Measurement of user experience

A Spanish Language Version of the User Experience Questionnaire (UEQ)

Maria Rauschenberger
MSP Medien-Systempartner
Oldenburg, Germany
maria.rauschenberger@gmx.de

Dr. Martin Schrepp
SAP AG
Walldorf, Germany
martin.schrepp@sap.com

Dr. Siegfried Olschner
DATEV eG
Nürnberg, Germany
siegfried.olschner@datev.de

Prof. Dr. Jörg Thomaschewski
Hochschule Emden/Leer
Emden, Germany
jt@imut.de

Prof. Dr. Manuel Pérez Cota
University of Vigo
Vigo, Spain
mpcota@uvigo.es

***Abstract* — To improve a product you will need most likely developers, managers and user feedback. Besides the basic software qualities other important properties are usability and user experience for developing a good product. Usability is well known and can be tested with e.g. a usability test or an expert review. In contrast user experience describes the whole impact a product has on the end-user. The timeline goes from before, while and after the use of a product. We present a tool that allows to evaluate the user experience of a product with little effort. We show in addition how this tool can be used for a continuous user experience assessment.**

Keywords: Software Quality; User Experience; Questionnaire; Usability; Test; Development;

I. INTRODUCTION

Is your redesign of the website better than the old version? Has the development effort spent to increase user experience really paid off? If you want to answer such questions you need a quantitative method to measure user experience [1]. An efficient and inexpensive method to do such measurements is the usage of well-constructed and validated questionnaires.

The concept of user experience combines well-known aspects, like efficiency, effectiveness, learnability with additional criteria like aesthetics, joy-of-use or attractiveness. The first group of criteria is often referred as pragmatic quality aspects [2] , while the second group is called hedonic quality aspects. Another often used terminology to distinguish both classes of quality criteria is usability goals versus user experience goals [3]. The dependency of pragmatic and hedonic quality is presented in Figure 1.

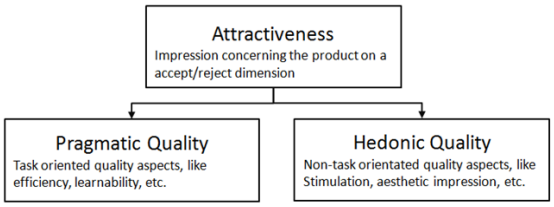

Figure 1: Grouping of different quality attributes.

One well investigated research question is the relationship of pragmatic and hedonic quality. There is empirical evidence that products that are perceived to show a high level of hedonic quality are also perceived as easy to use [4][5]. These and similar observations causes some authors [6] to state that 'What is beautiful is usable'. But in other studies [7] also an opposite dependency could be found. The perception of the aesthetic value of an user interface increased when the number of concrete usability problems decreased. These results indicate that it is necessary to consider both pragmatic and hedonic aspects if we want to measure how satisfied users are with a given product. This idea is underlying the construction of the *User Experience Questionnaire* (short UEQ) that is described in this paper.

This questionnaire allows a quick assessment of the user experience of a website or any other kind of interactive product. The scales of the questionnaire are designed to cover a comprehensive impression of user experience. The format of the questionnaire supports users to immediately express

feelings, impressions, and attitudes that arise when they use a product.

The UEQ is a semantic differential. For such questionnaires it is especially important that users see the items in their native language. So far the UEQ was available in German, English, French and Italian. We present in this paper the Spanish language version of the questionnaire.

We describe in the following how the UEQ was constructed and validated. In addition, the structure of the questionnaire and the meaning of the subscales are explained. We then show how the UEQ is applied for benchmarking in a big business software company. Finally, we describe the creation of the Spanish language version of the UEQ.

II. Construction and Validation of the User Experience Questionnaire (UEQ)

In this section we will provide a rough overview on the construction process of UEQ. Details concerning this process can be found in [8][9].

A. Generation of a pool of potential items

Within two brainstorming sessions with usability experts, 229 terms relevant for the assessment of user experience were proposed. This initial pool of potential items was reduced by experts to 80 adjectives in a multistep reduction process. Since the target format of the questionnaire was a semantic differential the best fitting antonym for each of the 80 adjectives was identified.

An example for an item is:

annoying	O	...	O	enjoyable	1
not understandable	O	...	O	understandable	2

Figure 2: Two examples for items.

Participants can rate a product concerning its position on the dimension spanned by the two adjectives on a 7 point scale.

B. Data collection

In order to examine the specific properties of the adjective pairs concerning the assessment of software products, the eighty items raw-version of the questionnaire was used in six studies focusing on the quality of interactive products, including e. g. a statistics software package, cell phone address book, online-collaboration software, or business software. 153 participants provided complete datasets.

C. Item reduction by factor analysis

A basic theoretic assumption underlying the construction of the questionnaire is that persons perceive several distinct aspects when they evaluate a software product [2]. The perceived attractiveness of the product is assumed to be the result of an averaging process from the perceived quality of the software concerning the relevant aspects in a given usage scenario.

For this reason the item set was split into two subsets. The first subset contains 14 items that represent an emotional reaction on a pure acceptance/rejection dimension. The second subset contains the remaining 66 items from the item pool.

A factor analysis (principal components, varimax rotation) of the first subset of items extracted one factor. This factor explained 60% of the observed variance in the data. This factor is called *Attractiveness*. To represent this factor in the questionnaire we picked the six items with the highest loading on the factor.

A factor analysis (principal components, varimax rotation) of the second subset of items extracted five factors. These five factors explain 53% of the observed variance in the data. We named these factors according to the items that showed the highest factor loadings as *Perspicuity*, *Dependability*, *Efficiency*, *Novelty* and *Stimulation*. Per factor, we chose four items to represent this factor in the questionnaire. All items that were not selected to represent one of these five factors were eliminated.

D. Validation studies for the first version of the questionnaire

The UEQ consists of 26 items assigned to 6 dimensions. The dimension Attractiveness was represented by 6 items all other dimensions by 4 items.

To guarantee that a questionnaire provides high-quality results it must be clarified if the scales are reliable (i.e. the scales are consistent) and valid (i.e. the scales do really measure what they intend to measure). The first version of the UEQ was used in a number of studies under quite different conditions. For example, as part of usability tests with business software applications (11 studies with a total number of 144 participants), or in online surveys (for example, a survey to evaluate the user experience of a social software with 722 participants). A review of all available studies showed that reliability (we used Cronbach's Alpha for an estimation of internal consistency) was sufficiently high.

In addition, the validity of the scales was investigated in a number of studies [8][9][10]. Results indicate good construct validity.

III. Structure of the Questionnaire

The user experience questionnaire contains thus in the final form 6 scales with 26 items in total:

- *Attractiveness*: General impression towards the product. Do users like or dislike the product? This scale is a pure valence dimension.
- *Efficiency:* Is it possible to use the product fast and efficient? Does the user interface looks organized?
- *Perspicuity:* Is it easy to understand how to use the product? Is it easy to get familiar with the product?
- *Dependability:* Does the user feel in control of the interaction? Is the interaction with the product secure and predicable?
- *Stimulation:* Is it interesting and exciting to use the product? Does the feel motivated to further use the product?

- *Novelty:* Is the design of the product innovative and creative? Does the product grab users attention?

The dependency of pragmatic and hedonic quality is presented in Figure 3.

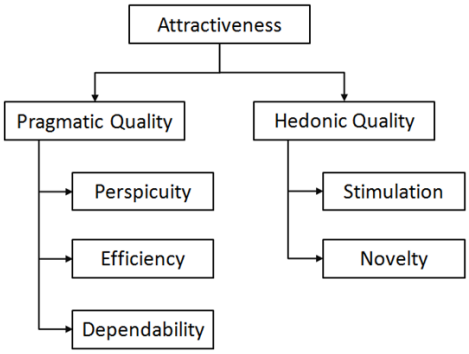

Figure 3: Scale structure of the UEQ questionnaire.

For the concrete questionnaire the order of the items and their orientation (starting with the positive or the antonym statement) is randomized. The concrete English questionnaire is shown in Figure 4.

	1	...	7		
annoying	○	...	○	enjoyable	1
not understandable	○	...	○	understandable	2
creative	○	...	○	dull	3
easy to learn	○	...	○	difficult to learn	4
valuable	○	...	○	inferior	5
boring	○	...	○	exciting	6
not interesting	○	...	○	interesting	7
unpredictable	○	...	○	predictable	8
fast	○	...	○	slow	9
inventive	○	...	○	conventional	10
obstructive	○	...	○	supportive	11
good	○	...	○	bad	12
complicated	○	...	○	easy	13
unlikable	○	...	○	pleasing	14
usual	○	...	○	leading edge	15
unpleasant	○	...	○	pleasant	16
secure	○	...	○	not secure	17
motivating	○	...	○	demotivating	18
meets expectations	○	...	○	does not meet expectations	19
inefficient	○	...	○	efficient	20
clear	○	...	○	confusing	21
impractical	○	...	○	practical	22
organized	○	...	○	cluttered	23
attractive	○	...	○	unattractive	24
friendly	○	...	○	unfriendly	25
conservative	○	...	○	innovative	26

Figure 4: English version of the UEQ

The items are scaled from -3 to +3. Thus, -3 represents the most negative answer, 0 a neutral answer, and +3 the most positive answer. When analyzed the following aspect should be considered. Scale values above +1 indicate a positive impression of the users concerning this scale, values below -1 a negative impression. Due to well-known answer effects, like the avoidance of extremes, observed scales means are in general in the range of -2 to +2. More extreme values are rarely observed, so a value near +2 represents a very positive near optimal impression of participants.

To reduce the effort for data analysis an excel file is created, that does all the necessary calculations. Only the raw data of the questionnaire results have to be entered into the tool. The tool then calculates the scale values, creates a bar chart to visualize the results and calculates some basic statistical indicators necessary for an interpretation of the data, for example confidence intervals for the scales. Figure 5 presents an example of a comparison of two product versions.

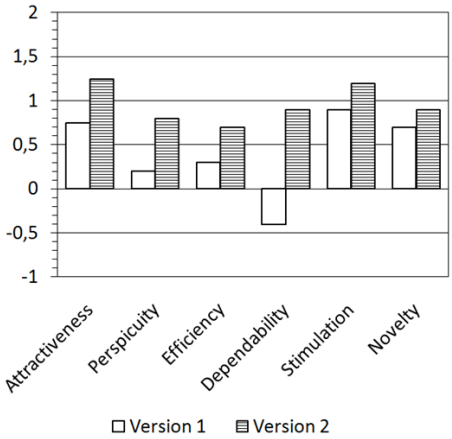

Figure 5: Example of a comparison *of two product versions concerning the UEQ scales.*

The scales can be grouped into three categories. attractiveness is a pure valence dimension. The scales efficiency, perspicuity and dependability describe the pragmatic quality of the product. The scales stimulation and novelty describe the hedonic quality of the product.

IV. Possible Application Scenarios

If a new product is rolled out or if an existing product is evaluated the first time typical questions are *Does the product create a positive user experience?* or *How do users feel about the product?*.

To answer such questions it is sufficient that a representative sample of users of the new product (for example participants of a usability test or pilot users) fills out the UEQ. Between 20 and 30 answers are typically enough to get a valid impression.

Another application is the continuous quality assessment of a software product within a development process [10]. In this approach a measurement with the UEQ is collected with each new version of the software. Thus, we can directly see if new versions bring an improvement in user experience if the scale values for the six scales of the UEQ increase with the new version (for an example on the concrete implementation of such a process, see Laugwitz et. al. 2009).

V. Application in DATEV

This part presents an example how the UEQ is applied for benchmarking in a big business software company.

A. About DATEV eG

The cooperative DATEV eG, Nuremberg (Germany), is a software company and IT service provider for tax consultants, auditors and lawyers as well as their clients. DATEV eG belongs to the main German information service providers and software companies. A huge range of applications allows to satisfy the individual needs of customers. Roughly 5800 employees produce more than 220 applications and provide service for about 39800 cooperative members.

B. Useage of UEQ within a defined User Centered Design Process

The concept of user centered design is meanwhile part of the official DATEV software development model and the UEQ is an integral component among other UCD methods like classical usability testing, focus groups, persona development and heuristic evaluation. The questionnaire is used to get user feedback at different development stages and all UEQ data are collected in one database. The evaluation is conducted with the IBM SPSS Statistics software.

C. Scenarios of use

One major goal is to perform a regular standardized survey with our users in consultant companies and enterprises. The challenge here is the integration into software release plans and market research activities. The UEQ is currently used successfully in three scenarios:

- Evaluation of new beta versions by selected beta testers
- Assessment of released software by randomly selected users
- At the end of a classic usability test to evaluate a new prototype

In the last scenario it is not the primary goal to get an accurate assessment, but the outcome will give an orientation whether the new software design will bring a significant improvement compared to the DATEV benchmark and previous measurements for the tested application. Of course one must be cautious, the tasks in a laboratory test do not represent the entire application and the demonstrated improvements in some parts will perhaps have no effect on the overall user experience of the complete application.

A current project is the test of the combination of online questionnaire and focus group. The outcome of the online-UEQ should be the base for questions in asynchronous online focus groups. Another example how to use the UEQ is described in an article concerning user experience for business software [11].

D. Additional DATEV specific Items

The quality of business software for tax consultants and auditors is not only determined by the UEQ factors. In addition to the scales attractiveness, efficiency, perspicuity, etc. the perceived quality is largely determined by the actuality of the application. Especially the integration of the latest tax laws in the program calculations is very important.

Based on this, the DATEV user experience design team made a few additions and some adoption of scale terms to fit management needs. In particular two additional single items are relevant. The question for "Overall satisfaction" and the question whether the software is “Up to date” regarding the latest changes in legislation texts. The two questions provide the opportunity to explain negative judgments. These reasons for the numeric ratings are very helpful during the interpretation of the UEQ values. Therefore, we follow the plan

to collect more free textual feedback by these additional questions.

Other additions are questions to control the sample (age, gender, occupation, etc.), questions about performance and questions about customer loyalty.

This in turn is important for one of our next projects, the use of the UEQ in combination with focus groups.

E. Acceptance

To achieve a high user acceptance of the UEQ you should take following points into account:

- Background and benefits of the method should be clear to the user
- A personal contact should be available for the responders
- The time interval between repeated measurements should be long enough

To achieve acceptance by managers and product managers consider to:

- Provide help during the interpretation of the UEQ factor values. Define your range of good, medium and bad and explain the theoretical background
- Combine old and new UEQ values in one picture and show the changes during the development in order to increase the intelligibility of the measurements
- Search for other user feedback that supports the interpretation of the UEQ outcome and integrate it into your report
- Adjust your wording to your company terminology
- Enhance the UEQ results with concrete enhancement suggestions based on user experience expertise and use this as a base for further discussions about the next development goals

F. Benefits of the process

In the long term the regular use of the questionnaire offers the possibility to find answers to important questions that come up during the user centered design process. In this context, you should be aware that the UEQ should only be a part within a mix of methods, but perhaps you will be able to answer following questions:

- Is your work effective? Regularly collected data will allow you to run pre-post-comparisons
- What are the strengths and weaknesses of the individual products? Perhaps one development team can learn from another team
- How will the user evaluate the final version? With a large data base you can calculate UEQ-factor correction values and you will be able to predict the ratings based on the feedback provided by the beta testers
- What do your users need? Determine the significance of individual UEQ-factors for different user groups

VI. Creation of a Spanish Language Version

First the German version of the UEQ was translated into Spanish by a native speaker and a bilingual person. After that the Spanish version had been retranslated into German. If the words turned out to match the original words the translation was declared to be successful. Otherwise the process was repeated until all words matched. To demand a one-to-one translation from one language into another is not entirely possible. The reason for that are the different meanings of one word, which make it difficult to find synonym in any language.

The translator was open minded and didn't know the questionnaire before. If you want to use the Spanish language version of the UEQ please contact one of the authors above. They can provide you with a paper version of the questionnaire together with a standard instructional text and a version of the Excel tool used to compute the results.

	1	...	7		
desagradable	○	...	○	agradable	1
no entendible	○	...	○	entendible	2
creativo	○	...	○	sin imaginación	3
fácil de aprender	○	...	○	difícil de aprender	4
valioso	○	...	○	de poco valor	5
aburrido	○	...	○	emocionante	6
no interesante	○	...	○	interesante	7
impredecible	○	...	○	predecible	8
rápido	○	...	○	lento	9
original	○	...	○	convencional	10
obstructivo	○	...	○	impulsor de apoyo	11
bueno	○	...	○	malo	12
complicado	○	...	○	fácil	13
repeler	○	...	○	atraer	14
convencional	○	...	○	novedoso	15
incómodo	○	...	○	cómodo	16
seguro	○	...	○	inseguro	17
activante	○	...	○	adormecedor	18
cubre expectativas	○	...	○	no cubre expectativas	19
ineficiente	○	...	○	eficiente	20
claro	○	...	○	confuso	21
no pragmático	○	...	○	pragmático	22
ordenado	○	...	○	sobrecargado	23
atractivo	○	...	○	feo	24
simpático	○	...	○	antipático	25
conservador	○	...	○	innovador	26

Figure 6: Spanish version of the UEQ

VII. Acknowledgements

We would like to acknowledge support of this work by Andrea Meyenburg and Markus Aulkemeier.

VIII. References

[1] DIN EN 9241_210, 2011-01, Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems, Berlin:Beuth.

[2] Hassenzahl, M., "The effect of perceived hedonic quality on product appealingness", International Journal of Human-Computer Interaction, 13, pp. 481-499, 2001.

[3] Preece, J., Rogers, Y., Sharp, H., "Interaction design: Beyond human-computer interaction" Wiley, New York , 2002.

[4] Kurosu, M., Kashimura, K., "Apparent usability vs. inherent usability: experimental analysis of the determinants of the apparent usability". In: Conference Companion of human factors in computing systems, pp. 292--293. Denver, Colorado, 1995.

[5] Tractinsky, N., "Aesthetics and Apparent Usability: Empirical Assessing Cultural and Methodological Issues". CHI'97 Electronic Publications http://www.acm.org/sigchi/chi97/proceedings/paper/nt.htm 1997.

[6] Tractinsky, N., Katz, A.S., Ikar, D., "What is beautiful is usable", Interacting with Computers, 13, pp.127--145, 2000.

[7] Ilmberger, W.; Schrepp, M. & Held, T., "*What kind of cognitive process causes the relationship between aesthetics and usability*". In: Holzinger, A. (Ed.): USAB 2008, LNCS 5298, pp. 43-54, 2008.

[8] Laugwitz, B.; Held, T. & Schrepp, M. (2008). *Construction and evaluation of a user experience questionnaire.* In: Holzinger, A. (Ed.): USAB 2008, LNCS 5298, pp. 63-76.

[9] Laugwitz, B.; Schrepp, M. & Held, T. "Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten [Construction of a questionnaire to measure the user experience of software products] ". In: A.M. Heinecke & H. Paul (Eds.): Mensch & Computer 2006 – Mensch und Computer im Strukturwandel. Oldenbourg Verlag, pp. 125 – 134, 2006.

[10] Laugwitz, B.; Schubert, U.; Ilmberger, W.; Tamm, N.; Held, T. & Schrepp, M., "Subjektive Benutzerzufriedenheit quantitativ erfassen: Erfahrungen mit dem User Experience Questionnaire UEQ [Quantitative measurement of subjective user satisfaction: Experiences with the user experience questionnaire UEQ]". In: Brau, H.; Diefenbach, S.; Hassenzahl, M.; Kohler, K.; Koller, F.; Peissner, M.; Petrovic, K.; Thielsch, M.;Ulrich, D. & Zimmermann, D. (Eds.); Usability Professionals 2009, pp. 220 – 225, 2009.

[11] Rauschenberger, M.; Hinderks, A.; Thomaschewski, T. (2011). Benutzererlebnis bei Unternehmenssoftware: Ein Praxisbericht über die Umsetzung attraktiver Unternehmenssoftware In: Brau, H; Lehmann, A.; Petrovic, K.; Schroeder, M.C. (Eds.); Usability Professionals 2011, pp. 158 - 163, 2011.