5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013

Efficient Measurement of the User Experience. A Portuguese Version

Manuel Pérez Cota a,*, Jörg Thomaschewski b, Martin Schrepp c, Ramiro Gonçalves d

a University of Vigo, Rúa Torrecedeira, 86; 36208 Vigo, Spain
b Hochschule Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany
c SAP AG, Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
d INESC TEC and UTAD, Quinta de Prados, 5001-801 Vila Real, Portugal

Abstract

For the international use of software products it is important to know the culture, language and behavior of the citizens. Means: internationalization and location. To be able to evaluate these products it is significant to know how people from a country behave and express their feelings. This article presents a questionnaire which was initially developed in German, English and Spanish and is now available after a complex transformation in Portuguese.

© 2013 The Authors. Published by Elsevier B.V.
Selection and peer-review under responsibility of the Scientific Programme Committee of the 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2013).

Keywords: User Experience; Measurement; Software Quality; Questionnaire

1. Introduction

The concept of user experience (see [1] or [2]) extends the classical usability concept with additional quality aspects. Usability concentrates on tasks users want to perform with a product. For example, the well-known ISO 9241-11 defines usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. Thus, task related quality aspects like

* Corresponding author. Tel.: +(34)-986-813933; fax: +(34)-986-813644.
E-mail address: mpcota@uvigo.es
efficiency, effectiveness, learnability, perspicuity, controllability, etc. subsume under this concept. These quality aspects are often called pragmatic or task-related quality aspects.

In the last couple of years there is a clear tendency to extend the concept of usability to a more holistic view on the interaction between humans and systems, which is often referred as user experience. User experience is a summary of the findings fun of use [3], aesthetics, emotions [4], stimulation or attractiveness. These quality aspects are not related to tasks users perform with a product and are thus called non-task related or hedonic aspects.

User experience describes the subjective impression of users towards a product before, during and after the use (ISO 9241-210). 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. For example, older users may experience a public web service quite different than younger technically highly skilled users. The target group of a product is essential to be looked at.

To create products or services that are successful in the long run it is therefore necessary to ensure that the product has a sufficiently high user experience for all relevant user groups. That is especially important for public services in the web. To avoid that certain user groups, for example older persons, less educated persons or persons with a certain disability, reject to use such a public web service because it has a too low user experience for them, it is important to control this quality aspect permanently.

Classical usability methods, for example usability tests, cannot be used for this purpose, since it is typically too expensive to do the test with a sample large enough to represent all intended target groups of users. In addition, it is not only necessary to detect and eliminate real usability problems. If a product is used or not depends not only on the existence of such usability problems, but also on the general impression users develop towards a product, i.e. the general user experience. Empirical results [5] show, for example, that even for products like business software, the attractiveness of the product is nearly equally affected by its perceived pragmatic and hedonic quality.

Since subjective impressions of larger user groups must be captured in order to control and improve a product or service, questionnaires are a natural way to do such measurements. A questionnaire can be applied online or as a print version. It is, for example, possible to ask users of a web service to fill out a short questionnaire when they leave the service. This allows data collecting from large samples of users with little effort.

The User Experience Questionnaire (UEQ) [6] was developed for such scenarios and the evaluation of usability and user experience. 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 format of the questionnaire supports the immediate user response to express feelings, impressions and attitudes that arise when they use a product. In addition, the questionnaire is short enough to be applied as an online form. It can be filled out by a participant in a few minutes even if demographical question are added at the beginning or end of the questionnaire.

### 2. User Experience Questionnaire (UEQ)

The original German version of the UEQ was created by a data analytical approach. An initial item set of 229 potential items related to the concept of user experience was created in several brainstorming sessions with usability experts. This initial set was then reduced to an 80 items raw version of the questionnaire by an expert evaluation.

This 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 book, 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 factor analysis (principal components, varimax rotation). Details concerning the construction process of the UEQ can be found in [6] and [7].

The reliability (i.e. the scales are consistent) 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 [6], [7], [8] showed a good construct validity of the scales.
The final version of the user experience questionnaire contains thus 6 scales with 26 items in total:

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

- **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.

- **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.

- **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.

- **Stimulation**: Is it interesting and exciting to use the product? Does the user feel motivated for a further using of the product? Items: valuable / inferior, boring / exiting, not interesting / interesting, motivating / demotivating.

- **Novelty**: Is the design of the product innovative and creative? Does the product grab the user’s attention? Items: creative / dull, inventive / conventional, usual / leading edge, conservative / innovative.

Fig. 1 shows the scales of the UEQ and their theoretical scale structure.

![Fig. 1. Scale structure of the UEQ.](image)

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. 2.
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. 3 shows an example for an overall result for a product. The graphic is automatically generated by the data analysis sheet that can be downloaded together with the questionnaire.
Currently the UEQ is available in German (original version), English and Spanish. We describe in this paper the creation of a Portuguese language version. All language versions of the UEQ, a MS Excel tool to analyze the data and material that describes how the results should be interpreted and reported are freely available on http://www.ueq-online.org.

3. Development of a Portuguese Language Version

To begin the process of translation into Portuguese, first a comprehensive study of the previous versions (German, Spanish and English) was done, to be very clear of what we want indicate with each word. The intention was to avoid problems we have had in previous translations.

The first trial was conducted in a similar way as we did the first version in Spanish (the complete process can be seen in “Measurement of user experience” [10]), but we should remark here the similarities: Firstly we worked with two bilingual persons, so that the translation process could be much more accurate. We always tried to work together so that, in real time, we could compare and review the use of certain words. One of the features we try to meet in all languages, in which we are working with, is that the chosen words mean and make everybody feel pretty much the same who would perform the test.

To confirm the correct choice it was necessary to carry out preliminary assessments with persons other than those who chose the original words. For that, men and women (with a total of 8 people) chose within the proposals that were given, the same words for the same concept, then the result was the same as that obtained by those who had already done with the first choice. Obviously these people were already working on words prefixed by the first translators. The specific Portuguese questionnaire is shown in Fig. 4.

After this process we carried out the test appearing in the next section, with the results shown.

<table>
<thead>
<tr>
<th>Desagradável</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Agradável</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incompreensível</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Compreensível</td>
</tr>
<tr>
<td>Criativo</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Sem criatividade</td>
</tr>
<tr>
<td>De Fácil aprendizagem</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Sem facilidade</td>
</tr>
<tr>
<td>Valioso</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Sem valor</td>
</tr>
<tr>
<td>Aborrecido</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Excitante</td>
</tr>
<tr>
<td>Desinteressante</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Interessante</td>
</tr>
<tr>
<td>Imprévável</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Previsível</td>
</tr>
<tr>
<td>Rápido</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Lento</td>
</tr>
<tr>
<td>Original</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Convencional</td>
</tr>
<tr>
<td>Obstrutivo</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Condutor</td>
</tr>
<tr>
<td>Bom</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Mau</td>
</tr>
<tr>
<td>Complicado</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Fácil</td>
</tr>
<tr>
<td>Desinteressante</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Atrativo</td>
</tr>
<tr>
<td>Comum</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Vanguardista</td>
</tr>
<tr>
<td>Incómodo</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Cémodo</td>
</tr>
<tr>
<td>Seguro</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Inseguro</td>
</tr>
<tr>
<td>Motivante</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Desmotivante</td>
</tr>
<tr>
<td>Atende as expectativas</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Não atende as expectativas</td>
</tr>
<tr>
<td>Ineficiente</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Eficiente</td>
</tr>
<tr>
<td>Evidente</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Confuso</td>
</tr>
<tr>
<td>Impraticável</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Prático</td>
</tr>
<tr>
<td>Organizado</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Desorganizado</td>
</tr>
<tr>
<td>Atraente</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Feio</td>
</tr>
<tr>
<td>Simpático</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Antipático</td>
</tr>
<tr>
<td>Conservador</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>Inovador</td>
</tr>
</tbody>
</table>

Fig. 4. Portuguese language version of the UEQ.
4. First Evaluation of the Portuguese Version

For questionnaires of this type (semantic differential) it is not sufficient to simply translate the items. It is important to translate the items with the same relation to the dimension as before i.e. measures the same qualities. At first the initial evaluation study of the Portuguese language version will be described. The evaluation started with 104 Portuguese students of degree and master in informatics and electrical engineering. All of them are native Portuguese speakers.

The result (Fig. 5) shows that the participants had a slightly positive or neutral impression concerning the user experience of Amazon. Their impression concerning the pragmatic quality (Perspicuity=Transparência, Efficiency=Eficiência and Dependability=Controlo) of the webshop is higher than the impression concerning the hedonic quality (Stimulation=Estimulação, Novelty=Inovação). This is somehow natural, since typical users visit Amazon with a clear goal (to buy something). This means, that the pragmatic quality is in this case in a higher importance than the hedonic quality.

An analysis of the Cronbach Alpha coefficient showed that the single scales revealed high consistency values (Attractiveness: 0.86, Perspicuity: 0.65, Efficiency: 0.77, Dependability: 0.51, Stimulation: 0.83, Novelty: 0.70). This is an indicator that the scales are sufficiently consistent.
The impression of the participants concerning the Skype user experience (Fig. 6) is generally positive and much better than their impression on Amazon (Fig. 5). Again pragmatic quality is judged a little bit better than hedonic quality aspects.

As in the previous study the alpha coefficients for the scales shows had high values (Attractiveness: 0.85, Perspicuity: 0.72, Efficiency: 0.74, Dependability: 0.64, Stimulation: 0.79, Novelty: 0.64) again indicating sufficient scale consistency.

Another interesting result is that the observed scale values are quite similar to the results obtained from a Spanish group of students (95 participants) that judged Amazon (Fig. 7) and Skype (Fig. 8) using the Spanish language version (that was already evaluated successfully, see [9, 10]).

![Fig. 7. Result of an evaluation of Amazon using the Spanish language version.](image1)

![Fig. 8. Result of an evaluation of Skype using the Spanish language version.](image2)

This comparison shows that the Portuguese version has sufficient scale consistency and seems to measure the same qualities than the Spanish language version that was already successfully evaluated against the original versions of the UEQ.
5. Conclusions

We showed how user experience of interactive products can be measured efficiently over well-designed questionnaires. This allows you to collect higher numbers of user feedback with little effort. The questionnaire can be used to control if the user experience of a product or service if sufficiently high for all intended target groups and can thus help to avoid that special user groups are excluded because the design of the product does not fit their needs.

The construction process and first validations of the new language version were described. The results of the first validations indicate that the scales of the questionnaire are sufficiently consistent and that the questionnaire seems to measure the intended quality aspects of products.

References