Evaluation of Anticipated User Experience

A method to compare user expectations and actual user experience

Dominik Oelfin
TI SAP CP Experience
SAP SE
Walldorf, Germany
dominik.oelfin@sap.com

Martin Schrepp
TI SAP CP Experience
SAP SE
Walldorf, Germany
martin.schrepp@sap.com

ABSTRACT

User experience encompasses all the effects that the use of a product has on users - before, during and after use. But what do future users expect from a product? There are various evaluation and measurement methods to examine actual user experience. However, only a few methods allow to explore the expectations of future users before they use a product. We describe a standardized method to explore user expectations, that is supported by a web tool for efficient data collection. The method is based on a list of positive UX properties that can be configured by a researcher. Participants in a study can indicate for each property whether they expect it from the product or not. The result can be visualized via a word cloud. The same properties can be used to rate actual user experience in a classical questionnaire format. Thus, the method allows to compare user expectations with the actual user experience and helps to generate a shared understanding about the important UX properties for a product. Finally, the method was tested in a concrete evaluation project.

CCS CONCEPTS

Human-centered computing → HCI design and evaluation methods

KEYWORDS

User Experience, User Expectations, Anticipated Use, Evaluation

1 Introduction

Good software is characterized by good user experience (UX). Thus, UX is an important factor that must be considered in any human-centered product design [1] process.

UX is a quite heterogeneous concept, which contains various quality aspects, for example efficiency, learnability, fun of use, originality, or aesthetics of the product design [2]. However, UX does not always have to be perfected in all its aspects to satisfy users [3]. Depending on the product some UX aspects maybe more important for user satisfaction than others. It is thus useful to determine the users' main UX expectations towards the product. The UX perception of the product does not start with the actual use, but even before that. An essential part of UX is the expectation towards the product before it is used for the first time [4, 5].

This paper investigates the concept of user expectation concerning UX and highlights how the structured capturing of user expectations can contribute to user-centered software development. Our goal is to elaborate an efficient method for evaluating user expectation and to support it with a web tool. The method allows to determine if the final design of a product matches the user expectations captured in the early design phase. We test and illustrate the method in two studies.

2 Anticipated use

UX impressions towards a product develop over time. There are several models that try to describe and explain this process. Before the actual use of a product, the so-called anticipation takes place. Future users develop expectations about the later experience of using the product [4, 7]. Motivation, needs and predispositions contribute to such expectations [6]. In the initial phase of use, users orient themselves in the product. Here the excitement of exploring new functions and frustration with unfamiliar or problematic processes is of particular relevance. In the second phase, the so-called incorporation, the relevance of these initial impressions decreases, and the users can assess the usefulness and the usability of the product. In the last phase the UX of the product influences the relation of the user to a product. For example, loyalty with the product or the brand arises in this phase [4, 7].

Desired product characteristics arise from various influences. Among other things, previous usage experience with products of the same brand contributes to user expectations. Both negative and positive previous experiences in the form of known and desired product characteristics have an influence here. Together with the profile of the users and the intended use of the product, user expectations arise among future users [8]. In addition to previous experiences with

Mensch und Computer 2021, Workshopband, Workshop on Quantitative Messung von User Experience.

https://doi.org/10.18420/muc2021-mci-ws01-363

products of a brand, experiences or reports of competing products also influence expectations towards a product [9].

Another interesting question is how user expectations or anticipated use affect the actual UX rating of a product?

A natural assumption is that users with low expectations can more easily be positively surprised if the UX of the product exceeds their expectations. But high expectations also offer advantages. Particularly in the case of well-known brands, it is important for the reputation of the brand and its products that the target group has high expectations towards the product. However, these expectations must also be met or exceeded during actual use [9].

In [10] it was shown that positive pre-use expectations increased the evaluation of the UX of the product during and after use. In addition, the confirmation of UX expectations increased the overall UX evaluation of the product.

3 A method to evaluate anticipated use

The goal of the method is to capture future users' expectations of a product's UX in a uniform and efficient way and to allow a comparison between the anticipated and the actual level of UX.

3.1 How to capture expectations?

Questionnaires are a frequently used method to measure subjective impressions of users towards the UX of a product. They consist of items and scales (subsets of items) that measure how well a product fulfills UX qualities, for example efficiency, ease of learning, controllability, fun of use or aesthetics.

However, in our case participants of a study have not yet used the product. We want to capture their expectations towards an anticipated use. Thus, classical item formats in questionnaires, for examples statements describing how well an UX aspect is realized to which participants can agree or disagree or semantic differentials will not really work here.

A common technique to prioritize requirements towards products are card sorting methods. Requirements are printed on cards and can be sorted into different categories, for example *important* and *unimportant* requirements. A similar idea is realized by the *Product Reaction Cards* [11], a set of 118 adjectives that describe positive and negative properties of a product, for examples: *clean*, *boring*, *fun*, *secure*, *novel*, *efficient*, *consistent*, *engaging*, or *slow*. Participants of a study choose which adjectives describe a product best.

Our method is based on these ideas. Since we want to capture expectations, negative adjectives make no sense. Thus, the basis for the evaluation of UX expectation is a set of positive adjectives. This list is presented to participants in a study and each participant can decide if the adjective describes an UX property that he or she expects from the product.

Since UX is a quite heterogeneous concept it does not make sense to formulate a set of adjectives that are able to describe all possible products. First, the list will be too long (this is already the case for the Product Reaction Cards) to be used in a practical study. Second, it makes no sense to ask participants to classify adjectives that make no sense for the product under investigation from a semantic point of view. This is also in line with research that shows that the importance of certain UX aspects depends strongly on the product type [2, 3].

Therefore, we follow a modular approach. The researcher defines the set of adjectives that he or she considers in principle as relevant for the product. To provide some assistance we have extracted a set of 58 positive adjectives from an analysis of several UX questionnaires [12, 13, 14, 15, 16; 17, 18] and the already mentioned Product Reaction Cards [11]. The researcher can state which type of product is investigated and per product type a preselection of adjectives is offered. Of course, the researcher can change this suggestion and it is also possible to add own adjectives or to use a complete custom list.

The adjectives are presented to participants before they actually use a product. Each participant can decide if the property described by the adjective is in his or her opinion important or unimportant for the expected user experience.

The results of such a study can be nicely visualized with a word cloud. The cloud shows the adjectives and their font size in the display correspond to the number of participants that considered this adjective as important. This gives a qualitative impression of the overall result and can be used to directly communicate which UX aspects of a product are important respectively unimportant for potential users.

3.2 How to measure actual UX?

Assume we have already captured UX expectation towards a product. Now the product is ready and delivered and we want to answer the natural question if the expectations are fulfilled. Thus, we need a method to measure UX of the product that can be compared to the expectations.

Of course, we can use a classical UX questionnaire format for this purpose. The questionnaire is linked to the adjectives used to capture the expectation. Each item contains the adjective and the participants can rate on a 7-point scale with the end points *does not apply at all* and *fully applies* if they think that the product can be described by this adjective. As an example, we show the first two items of such a questionnaire:

The product is …

valuable

does not apply at all o o o o o o o fully applies

predictable

does not apply at all o o o o o o o fully applies

…

We code the ratings (from left to right) as -3, …, 0, … , 3. Thus, we get a score between -3 and 3 per adjective that reflects how well the product property described by the adjective is realized in the evaluated version of the product.

3.3 Compare anticipated and actual UX?

Assume that A is one of the selected adjectives for the study. Assume further that we have n participants that judge if the adjectives are important (1) or unimportant (-1) for the expected UX. Let a_k be the judgement of the participant k. Then we define the expectation for A as:

$$Expectation(A) = \frac{1}{n}\sum_{k=1}^{n} a_k$$

Assume that we have m participants that judge how well the UX aspect described by adjective A is realized in the product. This is done by a 7-point scale ranging from -3 to 3. Let b_k be the rating of participant k. Then we can define the evaluation of A by:

$$Evaluation(A) = \frac{1}{m * 3}\sum_{k=1}^{m} b_k$$

Both measures are scaled from -1 to 1. The following scheme can thus be used to show how well expectation and evaluation of an adjective match.

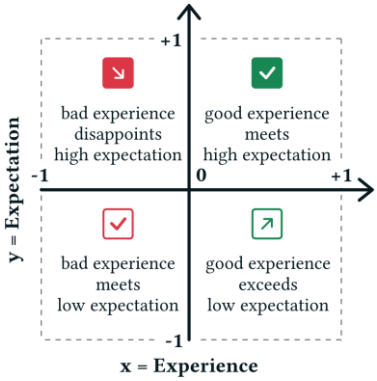

Figure 1: Comparison of UX expectation and actual experience.

3.4 Summary of the method

To sum up, our method consists of the following steps. First, the researcher selects a set of adjectives that describe those UX aspects that are relevant for the investigated product. Suggestions based on the product category are provided, but the researcher is free to remove suggested adjectives and add additional ones. Participants of a study can then select those attributes from this set that they expect from the product. In addition, the set of adjectives is used to create a classical UX questionnaire. Using this questionnaire, it is possible to measure how well the UX aspect described by an adjective is realized by the product. This allows to find out if user expectations are met.

4 Study 1 (Prestudy)

For our evaluation of user expectations, we came up with two different variants. Both variants equip the participants with an introduction to the method, a text which describes the evaluated product and a set of items. Variant 1 asks participants to classify each item into one of two boxes (either *unimportant* or *important* for the product). Variant 2 offers four different boxes to decide from (*unimportant*, *rather unimportant*, *rather important*, *important*).

The goal of our first study was to evaluate which variant provides the clearest results and if there are any general problems when using the method. In this first study we focused only on the evaluation of user expectations.

4.1 Participants

18 persons took part in this first study. Nine of them were students, the others were employees of a software company.

4.2 Setup

For our study we prepared two fictional products, both described in a short introduction text. Product 1 was a voice assistant for IT service requests. Product 2 was a web application that supports consumers to create and order an individually designed (book-)shelf. For each of the products we chose a set of 24 items from the initial item set of the method based on the matching UX aspects for the specific type of product. Besides the products we prepared a short instruction for the method.

All components were placed in an online board where later on the participants performed their evaluation.

Each participant was asked to perform two evaluations: First round with product 1 and variant 1, second round with product 2 and variant 2. To prevent any influences of the order, we interchanged the sequence of the variants for half of the participants.

Figure 2: Screen design used in study 1.

4.3 Results

Regardless of the variant, for the results of the expectation rating we only considered the words classified as “important”. For each item we counted how many participants classified it as “important”. Then we compared the results of the two variants. To decide which of them provides more differentiated results we considered the standard deviation as a measure of scatter across the different values. Hereby variant 2 delivers a higher standard deviation meaning that it better differentiates important from unimportant items. Therefore, we recommend variant 2 to be used in further studies. Furthermore variant 2

will provide more graduations within the expectation rating when comparing the anticipated UX to real UX.

To illustrate the expectations for both products we used word clouds. The following word clouds display the results of variant 2 (the study was done in German, we show the original German items). The bigger an item is being displayed the more participants consider it important for the product.

Figure 3: Word Cloud for best selection strategy (product 1).

Figure 4: Word Cloud for best selection strategy (product 2).

Overall, the participants found it easy to classify the items. Only a few items seemed to require a brief explanation. However, these could also be classified without additional explanation. The comment was raised if it would make sense to provide some explanation of the context of the adjectives. However, depending on the kind of product it would be very difficult to find universal explanations that fit every product.

Ten participants found it easier to classify the items into two levels. The clearer separation and simpler decision, since only an either-or decision had to be made, was emphasized positively. Eight participants, however, preferred the classification with four levels. This was mainly justified by the possibility of an easier gradation of the importance of the terms. Adding the higher deviation of the variant with four levels, this confirmed our recommendation of variant 2.

Many participants noticed that certain items described similar qualities and could be interpreted similarly. These items were mostly classified in the same level.

On average it took the participants 4:45 minutes to complete an evaluation with one of the variants. Since the participants were asked to "think out loud" during this study, it is assumed that the time would be significantly shorter in the independent setting.

5 Webtool to support the method

To support an easy usage of the method, our goal was to develop an associated webtool. This tool was realized with HTML, JS and PHP and can run on any PHP-enabled webserver.

The web tool allows to create, organize, and analyze an evaluation study with our method. After users log in to the tool, they see an overview of all their running evaluation projects. Here they can navigate into running projects, for example to get an overview over incoming data sets or to analyze and download the collected data.

In addition, they can setup a new evaluation. Therefore, they create a new project with name, description and language. The items are currently supported in German and English. In the following step the organizer can select which items should be part of the questionnaires, according to the type of the product (see Figure 5). By default, all items are selected. The organizer can deselect individual items or whole UX aspects with their respective items. Furthermore, it is possible to start with a preselection template that is linked to a specific product type (e.g. word processing, news platform, ...). Those templates were derived from the study of important aspects for different kinds of products in [3]. In addition to the given items the organizer can add individual items to the set.

Next, the organizer can edit all different texts that are displayed in the questionnaires (headline, introduction, label of the boxes / scales, thank you message).

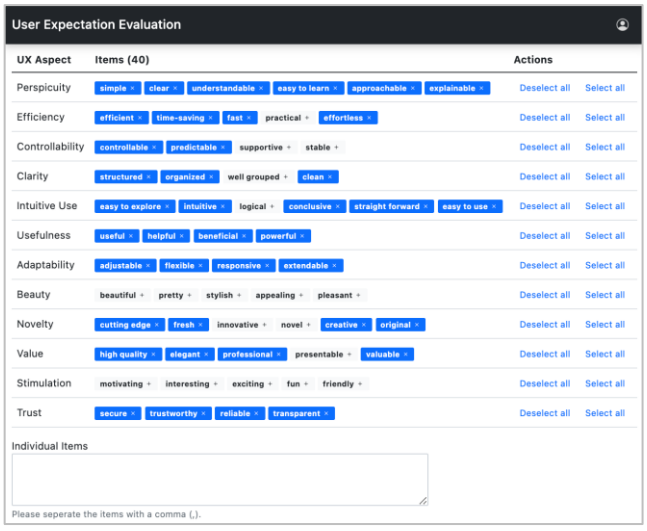

Figure 5: Part of the screen to create a new project. In this area the organizer can choose the items.

As soon as the project is saved, the web tool generates two pages: the page for the expectation query and the page for the experience questionnaire. Both can be accessed via links provided by the web tool.

Now the organizer sends the first link (expectation query) to the participants. The expectation query is displayed on a single page (see Figure 6). A heading and introduction text describes the evaluated product and the evaluation method itself on a high level. Underneath, the page displays the given items and the given classification boxes. The order of the items is randomized. The participants operate the following way: First they imagine the product based on the high-level information provided by the instruction. Based on that assessment they form their expectation. To express those expectations, they classify the given items by the importance for the product. To do so they put each item via drag-and-drop into the matching box (unimportant, rather unimportant, rather important, important). When the participants classified all items, they can submit the evaluation by clicking on the *Submit* button. Afterwards, a thank you message is displayed and asks the participants to close their browser window.

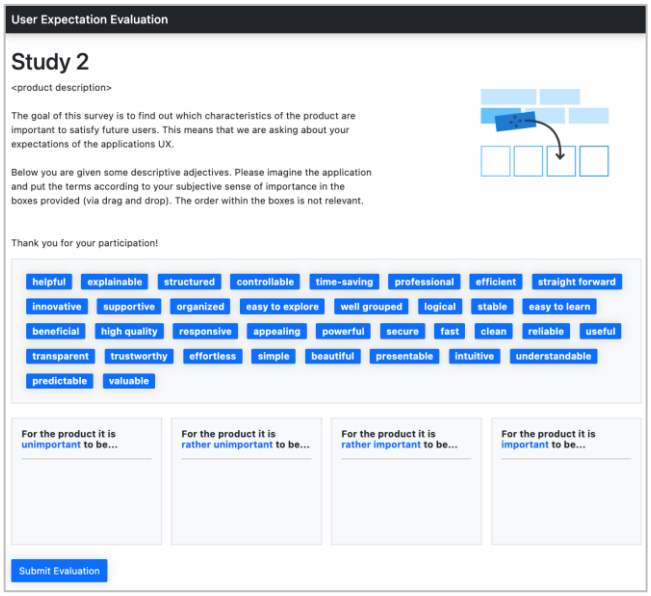

Figure 6: Screen for the evaluation of user expectations generated from a project.

In addition to the expectation query, the organizers can also send out the link to the experience questionnaire. This questionnaire is also displayed on a single page, providing all the different items and a seven-point scale for each item. The endpoints of the scale are by default named as *does not apply at all* and *fully applies*. The items are displayed in a random order.

During and at the end of the evaluation project, the organizer can have a look into the results of the evaluation as can be seen in Figure 7. Here all items are listed in a table providing three different values for each item. The first column provides an absolute value for the expectation / importance of the item. The second and the third column are filled as soon as the UX-questionnaire was used as well. The second column shows an absolute and a scaled value of the expectation-rating. The third column shows the same values for the UX-rating. To give a high-level comparison between expectation and experience the last column indicates the ratio of expectation and the real experience for each item via different icons.

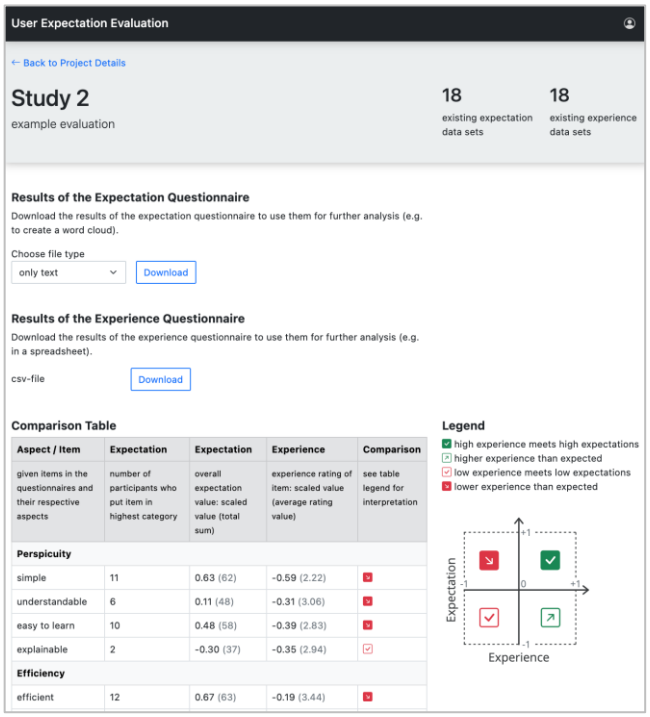

Figure 7: Screen that provides high level results and allows to download the data files.

6 Study 2

The goal of our second study was to use the method on a bigger sample, using both the expectation query and the UX-questionnaire and compare the results. Also, we wanted to test the web tool in a real study setting. Due to limitations in time we had to find a product to evaluate for which we had both future users (expectation) and current users (experience). The tool we evaluated is an internally used self-service tool that is used by employees on an infrequent basis. Students at the company don't use that tool yet and can thus provide their expectations.

6.1 Participants

36 persons took part in the second study. 18 of them provided their expectations, 18 others rated the UX. The sample consisted of both, students and employees. The study was done in English, since this was the official company language.

6.2 Setup

The items for the evaluation were selected based on the type of product. We used the matching UX-aspects for the product type based on the findings of [3], which were perspicuity, efficiency, controllability, clarity, intuitive use, value and trust. We deselected some of the items within those aspects to add additional items from the following UX aspects: usefulness, adaptability, beauty and novelty. In total we provided 36 items.

The invitation to the study was distributed via mail to a group of possible attendees, partly students and employees. The invitation mail provided both links, so the recipients could choose which query they participate in, depending on whether they did or did not use the evaluated application before.

6.3 Results

The word cloud displayed in Figure 8 shows the results of the expectation query. Above all it shows that participants expect an efficient, useful and secure application. Its use should be time-saving and easy to learn. Beauty and originality, on the other hand, are not expected.

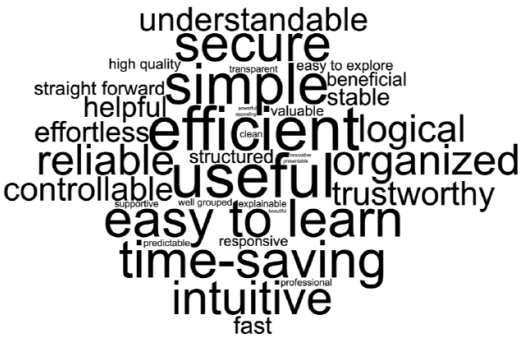

Figure 8: Word cloud visualizing the expected UX.

Next up, we compare the results of the expectations with the real UX ratings of our second group of participants. The results of the comparison are displayed in Figure 9 (a similar way to visualize the dependency between importance and realization of complete questionnaire scales can be found in [19]).

We can see that items that describe beauty and novelty are rather unimportant to the participants and are rated poorly. Efficiency, perspicuity and intuitive use are important to the participants. However, these expectations are disappointed as can be seen in the UX ratings of items of these aspects. Trust in the application is rather important and is also rated as rather good. Aspects of usefulness and adaptability are rated slightly better than expected.

However, analyzing the UX rating, it becomes clear that there are very few strongly positively rated items for the product. The UX is rated as rather poor overall. The average UX rating across all items is 3.39 on a scale of 1 to 7.

Overall, we can conclude that the product cannot meet most of the expectations.

Figure 9: Comparing expected UX with real UX.

7 Summary and outlook

We described a method that allows to capture user expectations concerning UX in a structured way. This method allows a comparison between anticipated UX qualities and their actual realization in the product. Thus, it allows to find out if user expectations are met by the design. We demonstrated the use of the method with a smaller pre-study and a larger study.

In practice our method can be used in the planning or early development phase of a product to evaluate the UX aspects that are especially important to fulfill the expectations of later users. The results can then be used to create a shared understanding of UX priorities between all members of the development team (designers, developers, product owners). Such a common understanding can help to set priorities and to make decisions in case of design conflicts. If, for example, a planned feature increases efficiency but on the same time decreases perspicuity, then a proper understanding of the expectations of the users towards these two UX aspects can help to decide if the feature should be included or not. After the product is finished the expectations can be compared to UX ratings to decide if the product properly supports what users expect.

It is also possible to use the method for already available products. In this case the expectations should be evaluated by a group of persons that do not use the product yet.

Since the method is supported via an easy-to-use web tool it is possible to use it efficiently in projects.

Of course, further studies are required to get a better picture towards possibilities and limitations of the method. In addition,

it must be evaluated if UX practitioners find the results easy to understand and also, even more important, easy to communicate to create a common understanding of the importance of different UX aspects in project teams.

REFERENCES

[1] Sproll, S, Peissner, M. & Sturm, C. (2010). From product concept to user experience: exploreing UX potentials at early product stages. Proceedings: NordiCHI 2010, pp. 473–483. Reykjavik, Iceland.

[2] Schrepp, M. (2021). User Experience Questionnaires: How to use questionnaires to measure the user experience of your products? Amazon KDP. ISBN-13 : 979-8736459766.

[3] Winter, D., Hinderks, A., Schrepp, M. & Thomaschewski, J. (2017). Welche UX Faktoren sind für mein Produkt wichtig? In Steffen Hess & Holger Fischer (Hrsg.), Mensch und Computer 2017 – Usability Professionals, pp. 191–200. Regensburg: Gesellschaft für Informatik e.V. und German UPA e.V. doi:10.18420/muc2017-up-0002

[4] Karapanos, E., Zimmerman, J., Forlizzi, J. & Martens, J.B. (2009). User experience over time: an initial framework. Proceedings of the 27th international conference on Human factors in computing systems -87 CHI 09, pp. 729–738. Boston, MA, USA: ACM Press.

[5] DIN EN ISO 9241-210 2019. Ergonomie der Mensch-System-Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210:2019); Deutsche Fassung EN ISO 9241-210:2019. Berlin.

[6] Hassenzahl, M. & Tractinsky, N. (2006). User experience - A research agenda. Behaviour and Information Technology, 25, pp. 91–97. doi:10.1080/01449290500330331

[7] Thüring, S. & Mahlke, S. (2007). Usability, aesthetics, and emotions in human-technology interaction. International Journal of Psychology, 42(4), pp. 253-264.

[8] Yogasara, T., Popovic, V. & Kraal, B. (2012). Anticipating User eXperience with a Desired Product: The AUX Framework. In David Durling, Pasin Israsena & Juthamas Tangsantikul (Hrsg.), DRS 2012 Bangkok - Research: Uncertainty, Contradiction and Value. pp. 1–14. Bangkok, Thailand: Chulalongkorn University.

[9] Kraft, C. (2012). User experience innovation: user centered design that works. New York, NY, USA: Apress.

[10] Kujala, Sari & Miron-Shatz, Talya 2015. The evolving role of expectations in long-term user experience. Proceedings of the 19th International Academic Mindtrek Conference, pp. 167–174. New York, NY, USA: Association for Computing Machinery. doi:10.1145/2818187.2818271

[11] Benedek, J., & Miner, T. (2002). Measuring Desirability: New methods for evaluating desirability in a usability lab setting. Proceedings of Usability Professionals Association, 2003, 8-12, 57.

[12] Brooke, J. (1996). SUS - A quick and dirty usability scale. In Patrick W. Jordan, Bruce Thomas, Bernard Weerdmeester & Ian L. McClelland (Hrsg.), Usability Evaluation in Industry. London, UK & Bristol, USA: Taylor & Francis.

[13] Laugwitz, B., Held, T. & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. Proceedings of the USAB 2008 (Band 5298, pp. 63–76). Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-89350-9_6

[14] Schrepp, M. & Thomaschewski, J. (2019). Design and Validation of a Framework for the Creation of User Experience Questionnaires. International Journal of Interactive Multimedia and Artificial Intelligence. DOI:10.9781/ijimai.2019.06.006

[15] Hurtienne, J. & Naumann, A. (2010). QUESI- A questionnaire for measuring the subjective consequences of intuitive use. Interdisciplinary College, pp. 536.

[16] Chin, J., Diehl, V, & Norman, K, (1988). Development of an instrument measuring user satisfaction of the human-computer interface. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 213–218. New York, NY, USA: Association for Computing Machinery. doi:10.1145/57167.57203

[17] Hassenzahl, M.; Burmester, M. & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Gerd Szwillus & Jürgen Ziegler (Hrsg.), Mensch & Computer 2003 (Band 57, pp. 187–196). Wiesbaden: Vieweg+Teubner. doi:10.1007/978-3-322-80058-9_19

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

[19] Hinderks, A, Domínguez-Mayo, F. J., Meiners, A. L., & Thomaschewski, J. (2020). Applying Importance-Performance Analysis (IPA) to Interpret the Results of the User Experience Questionnaire (UEQ). Journal of Web Engineering, 19(2), pp. 243-266.