Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)

Martin Schrepp¹, Andreas Hinderks², Jörg Thomaschewski³*

¹ SAP AG (Germany)  
² University of Seville (Spain)  
³ University of Applied Sciences Emden/Leer (Germany)

Received 19 June 2017 | Accepted 12 July 2017 | Published 11 September 2017

ABSTRACT

The user experience questionnaire (UEQ) is a widely used questionnaire to measure the subjective impression of users towards the user experience of products. The UEQ is a semantic differential with 26 items. Filling out the UEQ takes approximately 3-5 minutes, i.e. the UEQ is already reasonably efficient concerning the time required to answer all items. However, there exist several valid application scenarios, where filling out the entire UEQ appears impractical. This paper deals with the creation of an 8 item short version of the UEQ, which is optimized for these specific application scenarios. First validations of this short version are also described.

KEYWORDS

User Experience, UEQ, Questionnaire, Short Version.

Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)

Martin Schrepp¹, Andreas Hinderks², Jörg Thomaschewski³*

¹ SAP AG (Germany)  
² University of Seville (Spain)  
³ University of Applied Sciences Emden/Leer (Germany)

Received 19 June 2017 | Accepted 12 July 2017 | Published 11 September 2017

ABSTRACT

The user experience questionnaire (UEQ) is a widely used questionnaire to measure the subjective impression of users towards the user experience of products. The UEQ is a semantic differential with 26 items. Filling out the UEQ takes approximately 3-5 minutes, i.e. the UEQ is already reasonably efficient concerning the time required to answer all items. However, there exist several valid application scenarios, where filling out the entire UEQ appears impractical. This paper deals with the creation of an 8 item short version of the UEQ, which is optimized for these specific application scenarios. First validations of this short version are also described.

KEYWORDS

User Experience, UEQ, Questionnaire, Short Version.

The same is true for the perceived performance of a product. A product that a user perceives as slow and annoying can be seen as performing adequately by another user. In this respect, users vary widely in their expectations and personal preferences. Thus, any measurement of user experience must consider the feedback of a representative and large enough group of users.

Therefore, questionnaires are a simple method to collect such user feedback [3]. They can be distributed rather efficiently to larger groups of users, especially if they are designed as online questionnaires. In addition, analyzing the numerical data from such questionnaires is highly standardized and thus efficient as well.

In this paper we describe the design and evaluation of a short version of the User Experience Questionnaire, which is a widely used tool to measure user experience.

II. THE USER EXPERIENCE QUESTIONNAIRE (UEQ)

The objective of the UEQ is to allow a quick assessment done by end users covering a preferably comprehensive impression of user experience. It should allow the users to express feelings, impressions, and attitudes that arise when experiencing the product under investigation in a very simple and immediate way.

The UEQ can be used as a paper-pencil version, but is also short enough to be used as an online questionnaire. It consists of 26 items (Fig. 1) that are grouped into 6 scales.

Each item of the UEQ consists of a pair of terms with opposite meanings, for example:

<table>
<thead>
<tr>
<th>Not understandable</th>
<th>Understandable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficient</td>
<td>Inefficient</td>
</tr>
</tbody>
</table>

Participants rate each item on a 7-point Likert scale. The answers are scaled from -3 (fully agree with negative term) to +3 (fully agree with positive term). Half of the items start with the positive term, the others with the negative term (in randomized order).
The original version of the UEQ was designed in German [4], [5], but has so far been translated to several languages like Spanish [7] and Portuguese [11]. The English version of the UEQ is shown in Fig. 1.

| 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 |

Fig. 1. English version of the UEQ.

The original German version of the UEQ was designed using a data analytics approach to ensure the practical relevance of the constructed scales. Each scale represents a distinct UX quality aspect.

An initial set of more than 200 potential items related to UX was created with usability experts in two brainstorming sessions. A number of these experts then reduced the selection to a raw version of 80 items.

The raw version was used in several studies (with a total of 153 participants) on the quality of interactive products, including a statistics software package, cell phone address books, online collaboration software, or business software. Finally, the scales and the items representing each scale were extracted from this data set by principal component analysis [4], [5].

This analysis yielded the final questionnaire with 26 items arranged into six scales:

- **Attractiveness**: Overall impression of the product. Do users like or dislike it? Is it attractive, enjoyable or pleasing?
- **Perspicuity**: Is it easy to get familiar with the product? Is it easy to learn? Is the product easy to understand and unambiguous?
- **Efficiency**: Can users solve their tasks without unnecessary effort? Is the interaction efficient and fast? Does the product react to user input quickly?
- **Dependability**: Does the user feel in control of the interaction? Can he or she predict the system’s behavior? Does the user feel confident when working with the product?
- **Stimulation**: Is it exciting and motivating to use the product? Is it enjoyable to use?
- **Novelty**: Is the product innovative and creative? Does it capture the user’s attention?

Scales are not assumed to be independent. In fact, a user’s general impression is recorded by the Attractiveness scale, which should be influenced by the values on the other 5 scales (see Fig. 2).

![Fig. 2. Assumed scale structure of the User Experience Questionnaire (UEQ).](image)

**Attractiveness** is a pure valence dimension (emotional reaction on a pure acceptance/rejection dimension). Perspicuity, Efficiency, and Dependability are pragmatic quality aspects, i.e. they describe interaction qualities that relate to the tasks or goals the user aims to reach when using the product. Stimulation and Novelty are hedonic quality aspects, i.e. they do not relate to tasks and goals, but describe aspects related to pleasure or fun while using the product [2], [6].

For details concerning the design and validation of the UEQ see [4], [5]. Helpful hints on using the UEQ are also available in [7], [8]. There is a benchmark available as well, which is described in [12].

For a semantic differential like the UEQ, it is very important that participants can fill it out in their native language. Thus, several contributors created a number of translations.

The UEQ is currently available in 19 languages (German, English, French, Italian, Russian, Spanish [7], Portuguese [11], Turkish, Chinese, Japanese, Indonesian [10], Dutch, Estonian, Slovene, Swedish, Greek, Polish, Hindi, and Bulgarian).

The UEQ in all available languages, an Excel sheet for data analysis, and the UEQ Handbook are available free of charge at [www.ueq-online.org](http://www.ueq-online.org).

### III. Scenarios Requiring a Short Version?

Usually, 3-5 minutes are sufficient to fill out the UEQ including some demographic data [3]. Thus, the UEQ is already a quite efficient method to capture the opinion of a user towards the user experience of a product, leading to the obvious question why a shorter version is needed at all?

In the last couple of years we received a number of requests for a shorter version, and some users even created their own short version by removing a few items (which is not a recommended practice for a standardized questionnaire like the UEQ [9]). Accordingly, there seem to be some cases in which a full UEQ is considered to be too time consuming.

All these requests came from three different generic application scenarios in which only a very small number of items could be used to measure user experience.

1. The first scenario is collecting data when the user leaves a web shop or web service. For example, the user has just ordered something in a web shop and logs out. After pressing the log out button, the user is asked to fill out a short questionnaire concerning the user experience of the shop. In such scenarios it is crucial that the user has the impression that filling out the questionnaire can be
done extremely fast. Otherwise, users will refuse to give feedback (they are finished with their initial task and are in the process of leaving the shop, so motivating them to spend some more time on feedback is difficult). Presenting an entire UEQ with all 26 questions in such a scenario will severely reduce the number of users willing to give feedback.

2. In the second and quite frequent scenario, a questionnaire concerning user experience should be included in an already existing product experience questionnaire. Typically, such a questionnaire is sent out after a customer has purchased a product and has already used it for some time.

Such questionnaires try to collect data about the entire product experience, asking, for example, why the customer chose the product, if the functionality of the product fulfills the expectations, if the purchasing process was pleasant, if the customer wants to be informed about similar or other products of the company in the future, etc. As a result, such questionnaires tend to be quite lengthy. Thus, it is difficult to add a full 26 item user experience questionnaire in such cases.

On the other hand, it is often not possible to collect data concerning user experience in a separate questionnaire, since the number of customer interactions cannot exceed a certain limit (customers can get easily annoyed if they receive such marketing e-mails too often).

Thus, including a very short user experience section in such a customer experience questionnaire is often the only way for UX practitioners to collect feedback on their customers’ user experience.

3. A third scenario mentioned sometimes are experimental settings where a participant is asked to judge the user experience of several products or variants of a product in one session. In such scenarios the products or product variants are presented to the participant in a random order one after the other, and they have to fill out a questionnaire concerning user experience for each of them. In such a setting, the number of items must be kept to a minimum. Otherwise the participant will be stressed and the quality of answers will decrease quickly.

All of these scenarios share the requirement that the number of items must be small. In addition, any instruction must be simple and quick to read.

IV. CONSTRUCTION OF THE SHORT VERSION

The short version should contain only a limited number of items, but it should still cover the spectrum of product qualities measured by the UEQ.

To shorten the UEQ it was decided to skip the measurement of the single dimensions and to concentrate on the measurement of the two meta-dimensions pragmatic and hedonic quality. For each of these dimensions four items are chosen. Thus, the short version of the UEQ (henceforth: UEQ-S) will only contain eight items, grouped into two scales. In addition, the mean value of the eight items will be given as an overall UX value.

A data set with 1867 data records was collected by the German UEQ in previous studies. Each data record reflects the evaluation of a product by a participant. In total, 21 different products were assessed (business software, web shops, household appliances, etc.).

A main component analysis was performed on all twelve UEQ items from the Efficiency, Perspicuity and Dependability scales. For the analysis the number of factors was set to 1, and the four items that showed the highest loading on this factor were chosen. These were the items 11, 13, 20, and 21 (see Fig. 1) of the UEQ. They therefore represent the Pragmatic Quality scale of the short version UEQ-S.

The same procedure was repeated for the eight UEQ items from the Stimulation and Originality scales. The items 6, 7, 10, and 15 (see Fig. 1) showed the highest loadings and thus represent the Hedonic Quality scale of the UEQ-S.

The UEQ-S thus consists of items shown in Fig. 3:

![Fig. 3. The short version UEQ-S.](image)

Thus, the items show the intended scale structure. Only the item obstructive/supportive yields a relevant, but still relatively small cross-loading to the other factor. The other items load strongly on the factor they belong to and only weakly to the other factor.

In the original UEQ half of the items start with the positive term and the other half of items start with the negative term. In addition, the order of the items is randomized in the questionnaire. This was done to be able to detect participants that do not answer seriously [9] and to force users to carefully read the alternatives. However, this also has some disadvantages. The change of polarity must be explained in the instruction and in addition, it is cognitively more demanding for the participants.

In order to simplify the instruction and make it easier to fill in the questionnaire, it was decided that all items have the same polarity. The left side reflects the negative term and the right side the positive term (see Fig. 3). In addition, the order is not randomized: the first 4 items reflect the pragmatic quality and the items 5 to 8 the hedonic quality.

V. PREDICTION QUALITY

As a first evaluation we calculated how well the scales of the short version (UEQ-S) approximate the corresponding scales of the full version (UEQ). Therefore, in our data set that was used to design the short version, we calculated the difference between the mean value of all 8 items in the short version and the mean value of all 26 items (here 12 items belong to pragmatic quality, 8 belong to hedonic quality and...
belong to the scale Attractiveness, which is neither pragmatic nor hedonic) in the full UEQ for each participant. The same was done for the scales pragmatic and hedonic quality of the short version.

Regarding the pragmatic quality we calculated the difference between the mean of the four items of the pragmatic quality scale of the UEQ-S and all twelve items of the Efficiency, Perspicuity and Dependability scales in the full UEQ for each participant.

Following that pattern, we compared the difference between the four items of the hedonic quality scale of the UEQ-S with the mean of the eight items of the Stimulation and Originality scales in the full UEQ.

The distribution of these differences (kernel density plots) is shown in Fig. 4, Fig. 5 and Fig. 6.

The mean and standard deviation of the observed differences are 0.06 (0.39) for all items (Fig. 4), -0.09 (0.46) (Fig. 5) for the items concerning pragmatic quality and -0.03 (0.45) (Fig. 6) for the items concerning hedonic quality. Please note that the UEQ scale ranges from -3 to +3, so these differences concerning the scale means are quite small.

In all three cases the distribution of the observed differences is nearly symmetrical around zero, thus there is no systematic over- or underestimation based on the reduced number of items in the short version UEQ-S. It is evident that the short version is able to predict the values of the full version quite accurately.

VI. An Evaluation Study

In a first study with the short version 47 students judged the user experience of different well-known products. Each student could choose to judge either Amazon, Skype or Wikipedia with an online version of the UEQ-S. We only report results for Amazon, since for the other two products there were simply not enough data to draw meaningful conclusions.

The consistency of the pragmatic quality and hedonic quality scales was reasonably high. The corresponding Cronbach Alpha values were 0.85 (pragmatic quality) and 0.81 (hedonic quality).

The scale means for Amazon (N=31 students decided to judge Amazon) were 1.09 for pragmatic and 0.51 for hedonic quality. These values are quite similar to the values obtained in an older study with a similar target group (German students) and the full UEQ. In this study the mean value for the Efficiency, Perspicuity and Dependability UEQ scales was 1.17 and the mean value for Stimulation and Originality was 0.66. Thus, the short version UEQ-S seems to approximate the long version expectedly well.

A main component analysis of this data set shows the expected factor structure once more. The loading of the items of the two extracted factors (factors were extracted according to the Kaiser-Guttman criterium, loadings after varimax rotation) are shown in Table 2.

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1 pragmatic quality</th>
<th>Factor 2 hedonic quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>clear / confusing</td>
<td>0.825</td>
<td>-0.209</td>
</tr>
<tr>
<td>inefficient / efficient</td>
<td>0.833</td>
<td>0.255</td>
</tr>
<tr>
<td>complicated / easy</td>
<td>0.849</td>
<td>-0.033</td>
</tr>
<tr>
<td>obstructive / supportive</td>
<td>0.789</td>
<td>0.215</td>
</tr>
<tr>
<td>boring / exiting</td>
<td>0.274</td>
<td>0.701</td>
</tr>
<tr>
<td>not interesting / interesting</td>
<td>0.220</td>
<td>0.806</td>
</tr>
<tr>
<td>conventional / inventive</td>
<td>-0.047</td>
<td>0.841</td>
</tr>
<tr>
<td>usual / leading edge</td>
<td>-0.221</td>
<td>0.809</td>
</tr>
</tbody>
</table>

VII. Language Versions

The items of the UEQ-S are a subset of the UEQ items. Accordingly, all the available translations of the UEQ can be used, i.e. the UEQ-S is directly available in all languages for which a full version exists (German, English, French, Italian, Russian, Spanish, Portuguese, Turkish, Chinese, Japanese, Indonesian, Dutch, Estonian, Slovene, Swedish, Greek, Polish, Hindi, and Bulgarian).

The translated version of the UEQ-S can simply be created by choosing the corresponding items from the full UEQ of the desired language.

However, the question remains if the selected items predict the
behavior of the full UEQ as well as the German version. We cannot verify this for all translations yet, since we do not have access to sufficiently large data sets for all of them. So far, this is only possible for some languages.

Table 3 shows the measured deviation per participant between the short version and the corresponding values for the long version for five languages (same computation method as described in the validation of the German short version above).

<table>
<thead>
<tr>
<th>Language</th>
<th>All</th>
<th>Pragmatic</th>
<th>Hedonic</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portuguese</td>
<td>0.19 (0.33)</td>
<td>0.07 (0.41)</td>
<td>0.13 (0.41)</td>
<td>206</td>
</tr>
<tr>
<td>Turkish</td>
<td>0.13 (0.39)</td>
<td>-0.06 (0.44)</td>
<td>0.22 (0.47)</td>
<td>943</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.16 (0.38)</td>
<td>0.02 (0.49)</td>
<td>0.1 (0.38)</td>
<td>377</td>
</tr>
<tr>
<td>Indonesian</td>
<td>0.14 (0.33)</td>
<td>-0.15 (0.38)</td>
<td>0.25 (0.38)</td>
<td>212</td>
</tr>
<tr>
<td>English</td>
<td>0.15 (0.34)</td>
<td>0.03 (0.43)</td>
<td>0.17 (0.36)</td>
<td>224</td>
</tr>
</tbody>
</table>

The data shows that for these languages the fit between the short version and the full version of the UEQ is good enough to allow a practical application of the UEQ-S. For example, for the English version we can expect that the mean of the eight items of the UEQ-S deviates 0.15 (on average) from the mean of all 26 UEQ items. For the four UEQ-S items of the pragmatic quality scale the deviation from the mean of all twelve UEQ items of the Efficiency, Dependability and Perspicuity scales is 0.03 on average. For the four UEQ-S items of the hedonic scale the deviation from the mean of all eight UEQ items of the scales Stimulation and Originality averages at 0.17. Thus, as in the case of the German version, the approximation is quite good.

VIII. LIMITATIONS OF THE SHORT VERSION UEQ-S

We described the design of the UEQ’s short version UEQ-S. For a UX professional who wants to plan an evaluation the question arises which of the two versions should be used. Obviously, the short version has some advantages concerning the number of questions and accordingly the time the participants need to fill out the questionnaire.

However, this comes at a price. The full UEQ gives a detailed feedback concerning 6 different aspects of UX, i.e. measures on the Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation and Originality scales. This is lost in the short version that only distinguishes between pragmatic and hedonic quality. Given the fact that even a full UEQ requires only 3-5 minutes, the usage of the UEQ-S should be limited to the scenarios described in the beginning of this paper. The short version should only be used in situations where a full UEQ can not be applied at all. Otherwise, the loss of detailed information is not compensated by saving time in filling out the questionnaire.

IX. CONCLUSION

We described the design and validation of a short version of the UEQ. It consists of only eight of the 26 items of the UEQ. The short version, which is named UEQ-S, contains two subscales (pragmatic and hedonic quality; 4 items each) and a total value reflecting the overall user experience.

It was shown that the short version is able to predict the behavior of the full version concerning pragmatic and hedonic quality. The mean value obtained by the four items of the short version approximates the values obtained by averaging all 12 pragmatic items (from the Efficiency, Perspicuity and Dependability scales) and all 8 hedonic items (from the Stimulation and Originality scales) of the full version.

In a first application study concerning Amazon done with German students, the scales showed a high level of consistency. In addition, the measured mean for the pragmatic and hedonic quality approximates the values obtained by the full UEQ collected in a previous study.

The short version UEQ-S is only intended for specific scenarios which do not allow employing a full UEQ. The UEQ-S does not allow measuring the detailed UX qualities Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation and Novelty, which are part of the UEQ report. It is, in general, quite useful to gather these detailed values when it comes to interpreting the results and define areas of improvement [3].

Thus, the short version UEQ-S only allows a rough measurement on higher level meta-dimensions. Our recommendation therefore is to only use the short version UEQ-S in the scenarios described in this paper. The short version should not replace the usage of the full version in standard scenarios, for example after usability tests. In such scenarios, the small gain in efficiency does not compensate for the loss of detailed information on the single scales and therefore more detailed quality aspects.

REFERENCES

Martin Schrepp

Martin Schrepp has been working as a user interface designer for SAP AG since 1994. He finished his diploma in mathematics in 1990 at the University of Heidelberg (Germany). In 1993 he received 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 data analysis methods. He has published several papers concerning these research fields.

Andreas Hinderks

Andreas Hinderks graduated with a diploma in Computer Science and is Master of Science in Media Informatics by the University of Applied Science Emden/Leer. He has worked as a business analyst and a programmer from 2001 to 2016. Back then, he focused on developing user-friendly business software. Currently, he is a freelancing business analyst and senior UX architect. Furthermore he is a Ph.D. student at the University of Seville, Spain. He is involved in research activities dealing with UX questionnaires, process optimization, information architecture, and user experience since 2011.

Jörg Thomaschewski

Jörg Thomaschewski received a PhD in physics from the University of Bremen (Germany) in 1996. He became Full Professor at the University of Applied Sciences Emden/Leer (Germany) in September 2000. His research interests are Internet applications for human-computer interaction, e-learning, and software engineering. Dr. Thomaschewski is the author of various online modules, e.g., “Human-Computer Communication,” which are used by the Virtual University (online) at six university sites. He has wide experience in usability training, analysis, and consulting.