Revista Brasileira de Computação Aplicada, July, 2025

DOI: 10.5335/rbca.v17i2.16323
Vol. 17, Nº 2, pp. 34–45
Homepage: seer.upf.br/index.php/rbca/index

ORIGINAL PAPER

A methodology for assessing UX dimensions in standardized questionnaires

Erico de Souza Veriscimo [1], João Luiz Bernardes Júnior [1], Luciano Antonio Digiampietri [1]

[1]University of São Paulo

*ericoveriscimo@usp.br; jlbernardes@usp.br; digiampietri@usp.br

Received: 2024-09-25. **Revised:** 2025-07-12. **Accepted:** 2025-07-22.

Abstract

User Experience (UX) encompasses many and diverse variables, which contributes to making its evaluation a complex task. Self-evaluation through questionnaires is one of the most commonly used ways to evaluate UX, both in isolation and combined with other instruments or techniques, and there are currently many standardized and validated questionnaires that can be readily applied. Our work suggests a set of dimensions to describe different aspects of UX, and then proposes a methodology to quantify and compare how much different questionnaires emphasize each of these dimensions. We then select a set of ten questionnaires to exemplify the use of our methodology by comparing them in a particular case study. The results show that all selected questionnaires focus on only a few of the dimensions. Our research reinforces the importance of knowing different instruments and being able to compare them, choosing the one best suited to the context of use.

Keywords: Evaluation; Hedonic Criteria; Questionnaire; User Experience.

Resumo

A Experiência do Usuário (UX) abrange muitas e diversas variáveis, o que contribui para tornar sua avaliação uma tarefa complexa. A autoavaliação por meio de questionários é uma das formas mais comumente utilizadas para avaliar a UX, tanto de forma isolada quanto combinada com outros instrumentos ou técnicas, e atualmente existem muitos questionários padronizados e validados que podem ser facilmente aplicados. Nosso trabalho sugere um conjunto de dimensões para descrever diferentes aspectos da UX e, em seguida, propõe uma metodologia para quantificar e comparar o quanto diferentes questionários enfatizam cada uma dessas dimensões. Selecionamos, então, um conjunto de dez questionários para exemplificar o uso de nossa metodologia, comparando-os em um estudo de caso específico. Os resultados mostram que todos os questionários selecionados focam apenas em algumas das dimensões. Nossa pesquisa reforça a importância de conhecer diferentes instrumentos e ser capaz de compará-los, escolhendo o mais adequado ao contexto de uso.

Palavras-Chave: Avaliação; Critérios Hedônicos; Experiência do Usuário; Questionário.

1 Introduction

User Experience (UX) is among the chief factors used to design, describe, or improve how users interact with a system and feel when doing so (Rajeshkumar et al., 2013), particularly when user feedback may influence others, as in an app store (Mennig et al., 2019). UX is, therefore, extremely important for the acceptance, engagement, and competitive advantage of technological products, systems, or services (Veriscimo et al., 2020; Martinelli et al., 2022).

ISO 9241 (ISO, 2010) defines UX as the perceptions and responses from people resulting from the use of a product, system, or service, including all emotions, beliefs, preferences, perceptions, physical and psychological

responses, behaviors, and accomplishments that occur before, during and after use (ISO, 2010). With so many variables, choosing instruments to evaluate UX is not a simple task.

There are many ways to evaluate UX that may be combined or used in isolation, such as observation, event logs, self-evaluation through questionnaires, etc. Most evaluations, however, still focus more on user performance criteria without assigning more importance to criteria related to user emotions and pleasure (Veriscimo et al., 2020), which may be flawed since, according to ISO 9241 (ISO, 2010), emotion is fundamental to UX.

Among all these evaluation instruments, questionnaires are still used most often (Veriscimo et al., 2020). Silva et al. (2020) compare verbal and pictorial instruments and they argue that, while the latter may have fewer issues with language and cultural differences, verbal instruments can usually better measure feelings not so easily distinguished from each other. Using well-designed, validated, and standardized questionnaires is important to facilitate reproducibility, understanding, and sharing of results. By its fixed nature, however, each questionnaire covers distinct aspects of UX with more or less depth and that is by design since, for certain applications, some aspects may be of much greater importance than others. Martinelli et al. (2022) also report that, while most papers (60%) in their systematic review discussed some form of evaluation with users, industry professionals may struggle with how to conduct these types of evaluation. Therefore, it is important for anyone in this area not only to be aware of these instruments but also of which aspects of UX each favor (or to have a procedure to evaluate these favored aspects) to use the most adequate instrument for each evaluation and application.

Our goals, therefore, include:

- the definition and adoption of a set of general UX dimensions based on previous work (generic UX, affect/emotion, pleasure/fun, aesthetics/appeal, engagement/flow, hedonic qualities, pragmatic quality, which will be discussed in more detail in Section 3);
- the selection of a subset of relevant UX evaluation questionnaires with a particular context of use in mind;
- the presentation of a method to quantify dimension coverage in UX questionnaires (which we hope to be our main contribution) as well as their length (which may strongly impact certain evaluation strategies);
- the application of our proposed method to the previously selected set of questionnaires to illustrate how the method can be applied to facilitate a better choice of instrument on which dimensions an evaluation should emphasize. We hope our proposed set of dimensions and the method to quantify questionnaire coverage of these dimensions may be easily used to expand this analysis to more instruments.

Two points may require further clarification. First, our second goal mentions we have "a particular context of use in mind" and, indeed, the sort of analysis we propose to perform in our fourth goal must consider such a context, which will influence factors such as the dimensions that will be more important for a given experiment and the selection of a set of questionnaires to be compared. In this paper, we are considering the context of an experiment involving 3D interaction in which participants will need to answer questionnaires multiple times and the evaluation of the hedonic quality is of great importance. While this context will influence some decisions and discussions in this paper, it is just an example and the method we propose should be just as applicable to numerous other contexts of use.

The second point to clarify is that the analysis of instruments presented here is distinct from, and does not encompass, the evaluation of questionnaire psychometric properties such as reliability and concurrent validity. Ideally, these analyses should have been conducted before validation.

2 Related Work

Regarding UX questionnaire classification, Kocabalil et al. (2018) compared six instruments based on their coverage of ten proposed UX dimensions. Their analysis, however, focused only on speech-based interaction, and two of those dimensions are specific to that context and not generally applicable. Our work proposes a slightly simplified set of UX dimensions. We also illustrate our method by analyzing a broader and more general set of questionnaires than Kocabalil et al. (2018)

Schrepp (2020) analyzes 40 questionnaires to establish and compare their differences and similarities, positioning each questionnaire in a two-dimensional scaling according to their semantic distance and grouping them based on their coverage of certain aspects of UX. Their grouped presentation of the results, however, makes it somewhat difficult to choose a particular questionnaire for a study based on which aspects it should emphasize. Our goals for this paper include applying our method to the analysis of a considerably smaller set of questionnaires, but that allows us to discuss each in more detail.

Silva et al. (2020) also point out a lack of experimental studies comparing instruments to evaluate user subjective feelings and describe an experimental procedure to compare UX evaluation instruments, applying it to 4 of them, two pictorial and two verbal, but their analysis is entirely focused only on measuring hedonic aspect of UX. Thus, it compares aspects such as accuracy and descriptiveness of the instruments, instead of different dimensions of UX evaluation. In their experiment, they determined instruments using verbal terms to be more accurate and suggested this may be a consequence of how we normally refer to emotions in everyday life, through spoken or written verbal descriptions.

Kocabalil et al. (2018) and Schrepp (2020) propose the adoption of a set of UX dimensions (also referred to as aspects, factors or criteria) to compare different UX questionnaires. Based on these dimensions, each questionnaire item can be associated with one (or more) of them, depending on what the item is supposed to measure. In this way, questionnaires may be compared and classified based on which of these UX dimensions they

measure, or cover, in more detail.

Regarding which dimensions to choose, Winter et al. (2015) extract what they consider the 22 most important aspects of UX from an analysis of existing questionnaires, surveys, and discussion with specialists. Each of these aspects is described by a label: timeliness, adaptability, comfort, opacity, efficiency, immersion, intuitiveness, ease of use, usefulness, controllability, clearness, completeness, identity, novelty, originality, fun, stimulation, valence, connectedness, beauty, social aspects, and trust. Not all of these aspects are relevant to all use contexts (social aspects, for instance, hardly ever apply to single user experiences), and Schrepp (2020), a co-author of the previous study, uses only 16 of the 22 factors in his analysis. Given this relatively large number of factors, it is also difficult for shorter questionnaires to cover many of them in any depth.

Bargas-Avila and Hornbæk (2011) identify nine UX dimensions based on a systematic literature review: generic UX, affect/emotion, pleasure/fun, aesthetics/appeal, hedonic quality, engagement/flow, motivation, enchantment and frustration. Kocabalil et al. (2018) agree with these dimensions but, since instrumental and ergonomic factors are part of UX, added a tenth dimension: pragmatic quality. During their analysis, they also discarded the enchantment dimension since none of the instruments in their study included items relevant to it. Motivation and frustration also had a few items related to them, possibly because these three are closely related to other dimensions such as pleasure/fun and affect/emotion.

3 Proposed UX Dimensions

Based on the literature discussed in the previous section, in our method, we also adopt a set of UX dimensions on which to base our comparison of different questionnaires. Based on which UX dimensions questionnaires cover in more or less depth, it is possible to select which are more adequate for a given context, since different contexts of use may require a more detailed evaluation of different dimensions.

As part of our study, to select a set of UX dimensions to adopt, we analyzed the advantages and disadvantages of the sets proposed by Winter et al. (2015), Schrepp (2020), Bargas-Avila and Hornbæk (2011) and Kocabalil et al. (2018), briefly presented in the last section. As previously discussed, we believe the larger number of aspects proposed by Winter et al. (2015) and even by Schrepp (2020) bring some disadvantages (for instance, many aspects are irrelevant to many contexts of use and shorter questionnaires can cover very few of them in any depth) which make their use for our intended purpose, comparing the coverage of different questionnaires, less advantageous. So we favor the dimensions proposed by Bargas-Avila and Hornbæk (2011) and Kocabalil et al. (2018). Furthermore, we consider the latter proposal to be an improvement and evolution of the former.

Therefore, we adopt a set of dimensions based on the one proposed by Kocabalil et al. (2018), but go a bit further and group the two dimensions shown to be least useful in their work with those that are more closely related to motivation and frustration are included in affect/emotion and enchantment is included in pleasure/fun. We then end up with a set of only seven dimensions: affect/emotion, pleasure/fun, aesthetics/appeal, engagement/flow, hedonic quality, pragmatic quality, and generic UX. These seven dimensions are discussed in more detail in the following subsections and should be useful to analyze which aspects of UX a given instrument prioritizes. There is some unavoidable overlap between dimensions and an item can pertain to more than one, but, as will be shown below, each has unique characteristics that we aim to differentiate in our method of comparison.

3.1 Affect/emotion

Affect/emotion is related to feelings, and emotions, except for pleasure, fun, or enchantment, and positive or negative affect that arise from interacting with the system, including feelings of frustration and motivation. Statements such as "I was happy/I enjoyed/I was frustrated/I felt challenged/I felt pressured [interacting with the system]." are contemplated in this dimension. A statement such as "I had fun with the system" would not be included here because it is included in the pleasure/fun dimension. Physical sensations, such as fatigue, would also not be included, instead being part of the Pragmatic Quality dimension.

3.2 Pleasure/fun

The pleasure/fun dimension includes questions measuring how much or how little of these feelings of pleasure or fun the user derived from interacting with the system. It also includes questions regarding user enchantment or lack of it. "Using this system was fun", "I was bored", "I was impressed", and "I was interested" are examples of a questionnaire item related to this dimension. "I felt challenged" would not be included in this dimension, instead being part of the affect/emotion and engagement/flow dimensions.

3.3 Aesthetics/appeal

The aesthetics/appeal dimension includes items that measure system design and user perception of aesthetics and beauty, including the perception of creativity or originality in the interface. Questions related to this dimension often directly reference apparent qualities of the user interface or impressions caused by these qualities. "It was aesthetically pleasing", "It was impressive", "It was creative", and "It was original" are examples of questions related to this dimension. "It was easy to learn" is an example of an item that would not be included in this dimension, instead being included in the pragmatic quality dimension.

3.4 Engagement/flow

Engagement/flow: the concept of flow, rather popular in the videogame industry, originates from psychology studies involving challenge and/or creativity and describes

a state in which subjects are completely engaged with what they are doing, being able to shut off distractions and achieve peak performance (Frey et al., 2013). It is often a goal for video game designers and is closely linked to the perceived challenge of a task, which must neither be too low nor too high for the user to achieve this state. Engagement is a quality of UX characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect (O'Brien and Toms, 2008). Questionnaire items describing an experience as challenging (but in a positive rather than negative sense), for instance, are contemplated in this dimension, as are questions related to user focus in the task to the exclusion of distractions. "I was focused", "I was challenged", "I lost track of time", "I was bored", "I forgot about everything else around me", and, "I thought about other things" are examples of items related to this dimension. "It was innovative" or "It was original" are examples of statements unrelated to engagement/flow.

3.5 Hedonic quality

The hedonic quality supports stimulation, communicates identity, and provokes memory, referring to what a system is perceived to be, such as competent, related to other systems, special, interesting, exclusive, impressive, original, innovative, etc. (Hassenzahl et al., 2008). This dimension encompasses all items related to or lack of pleasure and emotions. Thus, this dimension has a significant overlap with the three previous dimensions, but also includes other aspects not covered by the previous ones, such as "I was good at it" and "I felt good". The only items not covered by this dimension would be those related to the pragmatic quality, such as "I thought it was difficult".

3.6 Pragmatic quality

Pragmatic quality measures the perceived capabilities of the system to support the execution of tasks and achievement of goals in an efficient manner, even including ergonomic aspects (Kocabalil et al., 2018). This dimension includes most items associated with classic usability attributes such as efficiency (particularly in task execution), ergonomics, learnability, rememberability, error handling, etc. "The system is efficient", "It was easy to learn", "It was complicated", and "It was practical" are usual examples of questionnaire statements related to this dimension. "It met my expectations" is an example of an item that would not be included in this dimension.

3.7 Generic UX

Generic UX refers to items that are not related to any particular aspect of UX but, instead, try to evaluate the system in broader, more general terms, such as "I would recommend this system to other users" or "I would not trade this product for any other".

4 Questionnaires and their selection

This section presents considerations about how we selected a set of questionnaires to compare, as well as a brief discussion of each questionnaire thus selected, so that in the next sections we can make use of this set when employing and analyzing our proposed comparison method.

4.1 Questionnaire Selection

In the present work, we must select set of questionnaires to illustrate and discuss the use of our proposed method, which compares them based on coverage of UX dimensions (and size). Thus, we do not necessarily need to select some set of "most important" questionnaires to compare, but only a moderately sized set of instruments which are often used for UX evaluation in different contexts (including, but not limited to, the context discussed in the end of the introduction). Furthermore, we did not intend to be as comprehensive as Schrepp (2020), mostly because we intend to discuss more information about each questionnaire. Based on these considerations, we chose a set of ten questionnaires reported to be used often in the UX evaluation literature (Kocabalil et al., 2018; Díaz-Oreiro et al., 2019; Veriscimo et al., 2020; Sandesara et al., 2022), particularly in systematic reviews. The questionnaires included according to this rationale were:

- Two of the six discussed by Kocabalil et al. (2018). SASSI, SUISQ, MOS-X, and Paradise were disregarded as these questionnaires and their analysis focused heavily only on speech-based interaction.
- AttrakDiff, UEQ and meCUE, reported as the most recognized instruments in a systematic review on standardized questionnaires for UX assessment (Díaz-Oreiro et al., 2019).
- Four of the five most frequently used according to a systematic review on UX assessment in 3D interaction (Veriscimo et al., 2020), two of which (UEQ and AttrakDiff) had already been included. The questionnaire disregarded was the SSQ, a disease simulation questionnaire.
- UMUX, cited (along with SUS, which was already included) in a review on mobile application design and experience (Sandesara et al., 2022) as one of the instruments most often used to evaluate experience and usability in that context.

So the questionnaires selected based on these papers were: AttrakDiff, SUS, NASA-TLX, UEQ, UEQS, GEQ, MeCUE and UMUX. Because both UEQ and UEQS (a short version of UEQ) were included, We added the In-Game GEQ and UMUX-Lite questionnaire to the set as well, which are also shorter versions of instruments already in the set. Despite opting not having included some of the questionnaires cited in the papers listed above, for the reasons just presented, we did analyze them using our method and they did not particularly stand out in any of our proposed dimension, cementing our decision. Before going further, we present a brief overview of each of the selected instruments.

4.2 AttrakDiff

AttrakDiff aims to measure the perception of hedonic and pragmatic qualities (Hassenzahl et al., 2003). It is comprised of 28 items divided into three broad categories: pragmatic quality, hedonic quality and attractiveness. Each item is comprised of a Likert scale with two words with opposing meanings and seven gradations between them, so respondents can indicate how close to each extremity is their perception regarding that item.

4.3 SUS

SUS or System Usability Scale (Brooke, 1996) performs an evaluation focusing mostly on perceived ease of use and system learnability and includes ten items, without categorization. Items are structured as statements with a five-point Likert agreement scale (which has been considered less reliable than a seven-point scale).

4.4 meCUE

MeCUE (Minge et al., 2017) is a standardized questionnaire measurement of UX. It consists of four separately validated modules which refer to instrumental and noninstrumental product perceptions, user emotions, consequences of usage, and an overall judgment of attractiveness, a total of 34 items. Its items are structured as seven-point Likert scales, with one extreme referring to “strongly disagree” and the other to “strongly agree”. Fig. 1 shows one of these items.

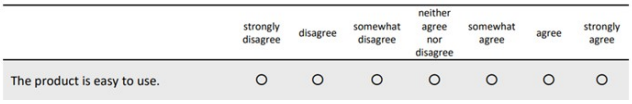

Figure 1: MeCUE item (Minge et al., 2017)

4.5 NASA-TLX

NASA-TLX is used for subjective evaluation of workload (Hart and Staveland, 1988). It contains only six uncategorized items structured as questions, with answers on a 21-point scale. Fig. 2 shows one of these items.

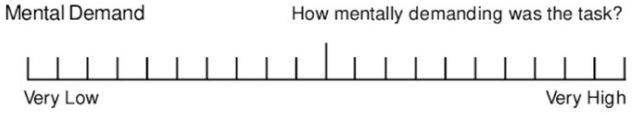

Figure 2: NASA-TLX item (Pilco et al., 2019)

4.6 UEQ and UEQS

UEQ and UEQS: the User Experience Questionnaire evaluates UX with 26 items organized in six categories: attractiveness, perspicuity, efficiency, dependability, stimulation and novelty (Laugwitz et al., 2008). UEQS is a shorter version of UEQ with only eight questions. Items are structured as seven-point Likert scales between antonyms, such as complicated and easy. Fig. 3 shows one of these items.

Figure 3: UEQ item (Laugwitz et al., 2008)

4.7 UMUX and UMUX-Lite

UMUX (Finstad, 2010) and UMUX-Lite (Lewis et al., 2013): The Usability Metric for User Experience is a questionnaire with four items used for the subjective assessment of user experience. UMUX-Lite is a shorter version of UMUX with only two items. Both with items structured as seven-point Likert scales, with one extreme referring to "strongly disagree" and the other to "strongly agree".

4.8 GEQ and In-Game GEQ

GEQ and In-Game GEQ, or the Game Experience Questionnaire (Norman, 2013), evaluates user experience in the particular context of assessing naturalistic gaming (i.e., when gamers have voluntarily decided to play) and thus gives a little more importance to factors such as engagement, immersion, flow and perception of competence. It includes 33 items divided in seven categories: competence, sensory and imaginative immersion, flow, tension/annoyance, challenge, negative affect and positive affect. Items are structured as statements about player feelings or perceptions during the experience, such as “I felt content”, about which the player selects an option in a five-point scale going from “not at all” to “extremely”. GEQ also has an “in-game version” (IJsselsteijn et al., 2013) with 14 questions developed to assess game experience at multiple intervals during a game session. GEQ may also include a social presence module and a post-game module that may be applied.

5 Method

Based on the dimensions discussed previously, we propose a simple method to evaluate UX questionnaires. Our method consists of 5 steps: 1) familiarization with the proposed dimensions; 2) classification of questionnaire items; 3) confronting and resolving differences in classification; 4) classifying by length and, finally, 5)

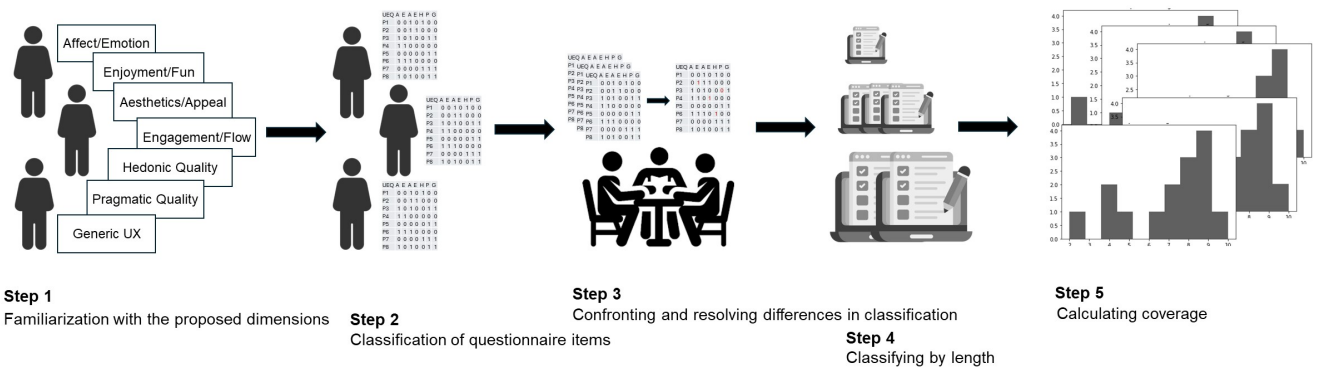

Figure 4: Method to evaluate UX questionnaires

calculating coverage. We discuss each of these steps in more detail in the following subsections. Fig. 4 shows the method to evaluate UX questionnaires.

One step that is not explicitly included in this method, but which must be accomplished before applying it, is selecting a set of questionnaires to compare. Often, researchers will already have several instruments with which they have experience and/or are licensed to use, and finding this set of instruments to compare is not a problem, the set is already a given. However, for cases in which the selection must be made from a broader set of instruments, for instance, found in the relevant literature, we have provided some relevant papers in Section 2 and Section 4.1 which discuss several questionnaires broadly used for UX evaluation. We also present in Section 4.1 our rationale for including or discarding instruments found in the literature (we merely discard the instruments which, based on the literature, are ill-suited to the task), which may serve as an example.

5.1 Familiarization with dimensions

First of all, of course, all participants should familiarize themselves as much as possible with those dimensions, their descriptions and examples and counter-examples. Classifying each questionnaire item within these dimensions may have a subjective component which makes it a complex task, which can be aggravated by an incomplete or inconsistent understanding of what each dimension encompasses.

5.2 Item classification

The familiarization step is then followed by classifying each questionnaire item for all questionnaires in the selected set as being related to one or more dimensions. We accomplish this classification by creating a table for each questionnaire, with each line representing an item and the dimensions placed in columns. For each table cell, we assign the value 1 if that item pertains to that dimension or 0 if it does not. Table 1 shows a sample classification of an UEQ item.

Because of the somewhat subjective nature of this classification, we recommend having the process performed by more than one person (at least two participants with knowledge of UX evaluation and familiarized with our proposed dimensions - in this particular study, we used exactly two UX researchers to perform this step). We also recommend that the classification process be done separately by the participants, only allowing them to confront their classifications in the next step. But each participant is free to go back to items previously classified, even in other questionnaires, and update their classification before the next step. In our experience, being confronted with new items or certain groups of items tends to improve a participant's understanding of the dimensions and it is useful to allow the better understanding thus acquired to reflect in updating classifications done previously.

5.3 Resolution of classification differences

The next step is confronting all classifications done by different participants and resolving items which were classified differently. In our experience, participants agreed for most items, but there were more than a few disagreements as well. We propose having another person who was not involved in the classification process to act as a judge (and in this study we did employ a third researcher to this end). For each case in which classifications by different participants disagree, each participant explains the rationale for their classification and the judge decides which stands. In practice, during the classification process we underwent while writing this paper, we actually reached consensus every time while explaining the rationales for each item.

5.4 Length classification

We then classify questionnaires by length, with those with less than 10 items being classified as short, those between 10 and 24 items as medium and those with 25 items or more as long. The length obviously impacts the time it takes to answer the questionnaire and may be an important factor to consider in certain experiments, particularly if the instrument will be used several times

Table 1: Dimension identification for an UEQ item.

UEQ	Affect/ Emotion	Enjoyment /Fun	Aesthetics /Appeal	Engagement /Flow	Hedonic Quality	Pragmatic Quality	Generic UX
Attractive/ Unattractive	0	0	1	0	1	0	0

over an experiment's course.

5.5 Coverage calculation

Finally, we calculate coverage for each dimension in a questionnaire by adding up the number 'n' of items that were classified as pertaining to that dimension and dividing 'n' by the total number 'ti' of items in the questionnaire, normalizing this result as shown in Eq. (1). We also add up these scores for each dimension to obtain a general coverage value.

$$f(x) = n/ti \quad (1)$$

5.6 How to use these results

Based on questionnaire length and coverage and knowing which dimensions they wish to prioritize and how often and for how long participants are expected to fill up questionnaires in a given experiment, researchers can more easily select the most appropriate instrument from the set they were initially considering.

To illustrate this proposed procedure, we applied it to NASA-TLX, UEQ, UEQS, GEQ, In-Game GEQ, UMUX, UMUX-Lite, SUS, AttrakDiff and MeCUE. The identification of dimensions for AttrakDiff and SUS was mostly extracted from Kocabalil et al. (2018), but adapting it to the lower number of dimensions in our study (including the enchantment dimension in pleasure/fun and the motivation and frustration dimensions in affect/emotion). We discuss these results in the next section.

6 Results

Following the methodology discussed previously, we present the results for the selected set of questionnaires. We compare questionnaires to each other in terms of coverage, individually and in general, and classify each by size, based on its number of items.

6.1 Comparing questionnaires

Table 2 and Table 3 show the comparison of all analyzed questionnaires. Each line represents a dimension and shows how many items of the questionnaire in that column are related to that dimension. Each item may be related to more than one dimension. In parenthesis is shown a normalized percentage value, calculated according to Eq. (1).

Another important characteristic of a questionnaire is its length, which impacts how long it takes to fill out during an experiment. Fig. 5 shows the number of items for each analyzed questionnaire. To aid in the visualization of this data, Fig. 6 illustrates the information from Table 2 and Table 3 comparing all questionnaires based on the normalized value for each dimension and Fig. 7 shows a radar plot comparing UX dimension coverage in the questionnaires.

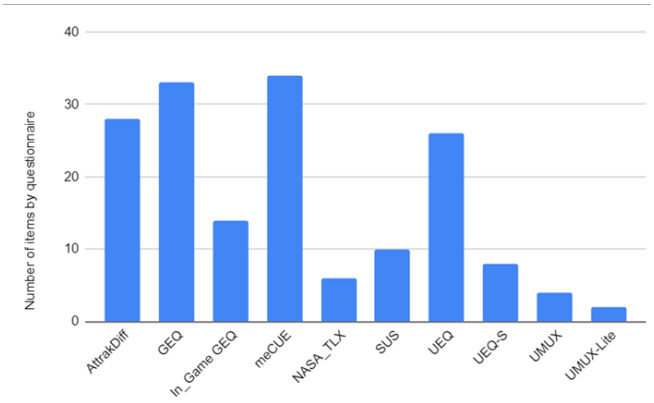

Figure 5: Items in each questionnaire

According to their length, questionnaires were classified as long (with 25 items or more, 4 of the analyzed questionnaires were classified as long), medium (between 10 and 24 items, 2 questionnaires) and short (less than 10 items, 4 questionnaires). For each questionnaire, their favored dimensions were determined in the following manner: we assessed each questionnaire separately assuring an evaluation that is independent of the set of selected questionnaires. A cutoff score was calculated for each questionnaire, thus, if the dimension contains more than the cutoff score, it is considered strong. The cutoff score was calculated as follows: the sum of the related dimensions of all items was performed and divided by the total number of items. For example: in the NASA-TLX questionnaire, all related dimensions were added up, totaling a value of 8, divided by 6 (the total number of items), resulting in a cutoff score of 1.14, thus, only the pragmatic quality dimension will be considered strong. The Generic UX dimension was not considered in this analysis. A final coverage score was also created, summing up the coverage for all dimensions in each questionnaire. Table 4 presents the results for all these metrics for the analyzed questionnaires.

7 Discussion

The results shown in the previous section indicate that, for our set of analyzed questionnaires, most focus their

Table 2: Comparing analyzed questionnaires

	Questionnaires (QA)				
Dimensions	AttrakDiff[1]	GEQ	IG-GEQ	NASA-TLX	SUS[1]
Affect/Emotion	1 (4%)	10 (30%)	5 (36%)	1 (17%)	0 (0%)
Enjoyment/Fun	0 (0%)	4 (12%)	2 (14%)	0 (0%)	0 (0%)
Aesthetics/Appeal	7 (25%)	2 (6%)	1 (7%)	0 (0%)	0 (0%)
Engagement/Flow	16 (57%)	11 (33%)	6 (43%)	1 (17%)	1 (10%)
Hedonic Quality	14 (50%)	15 (45 %)	7 (50%)	1 (17%)	0 (0%)
Pragmatic Quality	7 (25%)	5 (15%)	3 (21%)	5 (83%)	9 (90%)
Generic UX	0 (0%)	8 (24%)	2 (14%)	0 (0%)	1 (10%)
Total number of items	28	33	14	6	10

[1]Adapted from Kocabalil et al. (Kocabalil et al., 2018).

Table 3: Comparing analyzed questionnaires

	Questionnaires (QA)				
Dimensions	UEQ	UEQS	UMUX	UMUX-Lite	meCUE
Affect/Emotion	13 (50%)	5 (62%)	2 (50%)	1 (50%)	15 (44%)
Enjoyment/Fun	5 (19%)	2 (25%)	1 (25%)	0 (0%)	0 (0%)
Aesthetics/Appeal	10 (38%)	4 (50%)	0 (0%)	0 (0%)	3 (9%)
Engagement/Flow	6 (23%)	2 (25%)	0 (0%)	0 (0%)	2 (6%)
Hedonic Quality	17 (65%)	7 (87 %)	2 (50%)	1 (50%)	20 (59%)
Pragmatic Quality	7 (27%)	2 (25%)	3 (75%)	2 (100%)	8 (24%)
Generic UX	13 (50%)	5 (62%)	2 (50%)	1 (50%)	10 (29%)
Total number of items	26	8	4	2	34

Table 4: Other metrics

QA	Size	Strong dimensions	Coverage Score
NASA-TLX	Short	· Pragmatic Quality	133
UEQ-S	Short	· Hedonic Quality · Affect/Emotion · Aesthetics/Appeal	338
UMUX	Short	· Pragmatic Quality · Hedonic Quality · Affect/Emotion	250
UMUX-Lite	Short	· Pragmatic Quality · Hedonic Quality · Affect/Emotion	250
In-Game GEQ	Medium	· Hedonic Quality · Engagement/Flow · Affect/Emotion	186
SUS	Medium	· Pragmatic Quality	110
AttrakDiff	Long	· Engagement/Flow · Hedonic Quality · Pragmatic Quality · Aesthetics/Appeal	161
GEQ	Long	· Hedonic Quality · Engagement/Flow · Affect/Emotion	167
MeCUE	Long	· Hedonic Quality · Affect/Emotion	171
UEQ	Long	· Hedonic Quality · Affect/Emotion	273

inquiry on only a few of the adopted UX dimensions. Having too many items in a questionnaire makes its use cumbersome, tedious and tiresome, which might help explain why different instruments of lower length opt to focus on fewer dimensions. This, in turn, reinforces the importance of knowing different instruments and choosing well to use one more adequate to the interaction context being investigated.

None of the analyzed questionnaires had a good coverage for all dimensions, but UEQ (and UEQ-S) had the best coverage in this set, despite having a low coverage of the engagement/flow, enjoyment/fun and pragmatic dimensions, therefore it would be recommended for different UX studies in general, as long as these three dimensions are not of great importance for those studies. AttrakDiff and GEQ (and In-Game GEQ) had the best coverage in engagement/flow and, in general, had similar coverages, except for their weak points (affect/emotion for AttrakDiff and aesthetics/appeal for GEQ). Both are good options, therefore, for studies investigating engagement or flow with particular interest. To investigate the pragmatic dimension, SUS and NASA-TLX showed the best coverage in this set and would be recommended. Classical usability evaluation questionnaires tend to focus on this dimension to the almost entire exclusion of the others and are good options for studies majorly concerned with the pragmatic dimension.

These results seem to agree with those presented by Schrepp (2020), which place AttrakDiff and UEQ in the group of "questionnaires with a stronger focus on non-task related or hedonic UX aspects" and SUS in the group focusing more strongly on pragmatic aspects. Schrepp (2020) also points out how few instruments cover a large range of UX aspects, with the exception of a few rather large questionnaires, such as UEQ+.

While our set of selected questionnaires was different enough from that discussed by Kocabalil et al. (2018) to make a direct comparison of little use, results from both our work and theirs do indicate many of the

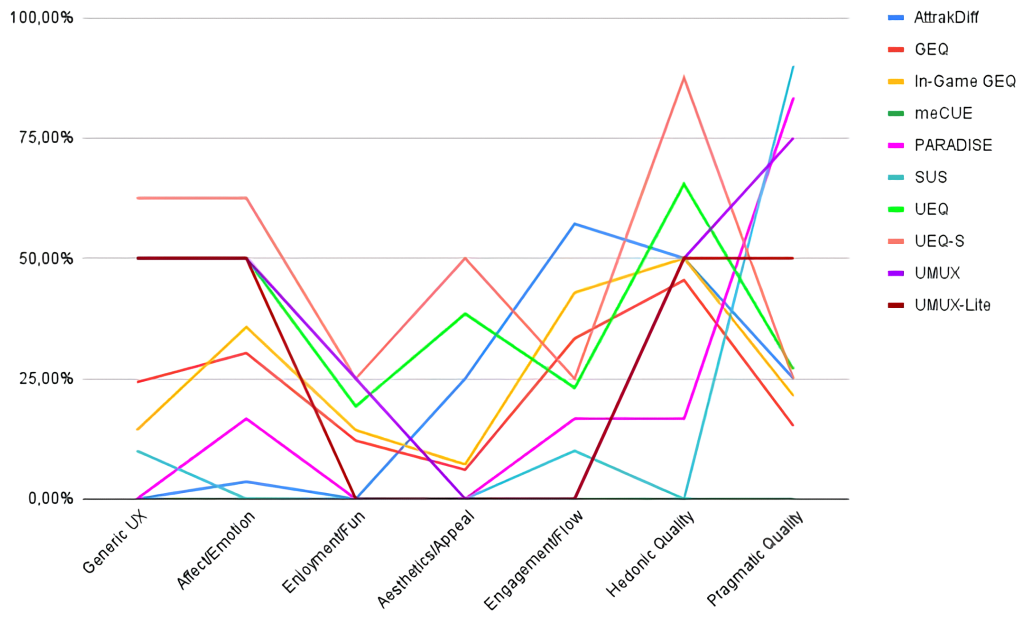

Figure 6: Comparison of analyzed questionnaires based on normalized values

same conclusions: that it is rare for a questionnaire to provide sufficient coverage across all UX dimensions unless it is rather long and therefore less practical; that questionnaires may be recommended to a particular study based on their coverage and which dimensions they favor; and, as we discuss below, that if feasible, combining multiple questionnaires can be very useful.

Table 5 shows recommended instruments from this set for evaluations with more emphasis on each of the dimensions. We recommend a few options for each dimension in order of best coverage. Based on this analysis, for the case we presented in the introduction in which a short instrument focusing on hedonic aspects was more desirable, we would select UEQ-S.

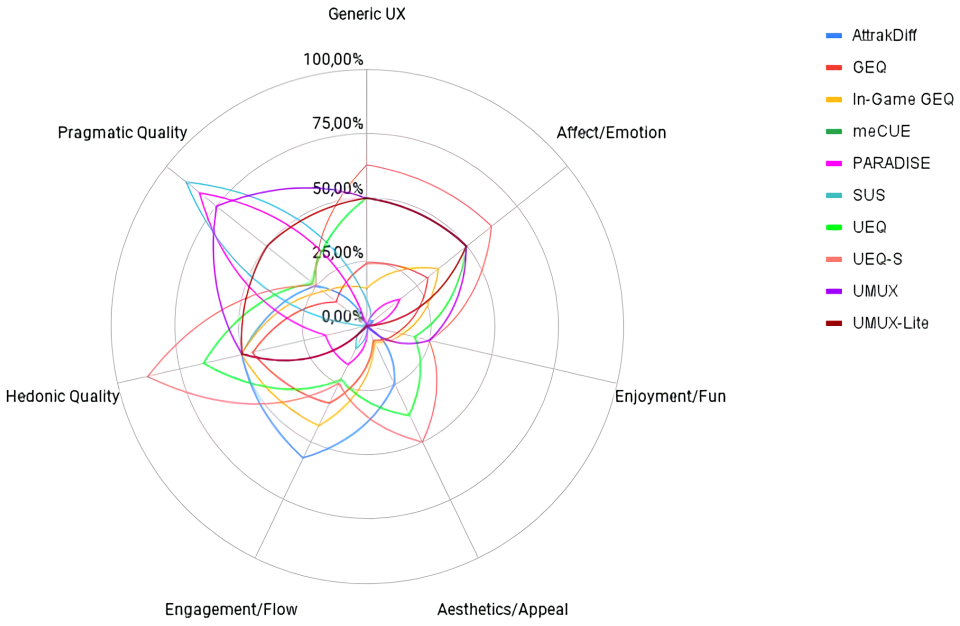

Figure 7: Radar plot comparing questionnaire coverage

Table 5: Recommendations based on favored dimensions

UX Dimensions	Questionnaires Short	Medium	Long
Affect/Emotion	· UEQ-S · UMUX · UMUX-Lite	· In-Game GEQ	· UEQ · GEQ
Enjoyment/Fun	· None	· None	· None
Aesthetics/Appeal	· UEQ-S	· None	· AttrakDiff · MeCUE
Engagement/Flow	· None	· In-Game GEQ	· AttrakDiff · GEQ
Hedonic Quality	· UEQ-S · UMUX · UMUX-Lite	· In-Game GEQ	· UEQ · AttrakDiff · MeCUE · GEQ
Pragmatic Quality	· NASA-TLX · UMUX · UMUX-Lite	· SUS	· None
Generic UX	· UEQ-S · UMU · UMUX-Lite	· None	· UEQ · GEQ · MeCUE

Table 6: Combining UEQ-S and NASA-TLX

	Questionnaires		
Dimensions	UEQS	NASA-TLX	UEQS + NASA-TLX
Generic UX	5 (62,5%)[2]	0 (00,0%)	5 (35,7%)
Affect/Emotion	5 (62,5%)[2]	1 (16,7%)	6 (42,9%)[2]
Enjoyment/Fun	2 (25,0%)	0 (00,0%)	2 (14,3%)
Aesthetics/Appeal	4 (50,0%)[2]	0 (00,0%)	4 (28,6%)
Engagement/Flow	2 (25,0%)	1 (16,7%)	3 (21,4%)
Hedonic Quality	7 (87,5%)[2]	1 (16,7%)	8 (57,1%)[2]
Pragmatic Quality	2 (25,0%)	5 (83,3%)[2]	7 (50,0%)[2]
Total number of items	8	6	14

If increasing coverage is important enough to merit increasing questionnaire length, a combination of instruments could be advantageous (particularly when combining shorter questionnaires). A combination of UEQ-S and NASA-TLX, for instance, would increase coverage for all dimensions compared to each individual instrument, however, its strong dimensions will change. Table 6 presents this comparison.

8 Conclusion

In this study, we proposed and applied a method for assessing UX dimensions in standardized questionnaires, facilitating the identification of the most appropriate choices.

To do this, we selected a set of seven UX evaluation questionnaires, proposed a division of UX into seven dimensions (Affect/Emotion, Enjoyment/Fun, Aesthetics/Appeal, Engagement/Flow, Hedonic Quality, Pragmatic Quality, and Generic UX) and compared these questionnaires based on their coverage of each of the dimensions. None of the questionnaires analyzed in the set had good coverage in all dimensions, illustrating how complex is the evaluation of such a broad concept as UX in all of its aspects and, therefore, how important it is to consider which dimensions are more important in each context and to choose adequate instruments based on that. UEQ and UEQ-S had the best coverage in general in this set and would be the most adequate for studies focusing on Affect/Emotion, Aesthetics/Appeal, and Hedonic Quality, i.e., not emphasizing Pragmatic Quality, in which case SUS and NASA-TLX would be the most adequate, or Engagement/Flow, for which we would recommend AttrakDiff, In-Game GEQ or GEQ. Considering the set of selected questionnaires, none of them stood out in the dimension of Enjoyment/Fun. Combining more than one of these standardized questionnaires may also be useful if done carefully, including an analysis of the impact of the increased length of the evaluation. In future work, we intend to include other questionnaires that are considered important and relevant in the UX evaluation, but that did not appear in our current research, such as the QUIS questionnaire (Harper and Norman, 1993). We also believe that using the seven dimensions proposed here and the procedure to evaluate and compare different questionnaires described in the Method section, based on these dimensions, can be a useful contribution to the area and a source of interesting future work, expanding this analysis to include more questionnaires.

As limitations, we highlight that the application of the method proposed in this study was conducted considering the seven specified dimensions and the chosen questionnaires. However, we assert that these limitations do not detract from the significance of

our contribution. Our selection process was based on publications from multiple well-established authors and the widespread adoption of the chosen questionnaires.

References

Bargas-Avila, J. A. and Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, Association for Computing Machinery, New York, NY, USA, p. 2689–2698. https://doi.org/10.1145/1978942.1979336.

Brooke, J. (1996). *Usability evaluation in industry*, CRC Press, London, chapter SUS-A quick and dirty usability scale, pp. 189–194.

Díaz-Oreiro, I., López, G., Quesada, L. and Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review, *Multidisciplinary digital publishing institute proceedings* **31**(1): 14. http://dx.doi.org/10.3390/proceedings2019031014.

Finstad, K. (2010). The usability metric for user experience, *Interacting with computers* **22**(5): 323–327. http://dx.doi.org/10.1016/j.intcom.2010.04.004.

Frey, J., Mühl, C., Lotte, F. and Hachet, M. (2013). Review of the use of electroencephalography as an evaluation method for human-computer interaction. http://dx.doi.org/10.48550/arXiv.1311.2222.

Harper, B. D. and Norman, K. L. (1993). Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5, *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference*, Vol. 224, Citeseer, VA, USA, p. 228.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research, *Advances in psychology*, Vol. 52, Elsevier, pp. 139–183. http://dx.doi.org/10.1016/S0166-4115(08)62386-9.

Hassenzahl, M., Burmester, M. and Koller, F. (2003). *Mensch & Computer 2003: Interaktion in Bewegung*, Springer, chapter AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität, pp. 187–196. http://dx.doi.org/10.3390/math10142380.

Hassenzahl, M., Schöbel, M. and Trautmann, T. (2008). How motivational orientation influences the evaluation and choice of hedonic and pragmatic interactive products: The role of regulatory focus, *Interacting with computers* **20**(4-5): 473–479. http://dx.doi.org/10.1016/j.intcom.2008.05.001.

IJsselsteijn, W., de Kort, Y. and Poels, K. (2013). The game experience questionnaire. http://dx.doi.org/10.1016/j.intcom.2010.04.004.

ISO, D. (2010). 9241-210: 2010. ergonomics of human system interaction-part 210: Human-centred design for interactive systems.
URL: *https://www.iso.org/standard/52075.html*

Kocabalil, A. B., Laranjo, L. and Coiera, E. (2018). Measuring user experience in conversational interfaces: a comparison of six questionnaires, *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*, pp. 1–12. https://doi.org/10.14236/ewic/HCI2018.21.

Laugwitz, B., Held, T. and Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire, *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*, Springer, pp. 63–76. http://dx.doi.org/10.1007/978-3-540-89350-9_6.

Lewis, J. R., Utesch, B. S. and Maher, D. E. (2013). Umux-lite: when there's no time for the sus, *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2099–2102. http://dx.doi.org/10.1145/2470654.2481287.

Martinelli, S., Lopes, L. and Zaina, L. (2022). Ux research in the software industry: an investigation of long-term ux practices, *Anais do XXI Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, SBC, Porto Alegre, RS, Brasil.
URL: *https://sol.sbc.org.br/index.php/ihc/article/view/22277*

Mennig, P., Scherr, S. A. and Elberzhager, F. (2019). Supporting rapid product changes through emotional tracking, *2019 IEEE/ACM 4th International Workshop on Emotion Awareness in Software Engineering (SEmotion)*, IEEE, pp. 8–12. https://doi.org/10.1109/SEmotion.2019.00009.

Minge, M., Thüring, M., Wagner, I. and Kuhr, C. V. (2017). The meCUE questionnaire: a modular tool for measuring user experience, *Advances in Ergonomics Modeling, Usability & Special Populations: Proceedings of the AHFE 2016 International Conference on Ergonomics Modeling, Usability & Special Populations, July 27-31, 2016, Walt Disney World®, Florida, USA*, Springer, pp. 115–128. http://dx.doi.org/10.1007/978-3-319-41685-4_11.

Norman, K. L. (2013). Geq (game engagement/experience questionnaire): a review of two papers, *Interacting with computers* **25**(4): 278–283. http://dx.doi.org/10.1093/iwc/iwt009.

O'Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology, *Journal of the American society for Information Science and Technology* **59**(6): 938–955. http://dx.doi.org/10.1002/asi.20801.

Pilco, H., Sanchez-Gordon, S., Calle-Jimenez, T., Pérez-Medina, J. L., Rybarczyk, Y., Jadán-Guerrero, J., Maldonado, C. G. and Nunes, I. L. (2019). An agile approach to improve the usability of a physical

telerehabilitation platform, *Applied Sciences* **9**(3): 480. http://dx.doi.org/10.3390/app9030480.

Rajeshkumar, S., Omar, R. and Mahmud, M. (2013). Taxonomies of user experience (ux) evaluation methods, *2013 International Conference on Research and Innovation in Information Systems (ICRIIS)*, IEEE, pp. 533–538. https://doi.org/10.1109/ICRIIS.2013.6716765.

Sandesara, M., Bodkhe, U., Tanwar, S., Alshehri, M. D., Sharma, R., Neagu, B.-C., Grigoras, G. and Raboaca, M. S. (2022). Design and experience of mobile applications: A pilot survey, *Mathematics* **10**(14): 2380. http://dx.doi.org/10.3390/math10142380.

Schrepp, M. (2020). A comparison of ux questionnaires - what is their underlying concept of user experience?, *in* C. Hansen, A. Nürnberger and B. Preim (eds), *Mensch und Computer 2020 - Workshopband*, Gesellschaft für Informatik e.V., Bonn, p. 6. https://doi.org/10.18420/muc2020-ws105-236.

Silva, L. G. Z. d., Guimarães, P. D., de Souza Gomes, L. O. and de Almeida Neris, V. P. (2020). A comparative study of users' subjective feeling collection instruments, *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*, IHC '20, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3424953.3426642.
URL: *https://doi.org/10.1145/3424953.3426642*

Veriscimo, E. d. S., Bernardes Junior, J. L. and Digiampietri, L. A. (2020). Evaluating user experience in 3d interaction: a systematic review, *XVI Brazilian Symposium on Information Systems*, pp. 1–8. https://doi.org/10.1145/3411564.3411640.

Winter, D., Schrepp, M. and Thomaschewski, J. (2015). Faktoren der user experience - systematische Übersicht über produktrelevante ux-qualitätsaspekte, *in* A. Endmann, H. Fischer and M. Krökel (eds), *Mensch und Computer 2015 – Usability Professionals*, De Gruyter Oldenbourg, Berlin, pp. 33–41.