Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version

Maria Rauschenberger, MSP Medien-Systempartner, Germany; Martin Schrepp, SAP AG, Germany
Manuel Pérez Cota, University of Vigo, Spain; Siegfried Olschner, DATEV eG, Germany
Jörg Thomaschewski, University of Applied Science Emden/Leer, Germany

Abstract — Developer, manager and user feedback is needed to optimize products. Besides the basic Software qualities – usability and user experience are important properties for improving your 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 you to evaluate the user experience of a product with little effort. Furthermore the tool is available in different languages and we are using the new Spanish Version. We show 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 rigorously constructed and validated questionnaires.

The concept of user experience combines well-known aspects like efficiency and effectiveness 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 Fig. 1.

One well investigated research question is the relationship of pragmatic and hedonic quality. Empirical evidence proves that products, which are perceived to show a high level of hedonic quality, are also perceived as easy to use [4], [5], [6]. These and similar observations cause some authors [7] to state that ‘What is beautiful is usable’. In contrast other studies point out [8], [9] an opposite dependency. The perception of the aesthetic value of a user interface increased when the number of concrete usability problems decreased. Thus, in this study a ‘What is usable is beautiful’ effect was observed.

Why are perceived hedonic and pragmatic quality aspects associated? As possible explanation for this connection halo-effects [10], mediation by the mood of the user [11] or mediation by other variables [6] have been suggested. Since it is quite difficult to separate these effects experimentally [8] it is currently unclear which of these hypotheses are able to explain this effect.

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 is the underlined idea of constructing the User Experience Questionnaire (UEQ) [12], [13] that is described in this paper. In the context of the questionnaire user experience is understood as the overall impression of a user when he or she interacts with a product, i.e. covers both pragmatic and hedonic quality aspects.

The UEQ allows a quick assessment of the user experience for any interactive product. The scales of the questionnaire are designed to cover a comprehensive impression of user experience. The questionnaire format supports the user response to immediately express feelings, impressions, and attitudes that arise when they use a product.

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 fill out the UEQ. 30 answers are usually enough to get a valid impression. For example, the answers can come from participants of a usability test or pilot users.
Another application is the continuous quality assessment of a software product within a development process [14]. 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 [14]). An application of the UEQ in the process of idea and innovation management is described in [15].

User experience is not only a snapshot of the present usage of a product has. It is an entire impression a product creates on the user. Even more, the user’s judgement starts before touching and using a new product. In addition the change of impression carries on during and after the usage [1]. The UEQ is able to present the distinct results over time for the result analysis.

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 should be applied in a company and how the results can be analyzed. Besides, the DATEV eG a big business software company is presenting their design process with the UEQ. Finally, we describe the creation of the Spanish language version of the UEQ.

II. CONSTRUCTION AND VALIDATION OF THE USER EXPERIENCE QUESTIONNAIRE (UEQ)

The items and scales of the UEQ were created by a data analytical approach. First, a set of 229 potential items was built as a result of several brainstorming sessions with usability experts. Second, this set was reduced to an 80 items raw version by an expert evaluation. Third, the eighty items raw version of the questionnaire was used in several 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. In total the data of 153 participants were collected for the initial data set. Finally, the scales and the items representing each scale were extracted from the data by factor analysis (principal components, varimax rotation). Six factors resulted from this analysis. Details concerning the process can be found in [12], [13].

The reliability (i.e. the scales are consistant) and validity (i.e. the scales do really measure what they intend to measure) of the UEQ scales was investigated in a number of studies [12], [13], [14]. Results indicate good construct validity.

III. STRUCTURE OF THE QUESTIONNAIRE

The user experience questionnaire contains 6 scales with 26 items in total:

1) Attractiveness: General impression towards the product. Do users like or dislike the product? This scale is a pure valence dimension. Items: annoying / enjoyable, good / bad, unlikable / pleasing, unpleasant / pleasant, attractive / unattractive, friendly / unfriendly

2) Efficiency: Is it possible to use the product fast and efficient? Does the user interface looks organized? Items: fast / slow, inefficient / efficient, impractical / practical, organized / cluttered

3) Perspicuity: Is it easy to understand how to use the product? Is it easy to get familiar with the product? Items: not understandable / understandable, easy to learn / difficult to learn, complicated / easy, clear / confusing

4) Dependability: Does the user feel in control of the interaction? Is the interaction with the product secure and predictable? Items: unpredictable / predictable, obstructive / supportive, secure / not secure, meets expectations / does not meet expectations

5) Stimulation: Is it interesting and exciting to use the product? Does the user feel motivated to further use the product? Items: valuable / inferior, boring / exiting, not interesting / interesting, motivating / demotivating

6) Novelty: Is the design of the product innovative and creative? Does the product grab users attention? Items: creative / dull, inventive / conventional, usual / leading edge, conservative / innovative

The dependency of the UEQ scale is presented in Fig. 2.

Fig. 2. Scale structure of the UEQ questionnaire.

For the specific questionnaire the order of the items and their orientation (starting with the positive or the antonym statement) is randomized. The specific English questionnaire is shown in Fig. 3 and the Spanish questionnaire is shown in Fig. 7.
IV. HOW TO APPLY THE QUESTIONNAIRE

For a successful application of the UEQ the acceptance of following two groups are needed: Users and Managers. 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 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
- 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

V. ANALYZING RESULT

After collecting the answers from the users a three step analysis as presented can follow. To reduce the effort for data analysis an MS Excel file is created, doing 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. Fig. 4 presents an example of a result and Fig. 45 shows an example of a comparison of two product versions.

A. Verifying the validation

The first step is to confirm the Cronbach’s Alpha data, which describes the consistency of the items of the scales (i.e. if all items in the scale measure the same quality). It is calculated automatically for each study in the excel sheet which can be downloaded from www.ueq-online.org.

If the Alpha value for a scale is small this is an indication that some of the items in this scale are possibly misinterpreted or interpreted in a direction that does not reflect their intention in the context of the UEQ. In this case it is questionable if this specific scale can be interpreted for the final result.

There are two well-known effects that can cause a small value of the Alpha-Coefficient for a scale. First, it is possible that the context in which the questionnaire is applied yields to a misinterpretation of some items in the scale. For example, in a study with informatics students the item ‘secure/not secure’ was referred from the users to the security (i.e. absence of malware or spyware) of the web-service and not to the dependability of the interaction.

Second, a scale may be simply irrelevant in the context in which the questionnaire is applied. Thus, the participants may have problems to interpret the items of the scale properly, which lowers the correlations between the items of the scale and thus decreases the Alpha-Coefficient.

If the alpha coefficient is higher or equal than 0,7 the scales show high consistency, i.e. all items in a scale measure the same aspect and it is unlikely that one of the items is misinterpreted in the given context.

But it can also happen that all items in a scale are influenced by a context specific effect, i.e. one of the scales differs highly from the other scales due to a special target group.

In a study with 20 participants the scale novelty had low results caused by a target group with different age. The VoIP-Software Skype was evaluated. The younger group had no enthusiasm about the technology, because they had known it for a long time. It was not exciting anymore. Elsewise the older group did not know Skype or any similar product. It was their first contact with this technology and they found it very fascinating. The consequence was that one group perceived Skype very stimulating and the other not.
After exanimating the Alpha value next step is the interpreting of the overall result as described in Chapter B.

**B. Interpretate the overall result**

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.

Fig. 4 shows an example for an overall result for a product. The graphic is automatically generated by the data analysis sheet (Excel) that can be downloaded together with the questionnaire.

![Example of an overview result](image)

Thus, this particular product created a slightly positive impression concerning Attractiveness and Stimulation, but is judged neutral concerning the other 4 scales. The error bars represent the 5% confidence intervals for the scale means, i.e. the probability that the true value of the scale mean lies outside this interval is less than 5%. The width of the error bars depend on the number of respondents and on the level of agreement between the respondents. Thus, the more the participants that filled out the questionnaire agree concerning their evaluation of the product the smaller are typically the width of the error bars. Thus, if there are many respondents to the questionnaire and the error bars are still wide, this can be an indication that there are different sub-groups of participants with quite opposite options about the product.

Two different products or product versions can thus easily be compared concerning their user experience by comparing the scale means. See Fig. 5 for a comparison of two product versions concerning the observed scale means.

![Comparison of two product versions](image)

In this example version 2 is much better concerning Attractiveness, Perspicuity Efficiency and Dependability. Concerning the hedonic scales Stimulation and Novelty both versions seems to be comparable.

To find out if the difference concerning the scale values is significant on the 5% level (or any other level you choose) it is necessary to apply a statistical test that compares the scale means (for ex. a t-test). It is not sufficient to check if the error bars do not overlap. If they do not overlap it can be concluded that the difference is significant at 5% level. But the opposite is not true. The error bars can overlap and the difference may still be significant!

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.

**C. Analyzing the results of the individual items**

After the overview the details have to be examined. First if you have two software versions with the UEQ results the items results are placed opposite each other. Items with extreme differences give a hint which areas have been improved or not. These way product versions can be compared easily and exact with one another. Also the detail analyzing shows, which areas should be improved for the next release (See Fig. 6). If it is the first product release see if some items show extreme results compared to other in the same UEQ results.

While analyzing each item the target group could give hints about what caused the significant distinction. Therefore the basic demographic data has to be collected with the UEQ results as well.

The UEQ exists in different languages which are tested reliably. Nevertheless, because of the complexity of language, it is also possible that translation deviance the results.
VI. APPLICATION IN THE COMPANY DATEV eG

This part presents an example how the UEQ is applied for benchmarking in a big business software company. A general impression of a process is presented in [16].

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. Roughly 5800 employees produce more than 220 applications and provide service for about 39800 cooperative members.

B. Usage of UEQ within a defined Design Process

The concept of user centered design is meanwhile part of the official DATEV eG 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.

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 eG 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 [16].

Because of the special form of the UEQ it is important that participants fill out the questionnaire in their natural language. Thus, it is for companies that use the UEQ on multi-national level important to have language versions of the questionnaire available.

VII. 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. For more information see [17].

![Spanish version of the UEQ](image-url)

---

Fig. 6. Example for the detail analyzing of the results from the UEQ-Excel-Sheet (a specimen of the first three items)

Fig. 7. Spanish version of the UEQ
The Spanish Language version of the UEQ is already validated in two bigger studies.

In the first study 94 students evaluated the user experience of the Amazon web-shop (www.amazon.es). The scale means and confidence intervals are shown in Fig. 8.

Fig. 8. Result for Amazon web-shop.

Thus, overall the participants had a slightly positive or neutral impression concerning the user experience of the Amazon web-shop. The impression concerning the pragmatic quality (Perspicuity, Efficiency and Dependability) is clearly higher than the impression concerning the hedonic quality (Stimulation, Novelty).

An analysis of the Cronbach Alpha coefficient showed that the single scales showed high consistency values (Attractiveness: 0.85, Perspicuity: 0.59, Efficiency: 0.74, Dependability: 0.48, Stimulation: 0.75, Novelty: 0.64). This is an indicator that the scales are sufficiently consistent.

In a second study 95 students evaluated the user experience of Skype. Again scale means and confidence intervals are shown in Fig. 9.

Fig. 9. Result for Skype.

The impression concerning the Skype user experience is quite positive. Again pragmatic quality is judged better than hedonic quality aspects. If we compare these evaluations to the results for the Amazon web-shop we clearly see that Skype creates a better user experience.

As in the first study alpha coefficient for the scales shows high values (Attractiveness: 0.83, Perspicuity: 0.71, Efficiency: 0.72, Dependability: 0.55, Stimulation: 0.78, Novelty: 0.71) again indicating sufficient scale consistency.

Of course further studies are necessary to finally judge if the psychometric properties of the Spanish version are identical to the existing and well-evaluated German and English version. But these first results are positive.

IX. AVAILABILITY

The UEQ questionnaire can be used free of charge. The questionnaire itself, a data analysis tool and literature describing the construction of the questionnaire can be downloaded from www.ueq-online.org. The questionnaire and the analysis tool are available in several languages. Currently German, English, French, Italian and the Spanish version are available. It is worked on a Portuguese Version as well.

X. SUMMARY

We described the construction, the result analyzing and the validation studies of the Spanish language version of the User Experience Questionnaire. This questionnaire allows a fast evaluation of the user experience of interactive products. It measures not only usability aspects like efficiency, perspicuity and dependability, but also user experience aspects like stimulation or originality.

Since the UEQ has the form of a semantic differential, it is quite important that participants can rate a product in their natural language. Thus, the new language version allows the application of the UEQ in Spanish speaking target groups.

The first available validation studies suggest that the scale quality of the Spanish version is sufficient to apply the questionnaire in projects to collect feedback about user impressions.

REFERENCES

Prof. Dr. Manuel Pérez Cota is Professor and Researcher at the state-owned university Universidad de Vigo (UVIGO) in Vigo, Spain. He is graduated with honors in Electrical Engineering (Universidad La Salle) and Electronics and Communications Engineering (Universidad Nacional Autónoma de México – UNAM, 1980), a PhD with honors in Industrial Engineering (Universidad de Santiago de Compostela, 1990). He is the director of the international research group SI1-GEAC (http://cuautla.uvigo.es/si1-geac/). He was the first director and developer of the Informatics (Computer Science School of the University of Vigo) and he was director, also, of the Informatics Department. He widely collaborates in different Master and PhD programs in Spain, Portugal, Germany, Argentina and Bolivia, and he has been supervisor in several PhDs, with some others in progress. He is part in different European and International projects. He has published quite extensively and has a lot of publications (including books, book chapters, Scientific Citation Index journal articles, and international journal articles, as well as publications in refereed conference proceedings). He is member of different international committees and associations (ACM, IEEE, AISTI, AIPO, ANALCT).

Prof. Dr. Jörg Thomaschewski. He became full professor at the University of Applied Sciences Emden/Leer, Germany from September 2000. He received the Doctor degree in Physics from the University of Oldenburg, Germany, in 1996. His research interests are in the fields Internet Applications focusing on human computer interaction, e-learning and software engineering. He is author of various online modules, e.g. “Human Computer Communication”, which is used in the Virtual University (Online) at six university sites. He has wide experience in usability training, analysis and consulting.

Dr. Siegfried Olschner is a senior member of the User Experience and Design team of DATEV eG, Nuremberg, Germany. His principal activities are dialogue design, workflow design and general user experience consulting for colleagues from the development departments. He also deals with the research and development of standardized test methods that are used as parts within the DATEV User Centered Design Process. One example is the improvement of the UEQ.

Maria Rauschenberger has a Bachelor of Engineering in Media technology and is studying the Master of Science in Media Informatics from the University of Applied Science Emden/Leer. Furthermore she is working as a Project manager for MSP Medien-Systempreter, Germany. Currently, she pursues the finals of Master degree and taking part of the regional user experience group Bremen (UPA). She is involved in research activities dealing with interaction design, process optimising, usability and user experience since 2011.

Dr. Martin Schrepp works since 1994 as user interface designer for SAP AG. He finished in 1990 his studies in Mathematics with a Diploma from the University of Heidelberg (Germany). In addition, he received in 1993 a PhD in Psychology (also from the University of Heidelberg). His research interests are the application of psychological theories to improve the design of software interfaces, the application of Design for All principles to increase accessibility of business software, measurement of usability and user experience and the development of general methods for data analysis. He published several papers in these research fields.