Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios

Martin Schrepp[1], Andreas Hinderks[2], and Jörg Thomaschewski[3]

[1] SAP AG, Walldorf, Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
[2] RMT Soft GmbH & Co. KG, Carl-Zeiss-Str. 14, 28816 Stuhr, Germany
[3] Hochschule Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany
martin.schrepp@sap.com, andreas@hinderks.org,
joerg.thomaschewski@hs-emden-leer.de

Abstract. A good user experience is central for the success of interactive products. To improve products concerning these quality aspects it is thus also important to be able to measure user experience in an efficient and reliable way. But measuring user experience is not an end in itself. Several different questions can be the reason behind the wish to measure the user experience of a product quantitatively. We discuss several typical questions associated with the measurement of user experience and we show how these questions can be answered with a questionnaire with relatively low effort. In this paper the user experience questionnaire UEQ is used, but the general approach may be transferred to other questionnaires as well.

Keywords: User Experience, Usability, Questionnaire, Pragmatic Quality, Hedonic Quality.

1 Introduction

To create successful products or services it is necessary to ensure that the product has a sufficiently high user experience. Different users or different groups of users may judge the same product quite differently concerning its user experience, for example because they have different needs or different abilities or skills to use the product. An efficient and inexpensive method to conduct such measurements is thus the usage of validated questionnaires. But before such a complex multi-dimensional construct like user experience can be measured in a meaningful way, it is very useful to clearly understand the meaning of the concept.

A well-known definition of user experience is given in ISO 9241-210 [1]. Here, user experience is defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service". Thus, user experience is seen as a holistic concept that includes all types of emotional, cognitive or physical reactions concerning the concrete or even only the assumed usage of a product. This is a quite general and abstract definition that is not helpful at all if we want to get an idea on how to measure this quality aspect of a product.

A different interpretation (which we adopt in this paper) is to define user experience as a set of distinct quality criteria [2] that includes classical usability criteria, like efficiency, controllability or learnability, and non-goal directed or hedonic quality criteria [3], like stimulation, fun-of-use, novelty, emotions [4] or aesthetics [5]. This has the advantage that it splits the general notion of user experience into a number of simple quality criteria, which describe distinct and relatively well-defined aspects of user experience that can be measured independently.

The measurement of user experience is not an end in itself. In fact, several different and quite natural questions can be the reason behind the wish to measure the user experience of a product quantitatively:

- *Continuous improvement by measuring the user experience of new versions:* Has the redesign of the product improved user experience compared to the previous product version? This question can be answered relatively simple by a statistical comparison of two measurements.
- *Comparison to the direct competitors in the market*: How good is the user experience of the product compared to the direct competitors in the market? This is similar to the question above, since here only the direct competitors, i.e. a special group of products, are of interest for a comparison.
- *Test if a product has sufficient user experience*: Does the product fulfill the general expectations of users concerning user experience? Such general expectations of users are formed by their usage of products that they frequently use. To answer this question it is thus necessary to compare the measured user experience of the product to results of other established products, for example from a benchmark data set containing quite different typical products.
- *Determine areas of improvement*: What should be changed in order to improve the user experience of the product? This question cannot be answered directly by a quantitative measurement of user experience. To answer this question, a connection of product features to the measurement is required.

We will discuss these different facets of user experience measurement using the example of the user experience questionnaire (UEQ) [6, 7].

2   Construction of the User Experience Questionnaire (UEQ)

The main goal of the UEQ is to allow a fast and immediate measurement of user experience. The UEQ considers aspects of pragmatic and hedonic quality [6, 7].

The original German version of the UEQ was created 2005 by a data analytical approach in order to ensure a practical relevance of the constructed scales, which correspond to distinct quality aspects. An initial item set of 229 potential items related to user experience was created in brainstorming sessions with usability experts. This item set was then reduced to an 80 items raw version of the questionnaire by an expert evaluation.

The 80 items raw version was used in several studies focusing on the quality of interactive products, including e.g. a statistics software package, cell phone address

books, online-collaboration software or business software. In these studies 153 participants answered the 80 items. Finally, the scales and the items representing each scale were extracted from this data set by principal component analysis [6, 7].

The reliability (i.e. the consistency of the scales) and validity (i.e. the scales really measure what they intend to measure) of the UEQ scales was investigated in 11 usability tests with a total number of 144 participants and an online survey with 722 participants. The results of these studies showed a sufficiently high reliability of the scales (measured by Cronbach's Alpha). In addition, a number of studies [7, 8], showed a good construct validity of the scales.

The user experience questionnaire contains 6 scales with 26 items:

- Attractiveness: Overall impression of the product. Do users like or dislike is? Items: annoying / enjoyable, good / bad, unlikable / pleasing, unpleasant / pleasant, attractive / unattractive, friendly / unfriendly.
- Perspicuity: Is it easy to get familiar with the product? Items: not understandable / understandable, easy to learn / difficult to learn, complicated / easy, clear / confusing.
- Efficiency: Can users solve their tasks without unnecessary effort? Items: fast / slow, inefficient / efficient, impractical / practical, organized / cluttered.
- Dependability: Does the user feel in control of the interaction? Items: unpredictable / predictable, obstructive / supportive, secure / not secure, meets expectations / does not meet expectations.
- Stimulation: Is it exciting and motivating to use the product? Items: valuable / inferior, boring / exiting, not interesting / interesting, motivating / demotivating.
- Novelty: Is the product innovative and creative? Items: creative / dull, inventive / conventional, usual / leading edge, conservative / innovative.

Attractiveness is a pure valence dimension. Perspicuity, Efficiency and Dependability are pragmatic quality aspects (goal-directed), while Stimulation and Novelty are hedonic quality aspects (not goal-directed). Figure 1 shows the assumed scale structure of the UEQ.
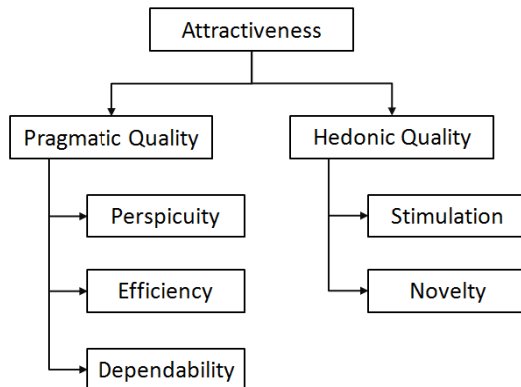

Fig. 1. Assumed scale structure of the UEQ

The questionnaire together with some information concerning its application and an Excel-Tool for data analysis is available free of charge under www.ueq-online.org. For semantic differentials like the UEQ it is of course important that participants read the items in their natural language. Thus, several language versions were constructed and validated (for example, English, Spanish [9], Portuguese [10], etc.). For German there is also a version for children and teenagers available [13] that uses a simplified language. These versions are also available under www.ueq-online.org [12].

Applying the UEQ does not require much effort. Usually 3-5 minutes [11] are sufficient for a participant to read the instruction and to complete the questionnaire. Analyzing the data can be done quite efficiently with the provided Excel-sheet.

3   Continuous Improvement by Measuring the User Experience of New Versions

Most software products undergo a number of redesigns during their lifetime. Typically a more or less complete first version is delivered and then refined based on customer feedback in a number of release cycles. A quite natural question is if the user experience of a revised version is better or at least comparable (for example, if the new version offers more functions and is more complex) to the previous version.

With a questionnaire like the UEQ it is quite simple to answer such questions. All one needs to do is to collect data from a representative sample of users and to compare both versions concerning the single scale means.

Figure 2 shows the results of the UEQ for two versions (a newer and an older version containing the same business functionality) of a business software product. For both versions, participants of a usability test completed the UEQ after they had finished their tasks in the test (20 participants for new version A and 19 participants for old version B).
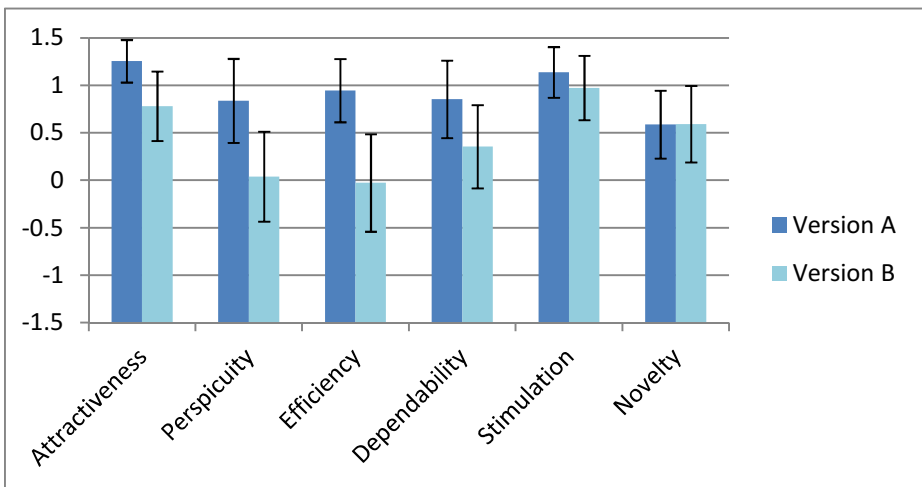

Fig. 2. UEQ result for two product versions. Error bars represent the 95% confidence intervals

Is seems that version A performs better than version B for all scales except novelty. But is this difference significant or only a more or less random deviation? Especially with smaller samples it is absolutely necessary to check if the observed differences are also statistically significant. If the scale mean of version A is higher than the corresponding scale mean of version B and the error bars do not overlap it is immediately clear that version A shows a significantly better result. However, the reverse statement is not true, so even if the error bars overlap (as in this example), the difference can still be significant − in this cases it is necessary to perform a classical significance test. For the example above such a test shows that the differences are despite the small sample size significant for Attractiveness and the pragmatic scales Perspicuity, Efficiency and Dependability at the 5% level.

Based on this simple possibility to compare two product versions it is straightforward to establish a continuous monitoring of user experience for a product. An example of an implementation of such a process is described in [11]. The availability of a quantitative measure for user experience also helps to define clear goals concerning the expected user experience of new or refined products.

4   Comparison to the Direct Competitors in the Market

Often is it not only the goal to be good, but to be better than the direct competitors in the market. The question if a new product outperforms competition with respect to user experience is related to the previous question. The only problem here is to collect data concerning the user experience of competitor products. With classical on premise software this is in most cases impossible due to practical problems to access users of the competitor products. For modern web-based applications this if often much simpler, since in many cases at least product demos are available on the web.

As an example we show a first evaluation of the currently available services for web automation. The three investigated services are *IFTTT* (www.ifttt.com), *Zapier* (www.zapier.com) and *We Wired Web* (www.wewiredweb.com). The basic function of these products is to connect different web services by user defined rules. It is, for example, possible to store a photo as file in a Dropbox when it is posted on Facebook. A rule is defined by a trigger associated with a channel (Facebook) and an action (store the photo in the Dropbox) that is fired when the trigger is activated (photo is posted).

82 students of the University of Applied Sciences Emden/Leer evaluated the services as part of a practical task with the UEQ. Each student had to use one of these web services to solve three different problems with the service, i.e. had to define three different rules. This forces the students to get familiar with the service and to get a realistic impression concerning its user experience. After this phase each student had to evaluate the service he or she used with the simplified German version [14] of the UEQ.

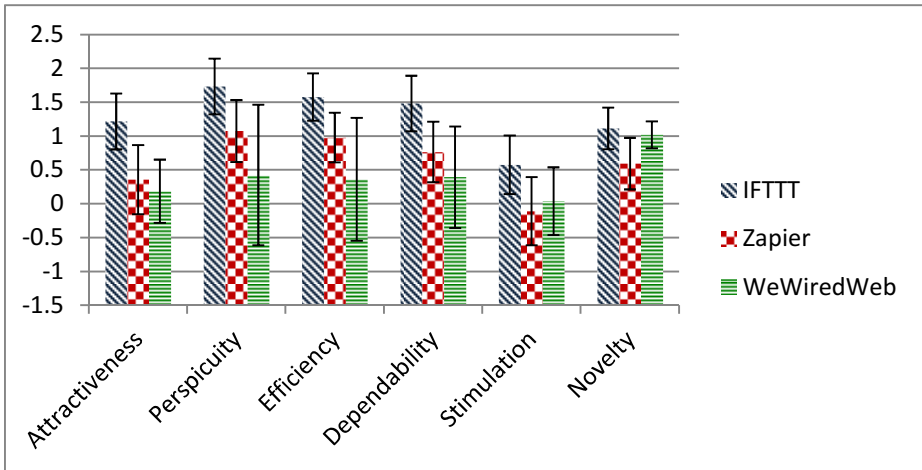Figure 3 shows the results for the three services.

Fig. 3. User experience evaluation for IFTTT, Zapier and We Wired Web

All three services offer quite similar functionality and interaction concepts. But their evaluation by students shows quite different results concerning user experience. Obviously IFTTT outperforms the other two services, especially with respect to pragmatic quality aspects.

The effort for such a comparative evaluation of different competing solutions is quite limited. As long as an access to the solutions is available or if it is possible to contact a sufficiently large sample of users, such an evaluation can be done in a couple of days. Suggestions on how to plan and perform such evaluations can be found in [14].

5   Test If a Product Has Sufficient User Experience

If a new product is launched, a typical question is if the user experience of the product is sufficiently high to fulfill the general expectations of users. Such expectations of users are formed during their interaction with other typical software products. These products must not belong to the same product category. For example, the everyday experience of users with modern web sites and modern interactive devices, like tablets or smartphones, has also heavily increased the expectations concerning the user experience of professional software, for example business applications, in the last couple of years.

Thus, the question if the user experience of a new product is sufficient can be answered by comparing the results for the product with the results of a large sample of other commonly used products, i.e. a benchmark data set.

For the UEQ such a benchmark was developed in the last couple of years [11]. The benchmark contains data from 163 product evaluations with the UEQ. These evaluated products cover a wide range of applications. The benchmark contains complex business applications (98), development tools (4), web shops or services (37), social networks (3), mobile applications (13), and a couple of other (8) products.

In total 4818 responses of subjects are contained in the benchmark. The number of respondents per evaluated product varied from extremely small samples (3 respondents) to huge samples (722 respondents). The mean number of respondents per study was 29.56 (std. deviation 73.5).

Many evaluations were done as part of usability tests, so the majority of the samples was in the range of 11 to 20 respondents (53%). The samples with more than 20 respondents (20%) were usually collected as online evaluations. Of course the studies based on tiny samples with less than 10 respondents (27%) do not carry much information. It was thus checked if these small samples had an influence on the benchmark data reported in the rest of this section. Since the results do not change much if studies with less than 11 respondents are eliminated, it was decided to keep them in the benchmark data set.

Since the benchmark data set contains currently only a quite limited number of evaluation results it was decided to limit the feedback per scale to 5 categories:

- *Excellent*: In the range of the 10% best results.
- *Good*: 10% of the results in the benchmark data set are better and 75% of the results are worse.
- *Above average*: 25% of the results in the benchmark are better than the result for the evaluated product, 50% of the results are worse.
- *Below average*: 50% of the results in the benchmark are better than the result for the evaluated product, 25% of the results are worse.
- *Bad*: In the range of the 25% worst results.

The following table shows the connection of these categories to the scale means for the 6 UEQ scales.

Table 1. Benchmark intervals for the UEQ scales

	Att.	**Eff.**	**Per.**	**Dep.**	**Sti.**	**Nov.**
Excellent	$\geq 1{,}72$	$\geq 1{,}64$	$\geq 1{,}82$	$\geq 1{,}6$	$\geq 1{,}50$	$\geq 1{,}34$
Good	$\geq 1{,}50$ $< 1{,}72$	$\geq 1{,}31$ $< 1{,}64$	$\geq 1{,}37$ $< 1{,}82$	$\geq 1{,}4$ $< 1{,}6$	$\geq 1{,}31$ $< 1{,}50$	$\geq 0{,}96$ $< 1{,}34$
Above average	$\geq 1{,}09$ $< 1{,}50$	$\geq 0{,}84$ $< 1{,}31$	$\geq 0{,}90$ $< 1{,}37$	$\geq 1{,}06$ $< 1{,}40$	$\geq 1{,}00$ $< 1{,}31$	$\geq 0{,}63$ $< 0{,}96$
Below average	$\geq 0{,}65$ $< 1{,}09$	$\geq 0{,}50$ $< 0{,}84$	$\geq 0{,}53$ $< 0{,}90$	$\geq 0{,}70$ $< 1{,}06$	$\geq 0{,}52$ $< 1{,}00$	$\geq 0{,}24$ $< 0{,}63$
Bad	$< 0{,}65$	$< 0{,}50$	$< 0{,}53$	$< 0{,}70$	$<0{,}52$	$< 0{,}24$

The benchmark is also included in the data analysis sheet and is automatically calculated together with the other statistics.
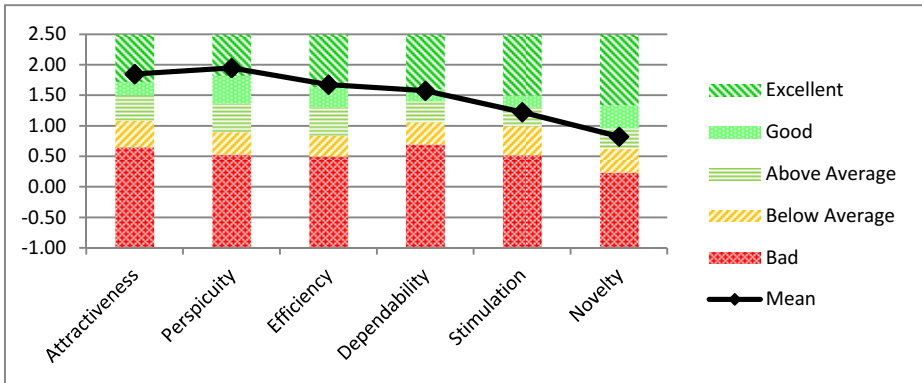
Fig. 4. Benchmark graph form the Excel tool

With the availability of a benchmark it is relatively easy to decide if a new product has sufficient user experience to be successful in the market. It is sufficient to measure the user experience with a large representative sample of users. A comparison of the results for the different scales with the results of the products in the benchmark allows then conclusions regarding the relative strengths and weaknesses of the product. However it must be noted that the general expectations concerning user experience grew over time. Since the benchmark contains also data from established products, a new product should reach at least the Good category in all scales.

6   Determine Areas of Improvement

Collecting quantitative data concerning user experience with a questionnaire like the UEQ is quite efficient. But this efficiency also has some drawbacks. We only get high level data concerning the scales of the UEQ, but the question which product features needs to be improved in order to increase user experience sometimes cannot be answered directly. If we compare these high level data with the results of a usability test, the situation is quite different. A usability test typically identifies a number of concrete problems, i.e. points that should be changed, but does not provide a good impression on how users feel about the product (especially since usability tests cause a lot of effort and can only be performed with a small sample of users).

However, with a questionnaire like the UEQ it is possible to make at least educated guesses about the areas where improvements will have the highest impact. For an evaluated product, the UEQ shows a pattern of 6 measured user experience qualities. From this pattern it is possible to make at least some assumptions where to look for improvements.

The evaluation of IFTTT shown in Figure 3 can serve as example. It shows good values (in the sense of the benchmark) concerning all pragmatic quality scales. Users seem to have the impression that it is easy to understand, efficient to use and offers a controllable interaction. On the other hand, the value for stimulation is not really

encouraging, so if effort is spent to increase the user experience of IFTTT it is quite clear that this effort should try to increase the fun of use of the service. A different pattern can be seen for product version B in Figure 2. Here it is quite clear that designers and developers first need to focus on an improvement of the pragmatic quality, especially efficiency and perspicuity.

7   Discussion

Obviously, a good user experience improves the chances of a product in the market. The UEQ offers the possibility to evaluate the user experience of a product quickly and efficiently. This simple and fast data collection with the UEQ offers the possibility to measure not only the current version of a product, but to establish a continuous measurement of different product versions for quality control. Another scenario that is possible due to this efficient measurement is to compare a product with its direct competitors to get information on the comparative position of the product.

The described benchmark offers an additional possibility to get an idea if the current user experience of a product is sufficient, by comparing it to a large number of different established products. Of course the benchmark offers just a high level impression about the position of a product in the market and should ideally be extended by a comparison to the direct competitors to get a clearer picture.

Clearly the efficiency of the UEQ has also the drawback that only high level information about strength and weaknesses of a product are provided. But since the different scales of the UEQ describe distinct quality aspects of an interactive product, some conclusions on concrete improvements are usually possible. This allows setting up special test activities that can help to get more detailed information about the problematic areas.

References

1. DIN EN 9241210, 2011-01, Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. Beuth, Berlin
2. Preece, J., Rogers, Y., Sharp, H.: Interaction design: Beyond human-computer interaction. Wiley, New York (2002)
3. Hassenzahl, M.: The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction 13, 479–497 (2001)
4. Norman, D.: Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books, Boulder Colorado (2003)
5. Tractinsky, N.: Aesthetics and Apparent Usability: Empirical Assessing Cultural and Methodological Issues. In: CHI 1997 Electronic Publications (1997), http://www.acm.org/sigchi/chi97/proceedings/paper/nt.htm
6. Laugwitz, B., Schrepp, M., Held, T.: Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In: Heinecke, A.M., Paul, H. (eds.) Mensch & Computer 2006 – Mensch und Computer im Strukturwandel, pp. 125–134. Oldenbourg Verlag (2006)

7. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 63–76. Springer, Heidelberg (2008)
8. Laugwitz, B., Schubert, U., Ilmberger, W., Tamm, N., Held, T., Schrepp, M.: Subjektive Benutzerzufriedenheit quantitativ erfassen: Erfahrungen mit dem User Experience Questionnaire UEQ. In: Brau, H., Lehmann, A., Petrovic, K., Schroeder, M. (eds.) Usability Professionals, pp. 220–225 (2009)
9. Rauschenberger, M., Schrepp, M., Cota, M.P., Olschner, S., Thomaschewski, J.: Efficient measurement of the user experience of interactive products - How to use the User Experence Questionnaire (UEQ). Example: Spanish Language Version. International Journal of Interactive Multimedia and Artificial Intelligence 2(1), 39–45 (2013)
10. Pérez Cota, M., Thomaschewski, J., Schrepp, M., Goncalves, R.: Efficient Measurement of the User Experience. A Portuguese Version. In: DSAI 2013 Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, Vigo, Spain, November 13-15 (2013)
11. Schrepp, M., Olschner, S., Schubert, U.: User Experience Questionnaire Benchmark - Praxiserfahrungen zum Einsatz im Business-Umfeld. In: Brau, H., Lehmann, A., Petrovic, K., Schroeder, M. (eds.) Usability Professionalspp, pp. 348–353 (2013)
12. UEQ, http://www.ueq-online.org ((last visited: January 20, 2014))
13. Hinderks, A., Schrepp, M., Rauschenberger, M., Olschner, S., Thomaschewski, J.: Konstruktion eines Fragebogens für jugendliche Personen zur Messung der User Experience. In: Brau, H., Lehmann, A., Petrovic, K., Schroeder, M. (eds.) Usability Professionals 2012, pp. 78–83 (2012)
14. Rauschenberger, M., Thomaschewski, J., Schrepp, M.: User Experience mit Fragebögen messen. Durchführung und Auswertung am Beispiel des UEQ. In: Brau, H., Lehmann, A., Petrovic, K., Schroeder, M. (eds.) Usability Professionals 2013, pp. 72–76. German UPA e.V., Stuttgart (2013)