Proceedings of ACM ECCE conference, Umeå, Sweden, September 2017 (ECCE’17)*, 8 pages.

<https://doi.org/10.1145/3121283.3121288>

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 the author(s) 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. ECCE 2017, September 19–22, 2017, Umeå, Sweden
© 2017 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-5256-7/17/09...\$15.00
<https://doi.org/10.1145/3121283.3121288>

1 INTRODUCTION

As research scientists, we regularly advocate for more rigor in UX evaluation processes and sometimes look at practitioners’ ad-hoc evaluation questionnaires with a suspicious eye. Without any theoretical model and any validation process applied to their items, how good can such measures be? Why don’t UX practitioners use a standardized UX scale rather than a quick and dirty home-grown instrument?

Surveys on UX practice indeed show that practitioners favor the use of low-cost in-house evaluation methods [1,8]. And this makes perfect sense for UX practitioners, who must cope with many constraints in order to design great products or services with limited resources. 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”. Frequently, the development and use of methods whose validity is doubtful might be explained by a focus on practicability over scientificity [25]. This is not surprising, yet the case of standardized and holistic UX questionnaires remains questionable since they claim to conciliate both scientificity and practicability but still have difficulties reaching a wide practitioners’ audience. At the time where academics attempt more and more to transfer standardized UX questionnaires to practitioners (e.g., the trend of having a dedicated wide audience website presenting the scale along with an easy to use manual or even an integrated administration platform), this requires critical reflection.

In this paper, we report on a real-world use case where we used the meCUE questionnaire [16] to assess employees’ experience with their organization’s intranet. Interestingly, it turned out that the participants reported as much criticism on the survey than on the system under study. Based on this instructive experience as researchers, this paper is an ideal opportunity to critically reflect on the usefulness and relevance of standardized UX scales for the evaluation of UX in business contexts.

2 HOLISTIC UX EVALUATION QUESTIONNAIRES

Standardized UX questionnaires are self-administered UX evaluation tools [25]. A questionnaire is said to be standardized when each respondent is exposed to the same questions in the same order and the same system of coding responses. A well-designed questionnaire offers the advantage of providing a valid and reliable assessment of a construct [17].

As compared to numerous standardized usability questionnaires, well spread amongst practitioners (e.g. the famous System Usability Scale by [4] or the Software Usability Measurement Inventory by [10]; see [19] for a review of existing scales), academia has only produced a handful of standardized generic UX questionnaires so far.

Main examples of holistic UX evaluation scales encompass: the AttrakDiff 2 scale [9], which provides a global subjective evaluation of both pragmatic and hedonic qualities of a system based on 28 semantic differentials; the relatively similar User Experience Questionnaire (UEQ) [14] relying on the same format but claiming a more balanced stance between pragmatic and hedonic dimensions; and finally the meCUE scale [16], which provides a self-reported assessment of product perceptions, users’ emotions, consequences of use and an overall evaluation. For the purpose of this paper, we will describe the latter in more detail.

2.1 meCUE: a Modular Evaluation of Key Components of UX

The meCUE questionnaire [16] is described by its authors as “a freely available, scientifically-founded questionnaire, which focuses on the modular acquisition of user-centered reviews and their experience of interactive technical products” (source: <http://mecue.de/english/home>). It is a modular UX assessment scale based on Thüring and Mahlke’s CUE-model [18] and composed of 34 items divided into 4 dimensions (Fig. 1): *instrumental* and *non-instrumental product perceptions*, *emotions*, *consequences* and *overall judgment*.

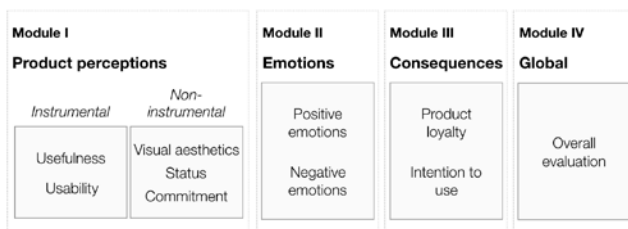


Figure 1: Structure of the UX questionnaire meCUE

Within each subscale, respondents are asked to assess their agreement level with statements on a 7-points Likert scale from 1 “strongly disagree” to 7 “strongly agree”. All items are positively worded and mandatory, as stated in the instructions: “Sometimes you may not be completely sure about your agreement with a particular statement or you may find the statement does not apply completely to the particular product. Nevertheless, please tick a circle in every line.”

According to its authors, the main advantage of meCUE as compared to existing questionnaires is to assess the major components of UX in a comprehensive manner [16]. The psychometric properties of the questionnaire have been assessed through several studies [11,15,16]. In addition, the authors claim that “as demonstrated in (their) studies, meCUE can be applied in UX surveys on all kinds of interactive systems”.

3 METHOD

3.1 Context

In the context of an internal system redesign, we used the meCUE questionnaire [16] to assess employees’ experience with their organization’s intranet. The request was made by the organization’s communications department, which was not satisfied with the intranet’s usage rate and asked us to investigate the reasons why the tool was not as successful as expected. We mainly intended to use the meCUE questionnaire as a way to understand employees’ experiences on several UX dimensions. Even though we expected the perceived quality of the current intranet to be rather low, we selected a UX scale to establish a baseline and allow for a before-after comparison score on both pragmatic and hedonic dimensions of the interaction, in accordance with the stakeholders’ request. Improving both hedonic and identity aspects has actually been an explicit objective of the organization; and this can be increasingly observed in intranet design nowadays, aka. the trend for “social intranets”. The meCUE questionnaire has documented scientific properties (cf. section 2.1), specifically allows for the assessment of both hedonic and pragmatic aspects and is described to be suitable for all contexts. It thus appeared as a perfect candidate to both assess the UX and gain methodological insight in how well such a scale really suits a business-Intranet context. The more “common-sensual” choice of a pragmatic-only scale (aka. usability scale) would thus not have been a good candidate for our purposes.

3.2 Pilot Test

Before deploying the questionnaire, we pilot tested the scale on 35 participants during face-to-face individual sessions in order to assess the understandability of the survey’s format and statements. The pilot test was done as part of a thorough translation and validation process and was therefore not focused on the intranet. At this stage, we identified some issues (e.g., perceived redundancy, inappropriate item formulation) based on users’ detailed feedback. As users’ comments were not related to translation issues, we decided to keep all items and to administer the scale in its original form in order to safeguard – and eventually check - its psychometric properties.

3.3 Participants

We later deployed the scale for our real-world use case for the redesign of an intranet. The communications department sent an email invitation to all employees (nearly two thousand), kindly asking them to complete our survey and evaluate their experience

with their organization’s intranet. The study was also broadcast on the intranet itself and advertised on flyers.

255 valid answers were collected. The mean age in our sample (based on 187 answers) was 37 (Min = 24, Max = 64). Due to the sensitive nature of the survey, and in order for the participants to feel comfortable sharing their experiences with the intranet, we decided to define demographic questions as optional in the survey. Respondents had the choice to answer the survey either in German (11%), English (62%) or French (27%).

3.4 Material

The survey consisted of four parts: (1) a screener asking employees whether they use the intranet or not and at what frequency, (2) the evaluation of the intranet using the meCUE questionnaire (each module being presented on a separate screen), (3) a qualitative and spontaneous UX assessment using the sentence completion method, (4) demographics. Finally, we added an open text field at the end of the survey: “Do you have other comments or suggestions about the intranet?” People who declared having never used the intranet were redirected to a shorter survey only investigating their needs and expectations.

4 RESULTS

Only the results related to the meCUE questionnaire and underlying users’ comments will be presented and discussed in this paper. We will first describe relevant observations made during the pilot test; we will then present how our participants assessed the intranet on the meCUE subscales, before placing an emphasis on the comments and criticism they addressed to the survey itself. While the pilot test and users’ comments directly inform us about the questionnaire’s adequacy, results related to the intranet might serve as a support to understand potential relations between the system evaluation and the meCUE questionnaire.

4.1 Pilot Test of meCUE on Several Application Contexts

During the pilot test, we asked participants to evaluate an interactive product that they used daily. We also asked them to think aloud while completing the questionnaire and to assess the level of understandability of each item on a 5-points Likert scale. From a strict language perspective, all items were assessed as understandable (average ratings ranging from 4 to 5), meaning that the items use everyday language. However, the majority of our respondents commented on several items as being “weird”, “exaggerated”, “ridiculous”, “out of scope”. Problematic items mainly include items from the *Commitment* and *Status* dimensions, the *Emotions* and the *Consequences* modules, as illustrated by these users’ comments:

- (Item C2) “That’s weird! it’s too strong and ridiculous (laughs). I empathize with those who will say that a product is like a friend to them.” (#26)
- (Item C3) “I could be annoyed, but, come on, “devastated”. No, no. I would be devastated if I lost a beloved one not a system or product” (#20)

- (Item PA2) “Makes me feel euphoric?? This is not a drug. Such items are really exaggerated!” (#25)

Interestingly, we also observed that the statements that participants considered as “exaggerated”, “ridiculous”, or “not applicable to the context of their experience” were either rated as a strong disagreement (1 “strongly disagree”) or as neutral (4 “neither agree nor disagree”). Pilot test participant #35 illustrates this dilemma by stating: “Whenever there is a super weird statement you really don’t know if you must tick the option “strongly disagree” because the statement has nothing to do with your experience or interaction, or the middle option “neither agree nor disagree” because you don’t want to bias the results with an answer that should obviously have been “not applicable”. Many others regretted and questioned the absence of a “not applicable” option.

Another major criticism addressed to the questionnaire was the perceived redundancy of items. The wide majority of our respondents were annoyed by the fact that several items were “exactly the same” (#20) and some even “were 100% sure that they already answered the same question a minute ago” (#25). Emotional nuances were particularly hard to distinguish.

Finally, some items were described as “ambiguous” (“What does it mean for a system to be creatively designed?” #26 about item A1) or “badly worded”, such as item IN_1 “If I could, I would use the product daily” where people said in almost every case “the question is irrelevant because I can use it daily” (#34)

4.2 Modular UX evaluation of the intranet

As shown in [Table 1](#), the intranet was negatively assessed on all meCUE modules, with average ratings ranging from 2.55 (product loyalty) to 3.65 (usability) maximum. These scores are therefore always below the neutral agreement level (which is 4 in a 7-points Likert scale). As the meCUE is positively worded (except for 6 items related to negative emotions), this means that on average our participants disagree or strongly disagree with all statements of the scale, except for the negative emotions dimension (M=4, SD=1.5). On module IV *Overall Judgment*, the intranet produced a mean score in the negative range (M=-0.52, SD=3), which confirms that the system does not address users’ needs in a satisfactory way.

Table 1: Descriptive statistics

meCUE modules	Mean	SD
Module I		
Usefulness	3.52	1.37
Usability	3.65	1.61
Aesthetics	3.55	1.46
Status	2.58	1.23
Commitment	2.36	1.38
Module II		
Positive emotions	2.46	1.15
Negative emotions	4	1.5
Module III		
Intention to use	2.78	1.23

Product loyalty	2.55	1.29
Module IV		
Overall evaluation	- 0.52	3

These results were not surprising as negative informal feedback about the intranet and a low usage rate observed through the platform analytics were the main reasons for launching a redesign of the system.

For the purposes of this paper, however, findings about the questionnaire’s adequacy are more interesting. The frequency distribution of responses for each item provides us with a more accurate view of the data and tends to confirm observations reported during the pilot test. Whereas non-problematic items show a relatively balanced frequency distribution, items identified as weak during the pilot test show unbalanced distributions with peak frequencies for answer option 1 “strongly disagree” and 4 “neither agree nor disagree” (see examples on Fig. 2).



Figure 2: Frequency distribution of responses for items U1, F1, C2 and C3

In addition to descriptive statistics, we also conducted a confirmatory factor analysis to verify the factorial structure. The analysis did not entirely confirm previous work and we were therefore not able to replicate the theoretical structure for modules I and III (e.g., usability and usefulness items tend to load on a single dimension; numerous problematic items related to status and commitment subscales). Due to space constraints, we are not able to report the analysis here, and this will be the subject of a future publication.

4.3 Users’ comments on the meCUE questionnaire

While the open-ended question at the end of the survey aimed at providing participants with a free space to express their view on the intranet, 16 participants (6.3% of valid answers) used it to comment on the meCUE questionnaire per se. This might seem hardly relevant when compared to the sample size; however, one should note that only those who answered all parts of the questionnaire had the opportunity to answer this question. The team also received some emails and oral comments on the survey.

4.3.1 Is this a joke? One of the most spontaneous reactions to the survey was to ask whether it was a joke:

- “Sorry, I don't get this survey! This seems like a joke! You are making fun of us, right?” (#161)
- I hope that this survey is a joke because it made me laugh so much... (#94)
- “I laughed so much while reading the questions that my daughters also wanted to “play”. Thank you.” (#166)
- “I don't have an additional comment on the intranet, but I DO have one on this survey. I suppose the idea behind these questions is a joke. If not, I'm quite disappointed. In my score, you will have many "strongly disagrees" because these questions don't make any real sense for me” (#82)
- “...the questions about "evaluation": where you serious??? they did not look very serious to me.” (#104)
- “WTF?! Who came up with the questions for this survey?!!” (#161)
- “I find the first part of this survey quite unprofessional. In the beginning, I thought it was a joke.” (#246)

4.3.2 The questions don't make any sense. On the same line, respondents criticised the statements for “being silly”, “exaggerated” and “not making any sense”, criticising us to “turn the matter into ridiculous” and for wasting the time of employees:

- “This is a silly survey, with questions that are relatively useless to anyone. Who talks about an intranet in terms of euphoria, cheerfulness, and exhaustion! No one I know, because the intranet is a somewhat useful tool to check my pay stubs and vacation hours. Other than that, I have no use for it at all.” (#231)
- “What have you smoked when you developed this questionnaire’s statements? The statements are completely exaggerated. With that, you turned your matter into ridiculous.” (#238)
- “This question makes no sense (like the whole first part of this questionnaire; why can't ONE person take the time to delete all standardized nonsense question instead of wasting the time of ALL participants who have to answer them?)” (#102)
- “Not about the intranet but about this survey and the questions that have been asked in it! I don't really see the point in asking bullshit questions like "the intranet is like a friend to me" or "I would be devastated if the Intranet didn't exist" or "it enhances my standing" or "my friends will be envious"... after all, the intranet is neither a status symbol to show off with nor a human being with whom one can have a relationship! In my opinion, the intranet should be a clearly structured, easy-to-use TOOL for the staff to help them deal with admin... so please stop doing silly surveys like this one and just let us get on with our work!” (#216)

4.3.3 *This survey won't help you to improve the system.* Respondents were also dissatisfied with the impression that the questionnaire was a useless basis to guide any suggestion for improvement:

- “Who the hell did select these questions???” “I feel euphoric when using the intranet???” “Wtf! I really don't see how these questions will help you improve the intranet” (#79)
- “Who designed this questionnaire? How will it be possible to improve the current intranet with such a questionnaire? Is this a joke? How many hours and EUROS of employees are wasted here? The questionnaire needs to be much more targeted.” (#83)

4.3.4. *The questions are redundant.* Perceived redundancy of the items, already identified as an annoyance factor during the pilot test, was also highlighted as an issue in the users' comments:

- “the questions at the beginning of the survey are very redundant. I've never seen such a questionnaire!” (#31)
- “I did not like the leading and partially redundant questions. Such a waste of time...” (#246)

As can be seen from the respondents' comments, there has been at least a sub-sample of respondents who had a rather negative experience of answering the meCUE questionnaire. Based on our previous work on standardized UX questionnaires [11,12], we argue that this is not an isolated phenomenon and that the issue of UX scale inadequacy has to be addressed more systematically by the community.

5 DISCUSSION

The results described above illustrate the following issues when deploying the meCUE questionnaire for the purposes of our use case: (1) qualitative data (critical feedback) gathered during pilot testing; (2) quantitative and qualitative data gathered during the main data collection (inadequately formulated items, perceived item redundancy, unusual response distributions, strong general criticism expressed by participants, factorial structure deviating from originally documented structure). Taken together, these issues explain why the validity of some answers might be compromised and why the participants' trust and willingness to take the data collection serious cannot be taken for granted. One might simplistically deduce that the meCUE is not suitable for such business-oriented use cases. Similarly, one might think that such a scale could be inappropriate for a system perceived as negative on all UX dimensions. However, our findings suggest more complex answers to this problem, rather than reducing it to the only context of application or specific system under evaluation.

From a respondent's perspective, the item formulation seems most impacting. As generic UX questionnaires promote a “one size fits all approach”, it seems obvious that the wording of their items should tend to be applicable to a majority of contexts. To do so, not only should the items be checked for face validity, but their understandability and applicability should also be extensively tested through qualitative pilot testing. Beyond numerous items that were perceived inappropriate or even

exaggerated to assess an interactive system, we observed that other meCUE statements, while being reasonable in terms of meaning, are debatable in terms of their formulation (e.g., item A1). In Minge and al.'s recent paper [15] on the development and validation of an English version of the meCUE, the authors go from the back-translation process to the deployment of the scale in an online survey (p. 2058). They rely on this process to demonstrate that the meCUE is a validated questionnaire having a good internal consistency and reliability. However, as highlighted by work on scale construction and translation [22,23,24], both scale construction and scale translation processes require pilot studies in which each participant is encouraged to think aloud in order to indicate how he or she interprets the stimuli and to motivate responses. Thanks to this process, “poorly formulated questions or other problems with the instrument may be quickly revealed” [24] (p. 129). This claim is indeed backed up by our findings, illustrating what potential problems can be flagged during pilot-testing.

The same critical perception applies to the redundancy in items, especially emotional items (yet not only). Our observations showed that people were not able to clearly distinguish between emotional nuances or facets of usability and therefore complained about answering the same question at least twice. In the construction of a questionnaire, multiple-items scales aim at covering different facets of a construct in order to safeguard construct validity [17]. However, when items are so close that target respondents describe them as “synonyms”, the objective is not reached. This leads to perceived redundancy, an increase in the length of the questionnaire and also an “artificial” increase in the internal consistency of a scale.

5.1 What about other standardized UX questionnaires?

The issues highlighted here are not limited to the meCUE questionnaire. Regarding the issue of applicability of UX tools in industrial settings, and especially for the evaluation of business-oriented systems, one should note for instance that the User Experience Questionnaire has been designed as a response to the “limitations” of the Attrakdiff scale [9]. According to Laugwitz et al. [14], Attrakdiff “lays a greater emphasis on the hedonic aspects of product quality than on the pragmatic aspects. This may not be perfectly appropriate for a comprehensive evaluation of professional software” (p. 65). Moreover, in a majority of cases, holistic UX scales validation studies are based on the assessment of leisure-based or personal interactive products. A common paradigm is to let participants free to choose which product they want to rate [15]. While being more convenient from the researcher's perspective, this approach might lead to bias and should be complemented by additional studies if the objective is to safeguard that a questionnaire can be applied in UX surveys on all kinds of interactive systems.

Discussions about the scientific rigor in the development and validation of a questionnaire are obviously not solely focused on the meCUE. While the three main holistic UX scales (AttrakDiff, meCUE, UEQ) are considered by the research community as

validated tools, a previous study showed that the initial validation of the Attrakdiff scale and subsequent English translation can also raise issues and needed more consolidation from the UX community [12,13]. Too often, standardized scales are considered validated after a single validation study where their authors conclude that their questionnaire has good psychometrics properties and can therefore be considered as valid. The issue here is not really that the instruments we develop have some flaws and that the validation studies that we conduct have some limitations; they of course do. The issue is much more related to the general lack of consolidation (e.g., replication) and the pressure for publishing innovative and fully-fledged scales. Developing a new measurement scale should be done in a lean way, relying on iterative community efforts. Of course, many reasons might explain these practices, including the pressure imposed by the “publish or perish” model, the interdisciplinary nature of the Human-Computer Interaction field and associated differences in quality standards or even the limitation of page number imposed by publication formats. In any case, this observation confirms the need for more methodological consolidation and the development of better quality and publication standards to document the development of novel tools.

5.2 Why standardized UX scales don’t always fit?

In their quest for maximum transferability, holistic UX questionnaires tend to follow a “one size fits all” trend. However, is a generic measurement of UX realistic? 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 [2]. 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 [2], Web 2.0 services [21], virtual environments [6], or games [3], this seems insufficient with regards to the industrial needs. The creation of more specific tools, however, is a promising lead towards more adequately formulated sets of items used to assess user experience for given types of systems. This approach would also support the refinement of theories underlying UX; the three standardized scales mentioned in this paper all rely on a similar and generic UX model that might be challenged and refined.

The negative users’ comments collected in our study also bring additional questions to the table. For instance, it looks like UX scales were designed to assess systems that already have a good UX baseline. But how is one supposed to really know if a system already is any good or not? And how should we cope with this inadequacy when one’s objective is to establish a baseline and allow for a before-after comparison score on both pragmatic and hedonic dimensions of the interaction? UX scales are easily and widely used by non-expert people, for the good reason that their application is (perhaps naively) perceived as requiring almost no expertise. This brings us to a second concern related to standardized UX scales: their practicability and the way we

transfer these tools to practice. Authors should be more cautious when advertising their scales by clearly highlighting the limitations of each study and tool without overestimating the power and applicability of a scale. Several validation studies have been conducted to assess the psychometrics properties of the meCUE scale [11,15,16]. These studies have led its authors to claim that “as demonstrated in (their) studies, meCUE can be applied in UX surveys on all kinds of interactive systems”. In the paper reporting on the validation of the English version of the scale, one can actually not find any mention to a limitation related to the process or the tool and no single item is described as problematic. However, the present use case shows that the applicability of the meCUE might be debatable for business-oriented systems and that its overall quality might be improved on several dimensions.

6 CONCLUSION

Unsuccessful cases are not usually reported in Academia, yet we believe they often allow to uncover issues and ultimately to serve as a basis for improving our tools and methods. Through this paper, we intended to share the lessons we learned by applying a standardized UX scale on a business-oriented use case. The generalizability of the results seems of course questionable as we only report a single use case focused on one UX scale. However, our goal here was to leverage one illustrative use case in order to extend the debate to other scales, as it appears that some of the issues we emphasize are quite general.

The use of standardized UX questionnaires seems trivial enough to fall into many traps. This is true for scientists, and probably even more for practitioners, who are not necessarily trained to understand all requirements for the psychometrically accurate application of the scale and interpretation of data. Interestingly, while we tended to believe that the weaknesses of standardized scales had only a medium impact in practice; and that using an academic tool was better than using an ad hoc questionnaire not based on any model and without any concern for validity and reliability, this paper attempts to raise awareness on a misconception that we initially shared.

To conclude this paper, we can wonder: what if current standardized UX scales sometimes just don’t fit real-life cases? What if their cost-effectiveness and relevance to practice were too low at the moment? As a community, we easily acknowledge that there are things we are doing wrong and could do differently. Yet more and more of our contributions - even the most inspiring ones - sometimes omit key quality components. As stated by Dix [7], “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 such as the questionnaires mentioned in this paper 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 [5]. It is also our duty to better inform practitioners of the limitations of a tool and also of

course, as a community, to keep working iteratively on their continual improvement.

A APPENDIX: meCUE QUESTIONNAIRE

Appendix A presents the French and English versions of the meCUE questionnaire used in our multilingual survey in addition to the original German version published by Minge & Riedel [16]. We translated the questionnaire into French following a rigorous translation process. The English version is a slightly adapted translation based on Minge, Thüring & Wagner [15]. Note that the items were contextualized to the purpose of our study, by replacing the generic term “product” by the term “intranet” – the table below shows the non-contextualized items.

Module	Dimensions	meCUE English items	meCUE French items
Module I Product Perceptions	Usefulness	F1 The functions of the product are exactly right for my goals.	Les fonctionnalités du produit sont parfaitement adaptées à mes objectifs.
		F2 I consider the product extremely useful.	Je considère ce produit comme extrêmement utile.
		F3 With the help of this product I can achieve my goals.	A l'aide de ce produit, je peux atteindre mes objectifs.
	Usability	U1 The product is easy to use.	Le produit est facile à utiliser.
		U2 It is quickly apparent how to use the product.	On perçoit rapidement comment utiliser le produit.
		U3 The use of the product is easy to understand.	L'utilisation du produit est facile à comprendre.
	Visual aesthetics	A1 The product is creatively designed.	Le produit est conçu de manière créative.
		A2 The design looks attractive.	Le design a l'air attrayant.
		A3 The product is stylish.	Le produit est élégant.
	Status	S1 The product enhances my standing among peers.	Le produit me donne une meilleure image auprès des autres.
		S2 With this product, I am perceived differently.	A travers ce produit, on me perçoit différemment.
		S3 My friends may well be envious of this product.	Mes amis peuvent bien être envieux de ce produit.
	Commitment	C1 I cannot live without this product.	Je ne peux pas vivre sans ce produit.
		C2 The product is like a friend to me.	Le produit est comme un ami pour moi.
		C3 If I lost the product, I would be devastated.	Si je perdais le produit, j'en serais dévasté.
Module II Emotions	Positive Emotions	PA1 The product excites me.	Le produit m'enthousiasme.
		PA2 The product makes me feel euphoric.	Le produit me rend euphorique.
		PA3 When using this product, I feel cheerful.	Quand j'utilise ce produit, je me sens joyeux.
		PD1 The product relaxes me.	Le produit me détend.
		PD2 When using this product, I feel at peace.	Quand j'utilise ce produit, je me sens serein.
		PD3 The product calms me.	Le produit m'apaise.
	Negative Emotions	NA1 The product annoys me.	Le produit m'énerve.
		NA2 The product frustrates me.	Le produit me frustre.
		NA3 The product makes me angry.	Le produit me met en colère.
		ND1 The product makes me tired.	Le produit me fatigue.
		ND2 When using this product, I feel exhausted.	Quand j'utilise le produit, je me sens épuisé.
		ND3 When using this product, I feel passive.	Quand j'utilise le produit, je me sens passif.
Module III Consequences	Product Loyalty	L1 I would not swap this product for any other.	Je n'échangerais le produit contre aucun autre.
		L2 In comparison to this product, others look incomplete.	Par rapport à ce produit, les autres produits ont l'air moins perfectionnés.
		L3 I would get exactly this product (again) everytime.	Je n'hésiterais pas à choisir exactement ce produit (à nouveau).
	Intention to use	IN1 If I could, I would use the product daily.	Si je le pouvais, j'utiliserais le produit chaque jour.
		IN2 I can hardly wait to use the product again.	Je suis impatient d'utiliser le produit à nouveau.
		IN3 When using the product, it happens that I lose track of time.	Quand j'utilise ce produit, il m'arrive de perdre la notion du temps.
Module IV	Global	Finally, how would you rate the product overall?	Enfin, comment évaluez-vous ce produit dans son ensemble ?

ACKNOWLEDGMENTS

We would like to thank Sophie Doublet and Marie Behaegel for their contributions to this work. We also express our gratitude to Martin Schrepp, Sari Kujala, Aliaksei Miniukovich and Noam Tractinsky for their valuable comments on a previous version of this work.

REFERENCES

[1] R. Alves, P. Valente, and N. J. Nunes. 2014. The state of user experience

- evaluation practice. In *Proc. of NordiCHI'14*. New York, NY: ACM Press (pp. 93-102). doi:10.1145/2639189.2641208
- [2] Regina Bernhaupt and Michael Pirker. 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, 642-659. Berlin/Heidelberg: Springer.
- [3] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny. 2009. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45, 624-634.
- [4] John Brooke. 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.

- [5] Paul Cairns. 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.
- [6] D. B. Chertoff, B. Goldiez, and J. J. LaViola. 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.
- [7] Alan Dix. 2010. Human-Computer Interaction: a stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers*, 22(1), 13-27.
- [8] C. M. Gray, E. Stolterman and M. A. Siegel. 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
- [9] M. Hassenzahl, M. Burmester, and F. Koller. 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
- [10] Jurek Kirakowski and Mary Corbett. 1993. SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24, 210–212.
- [11] Carina Kuhr. 2013. Measuring the User Experience of Mobile Applications - an Empirical Validation of a Quantitative Method. (*Master thesis*). Technische Universität Berlin. http://mecu.de/Homepage%20Content/02%20Links%20%26%20Literatur/M_A_Kuhr.pdf
- [12] Carine Lallemand. 2015. Towards Consolidated Methods for the Design and Evaluation of User Experience. (Doctoral dissertation). University of Luxembourg. <https://publications.uni.lu/handle/10993/21463>
- [13] C. Lallemand, V. Koenig, G. Gronier, and R. Martin. 2015. Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. *Revue européenne de psychologie appliquée*, 65(5), 239-252. doi:10.1016/j.erap.2015.08.002
- [14] B. Laugwitz, T. Held, and M. Schrepp. 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.
- [15] M. Minge, M. Thüring, and I. Wagner. 2016. Developing and Validating an English Version of the meCUE Questionnaire for Measuring User Experience. In Proc. of the HFES'16. doi: <http://dx.doi.org/10.1177/1541931213601468>
- [16] Michael Minge and Laura Riedel. 2013. meCUE – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In: S. Boll, S. Maaß & R. Malaka (Hrsg.): *Mensch und Computer 2013: Interaktive Vielfalt* (S. 89-98). München, Oldenbourg Verlag.
- [17] Paul E. Spector. 1992. Summated rating scale construction. An Introduction. Newbury Park: Sage.
- [18] Manfred Thüring and Sascha Mahlke. 2007. Usability, aesthetics and emotions in human-technology interaction. *International Journal of Psychology*, 42(4), 253–264. doi:10.1080/00207590701396674
- [19] Tom Tullis and William Albert. 2013. *Measuring the User Experience* (2nd ed.): Collecting, Analyzing, and Presenting Usability Metrics. Burlington, MA: Morgan Kaufmann Publishers.
- [20] K. Väänänen-Vainio-Mattila, V. Roto, and M. Hassenzahl. 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.
- [21] Kaisa Väänänen-Vainio-Mattila and Minna Wäljas. 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
- [22] Robert J. Vallerand. 1989. Vers une méthodologie de validation transculturelle de questionnaires psychologiques : implications pour la recherche en langue française. *Psychologie Canadienne*, 30(4), 662-689.
- [23] Fons Van de Vijver and Ronald K. Hambleton. 1996. Translating tests: Some practical guidelines. *European Psychologist*, 1, 89-99.
- [24] Fons Van de Vijver and Norbert K. Tanzer. 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
- [25] A. Vermeeren, E. Law, V. Roto, M. Obrist, J. Hoonhout and K. Väänänen-Vainio-Mattila. 2010. User Experience Evaluation Methods: Current State and Development Needs. *Proceedings of NordiCHI 2010*, New York, NY: ACM Press. doi :10.1145/1868914.1868973