Results of a Study to Improve the Spanish Version of the User Experience Questionnaire (UEQ)

Mónica Hernández-Campos[1]*, Jörg Thomaschewski[2], Yuen C. Law[1]

[1] Instituto Tecnológico de Costa Rica, Cartago (Costa Rica)
[2] University of Applied Sciences Emden/Leer, Emden (Germany)

Abstract

This paper analyses changes in some items of the User Experience Questionnaire (UEQ) for use in the context of Costa Rican culture. Although a Spanish version of the UEQ was created in 2012, we use a double-translation and reconciliation model for detecting the more appropriate words for Costa Rican culture. These resulted in 7 new items that were added to the original Spanish version. In total, the resulting UEQ had 33 items. 161 participants took part in a study that examined both the original items and the new ones. Static analyses (Cronbach's Alpha, mean, variance, and confidence interval) were performed to measure the differences of the scales of the original items and the new UEQ variant with the Costa Rican words. Finally, confidence intervals of the individual items and Cronbach's Alpha coefficient average of the affected scales were analysed. The results show, contrary to initial expectations, that the Costa Rican word version is neither better nor worse than the original Spanish version. However, this shows that the UEQ is very robust to some changes in the items.

Keywords

I. Introduction

Nowadays, users expect devices, products and services that offer quite natural and easy-to-learn interactions. Especially the daily use of smartphones or tablets have brought the general expectation of users regarding the user experience of user interfaces to a high level, even if it is a complex business application. Simply said, users today expect a perfect user experience. A well-known definition of user experience is given in ISO 9241-210 (2019) [1]. Here, user experience is defined as "*user's perceptions and responses that result from the use and/or anticipated use of a system, product or service*" [1]. Thus, user experience is seen as a holistic concept that includes all types of emotional, cognitive, or physical reactions regarding the actual or even perceived use of a product or service that occur before, during, and after use. Still, the standard does not provide a clear list of factors or methods for measuring user experience.

In many cases, questionnaires are used to measure the user experience of products or services because UX questionnaires are easy to use, and a common quantitative way to measure user experience [2]. There are various UX questionnaires, such as meCUE [3], SUPR-Q [4], UEQ [5], [6], VisAWI [7], and Web-CLIC [8]. One goal of using a UX questionnaire is the idea of getting a better understanding of the own product or service and making appropriate improvements.

All steps in a testing process, including design, validation, adaptation, administration, and scoring, should be designed to minimize construct-irrelevant variance and promote valid score interpretations for all examinees in the intended population. Removing all barriers

* Corresponding author.

E-mail address: mohernandez@itcr.ac.cr

allows for the comparable and valid interpretation of test scores for all examinees, which is central to the validity and comparability of test scores. For this reason, those responsible for all steps in the testing process should guarantee to minimize the potential threats to validation such as linguistic, communicative, cognitive, cultural, or age matters. These characteristics can impede some individuals in demonstrating their standing on intended constructs. Often, a product or service must be offered in different languages. Thus, the measurement of the user experience should also be carried out in the languages in which the tool is available, so that users can do the evaluation in their native language. One of the most critical aspects is language and its cultural variations. According to international standards in testing, it is necessary to avoid the use of language that has different meanings or connotations for the test-takers as well as the use of unfamiliar words [9].

The User Experience Questionnaire (UEQ) is the one of very few standard UX questionnaires available in many different languages. At the moment, 36 language versions are offered (see ueq-online. org). The language versions are usually conscientiously constructed and evaluated in the individual countries by local scientists. The UEQ maintainers then include the language version on their website ueq-online.org and often stay in touch with local language version scientists beyond that. The Spanish version of the UEQ has been carefully created and evaluated (see [10], [11]). Particularly in Latin America, regional variations of the language have developed in each country. Although the words in each variation are generally understood by native speakers in other countries, slight differences in usage and meaning might hinder communication. Differences between European Spanish and American Spanish are even greater.

In the case of the UEQ, unwanted and unknown effects of different meanings might exist for some of the items. Due to this

large cultural Spanish language area and the related different use of words and meanings, requests for changes to the Spanish version of the UEQ are sent to the UEQ maintainers from different research groups, and for the case of Costa Rica, this is no exception. In order to ensure the fairness and validity of the test and to avoid a language bias, a new set of words are proposed for some of the items. There are two possible outcomes from this investigation: 1) The proposed new words are a better fit, in which case the results of the UEQ will better represent the users' experience; and 2) The UEQ is robust enough to accept modifications of some words, which will allow the use of words that are more familiar in the region, hence reducing the risk of item misinterpretation. Furthermore, there are also requests for adaptation of various items in other languages (e.g., for the French and Arabic versions), so the procedures and findings described here about Spanish adaptation are of more global importance.

This article analyses changes to the UEQ items to better understand the items in Costa Rica. For this purpose, 163 participants took part in a study that examined both the original items and the items with more culturally appropriate words.

II. Construction of the German Version and Spanish Version of the UEQ

The original German version of the UEQ was created by Laugwitz et al. in 2006 [5] using 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. These 80 items raw version was used in several studies. 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 the works of Laugwitz and colleagues [5], [6].

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 were 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). As a result of this questionnaire construction, 6 scales with the following items were obtained.

Attractiveness: General impression towards the product. Do users like or dislike the product? The scale is a 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 look 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 predicable? 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 use 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.*

Attractiveness is a pure valence dimension and consists of 6 items. Perspicuity, Efficiency and Dependability measure the goal-directed aspects, while Stimulation and Novelty measure the non goal-directed aspects. These scales are each measured with 4 items (see list above). In total, there are 5 scales with 4 items each and the scale attractiveness with 6 items. The entire questionnaire thus consists of 26 items.

It is easy to see in the list above that each item of the UEQ consists of a pair of terms with opposite meanings. So, a semantic differential was chosen as item format, since this allows a fast and intuitive response. Each item can be rated on a 7-point Likert scale. Answers to an item therefore range from -3 (fully agree with negative term) to +3 (fully agree with positive term). Half of the items start with the positive term, the rest with the negative term (in randomized order).

Examples:

Not understandable o o o o o o o *Understandable*

Efficient o o o o o o o *Inefficient*

Applying the UEQ does not require much effort. Usually 3-5 minutes are sufficient for a participant to read the instructions and complete the questionnaire. The UEQ can either be used in a paper-pencil form or as an online questionnaire. Analysing the results of the UEQ is also no effort, as a comprehensive Excel tool is available for this purpose on the website. This Excel tool also contains a Benchmark [12] for a better interpretation of the result.

As described in Rauschenberger et al. [10], a Spanish version of the UEQ was created in 2012. First, the German version of the UEQ was translated into Spanish by two scientists with human computer interaction (HCI) and UEQ experience, a native Spanish speaker (living in Spain) and a bilingual scientist (native German, Spanish level C1, living in Germany). The translation was done in joint discussion for each item. During translation, the English version was also used to better align the items. Afterwards, the Spanish version was back-translated into German by an independent scientist (native German, Spanish level C2, living in Spain). If the words matched the original words, the translation was considered successful. Otherwise, the process was repeated until all words matched.

In a next step, the translation was checked with two different studies [11]. The web shop amazon.de and the communication software Skype were used, each with 94 participants. The two studies were conducted in Spain (Vigo) and found to have good internal consistency, determined with the Cronbach's Alpha [11].

Later, international comparative studies with different test objects have also confirmed the good appropriability of the results of the Spanish UEQ version and its internal consistency (e.g. [12]).

III. Methods

As described above, the main purpose of this work was to adapt and validate the Spanish version of the original UEQ to Costa Rican culture. For this matter, we first translated the original German words to Costa Rican Spanish, using a double-translation and reconciliation model [14]. A native Costa Rican Spanish speaker with a C1 German level translated the words to Spanish, these were then translated back to German by a native German speaker who is familiar with Costa Rican Spanish (double-translation). From the resulting back-translated words, four pairs were completely different to the original German words. We reviewed and corrected the translations for these pairs (reconciliation). The resulting Spanish word list was finally compared to the Spanish version available at the UEQ website and the pairs that were completely different were selected. These resulted in 7 new items that were added to the original Spanish version. In total, the resulting UEQ had 33 items. The original items, the corresponding new items

and the affected scales are shown in Table I. The new UEQ was then applied in a study to compare the new items with their existing counterparts.

No	New items & (Scale)	No	Original Items & (Scale)	Engl. Items & (Scale)
27	Tedioso/ Ameno (Atracción)	1	Desagradable/ Agradable (Atracción)	Annoying/ Enjoyable (Attractiveness)
28	Incomprensible/ Comprensible (Transparencia)	2	No entendible/ Entendible (Transparencia)	Not understandable/ Understandable (Perspicuity)
29	Estorboso/ Facilitador (Controlabilidad)	11	Obstructivo/ Impulsor de apoyo (Controlabilidad)	Obstructive/ Supportive (Dependability)
30	Repugnante/ Llamativo (Atracción)	14	Repeler/ Atraer (Atracción)	Unlikable/ Pleasing (Attractiveness)
31	Molesto/ Placentero (Atracción)	16	Incómodo/ Cómodo (Atracción)	Unpleasant/ Pleasant (Attractiveness)
32	Según lo esperado/ Contrario a lo esperado (Controlabilidad)	19	Cubre expectativas/ No cubre expectativas (Controlabilidad)	Meets expectations/ Does not meet expectations (Dependability)
33	Poco práctico/ práctico (Eficiencia)	22	No pragmático/ Pragmático (Eficiencia)	Impractical/ Practical (Efficiency)

As described previously, the scale Attractiveness consists of 6 items and all other scales consist of 4 items. In Table I, it is seen that 3 items of the scale Attractiveness were modified (= 50%), 2 items of the scale Dependability were modified (= 50%), 1 item each of the scale Perspicuity (= 25%), and Efficiency (= 25%) were also modified. The items of the Stimulation and Novelty scales remained unchanged.

A. Procedure and Materials

The study was performed virtually, and all participants were asked to fill an online form. To start, participants read information and instructions about the study. This was followed by a short demographic questionnaire. Finally, they were presented with the 33 UEQ items.

Participants were explicitly asked to evaluate the "Netflix" application, but were also asked in the online form to write the name of the application they were evaluating. This was then used to validate the data (see Methods subsection).

B. Participants

163 participants were recruited during 2020 through the snowball strategy (56,4% male and 42,9% female). We shared the UEQ using social media asking for volunteers over 18 years old. They were not paid for their participation. All participants reported 100% experience using computers, and experience with the evaluated software "Netflix".

C. Methods

From the 163 completed questionnaires, we filtered out those who wrote something different than "Netflix" in the corresponding field. The questionnaires of those participants who did not complete the instrument seriously, for example, if all answers had the same value or if they were random, were also filtered out. For the latter case, we used a simple heuristic and checked the best and worst evaluations for the items in the same scale, if the difference was greater than 3 in any scale, all answers for that participant were discarded. In total, 2 questionnaires were excluded from this study, for a total of 161 valid questionnaires analysed.

To compare the new words to the original ones, we performed a series of tests first with the original set and then exchanging each of the seven items mentioned previously with its corresponding new word pair, for one-to-one comparisons, while for aggregated and average comparisons, all 7 items were substituted.

First, we compared the means, variance, standard deviation, and confidence intervals of the answers for the affected scales.

Following this, a Cronbach's Alpha Coefficient was calculated in order to measure the consistency of the scales of the new UEQ variant with the Costa Rican words. This was compared to the consistency of the scales of the original UEQ. The user experience questionnaire contains 6 scales: attractiveness, perspicuity, efficiency, dependability, stimulation and novelty, but only 4 of these (attractiveness, perspicuity, efficiency, dependability) were affected by the new proposed items. Cronbach's Alpha coefficient and confidence intervals were calculated for each of these scales according to Bonett [15].

Finally, sample sizes (precision, error probability) were used to compare both versions and factor analyses were carried out to find differences.

IV. Results

The results are split into two parts. First, the mean values of the items and the scales of the Spanish original UEQ are compared with the Costa Rican UEQ. Then, a comparison of the Cronbach's Alpha coefficients is made.

A. Results of the UEQ Comparing Mean Values

To compare the original UEQ and the new UEQ variant with the Costa Rican items, Table II shows the descriptive statistics of the original items compared to the new ones. Mean (M), standard deviation (SD) and variance (V) were calculated for the answers of the participants for each of these items. It can be seen in Table II that some mean values barely differed (Item No 28, 30, 31, 32), but other mean values lead to a noticeable difference (Item No 27, 29, 33). Thus, changes can be seen at the level of the individual items.

TABLE II. Descriptive statistics (Mean, Standard Deviation and Variance): comparison between new and original items

Item No	New			Item No	Original		
	M	SD	V		M	SD	V
27	1,5	1,2	1,3	1	2,4	0,5	0,7
28	2,1	0,7	0,9	2	2,4	0,6	0,8
29	1,8	1,1	1,2	11	0,6	1,3	1,6
30	1,9	0,9	0,7	14	1,9	1,0	1,0
31	1,9	0,9	0,9	16	2,1	0,9	0,8
32	1,2	1,5	2,3	19	1,2	1,5	2,1
33	1,9	1,2	1,4	22	1,3	1,1	1,2

An examination of the UEQ scales shows that the differences in the individual items can also result in different mean values in the overall result of the scales. Fig. 1 shows the mean values and the confidence interval with the changed items.
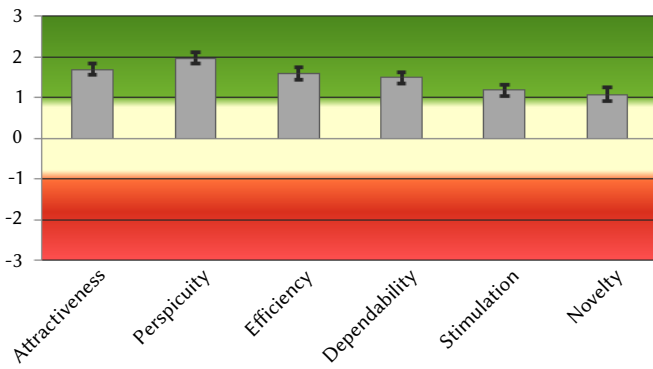
Fig. 1. Results of the evaluation of the test object "Netflix" by 161 subjects with the modified items with 5% confidence interval as error bar.

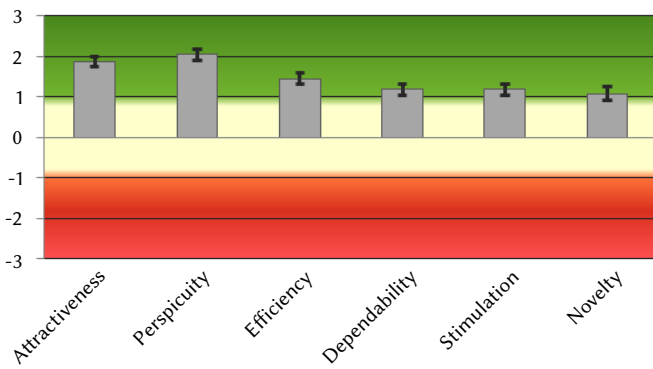The mean values and the confidence interval of the original UEQ are shown in Fig.2.

Fig. 2. Results of the evaluation of the test object "Netflix" by 161 test persons with the original Spanish version of the UEQ with 5% confidence interval as error bar.

For a more detailed comparison, the results of Fig. 1 and Fig. 2 have been combined in the Table III. Note that as previously described 3 items of the scale Attractiveness were modified (= 50%), 2 items of the scale Dependability were modified (= 50%), 1 item each of the scale Perspicuity (= 25%), and Efficiency (= 25%) were also modified. The items of the Stimulation and Novelty scales remained unchanged.

TABLE III. DESCRIPTIVE STATISTICS (MEAN, 5% CONFIDENCE): COMPARISON BETWEEN NEW AND ORIGINAL VERSION OF THE UEQ

Scale	New		Original	
	M	Conf	M	Conf
Attractiveness	1,69	0,13	1,87	0,12
Perspicuity	1,97	0,14	2,04	0,14
Efficiency	1,60	0,15	1,45	0,15
Dependability	1,49	0,14	1,18	0,13
Stimulation	1,18	0,14	1,18	0,14
Novelty	1,07	0,17	1,07	0,17

Further statistical results, in addition to the mean value, the standard deviation and the variance were also examined (see Table IV).

Since the same participants answered the questionnaire in both cases, a smaller variance can be interpreted as a better quality of a scale. According to this, the scales Attractiveness, Perspicuity, and Efficiency are better in the original version (see Table IV).

TABLE IV. DESCRIPTIVE STATISTICS (MEAN, STANDARD DEVIATION AND VARIANCE): COMPARISON BETWEEN NEW AND ORIGINAL VERSION OF THE UEQ

Scale	New			Original		
	M	SD	V	M	SD	V
Attractiveness	1,69	0,82	0,67	1,87	0,75	0,57
Perspicuity	1,97	0,93	0,86	2,04	0,88	0,78
Efficiency	1,60	0,99	0,98	1,45	0,95	0,90
Dependability	1,49	0,87	0,76	1,18	0,87	0,75
Stimulation	1,18	0,92	0,85	1,18	0,92	0,85
Novelty	1,07	1,09	1,20	1,07	1,09	1,20

By comparing the mean values, no statement can be made as to whether one of the two questionnaires is better suited to measuring "Netflix". Therefore, a comparison of the Cronbach's Alpha coefficients is made in the following section.

B. Comparison of the Cronbach's Alpha Coefficients

The value of the Cronbach's Alpha coefficient can be used as a degree of reliability. A significant higher value of the Cronbach's Alpha coefficient can be a signal for an improvement of the UEQ scales.

In Table V, the average Cronbach's Alpha coefficient is presented with 5% confidence interval for each scale.

TABLE V. AVERAGE CRONBACH'S ALPHAS AND CONFIDENCE INTERVALS FOR THE UEQ WITH NEW AND ORIGINAL ITEMS

Scale	New			Original		
	Avg. Alpha	Confidence interval		Avg. Alpha	Confidence interval	
Attractiveness	0,87	0,83	0,90	0,84	0,80	0,87
Perspicuity	0,73	0,65	0,79	0,66	0,56	0,73
Efficiency	0,66	0,57	0,74	0,63	0,52	0,71
Dependability	0,57	0,45	0,67	0,53	0,40	0,64
Stimulation	0,69	0,60	0,76	0,69	0,60	0,76
Novelty	0,72	0,64	0,78	0,72	0,64	0,78

Again, the same participants answered the questionnaire and thus a higher value for the Cronbach's Alpha coefficient can be interpreted as better reliability of a scale. All Cronbach's Alpha values have slightly improved in the new UEQ version, but are within the confidence interval of the values of the original UEQ version (see Table V).

The data of the Table V are shown as graph in Fig 3 too for easier comparison.
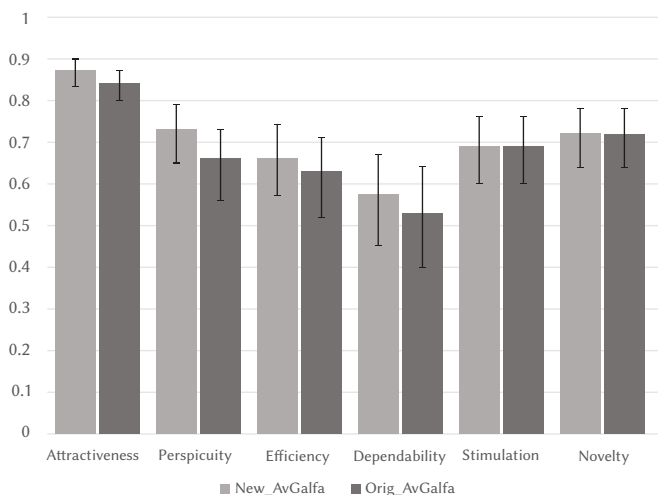

Fig. 3. Average Cronbach's Alphas and confidence intervals (shown as error bars) for the UEQ with new and original items.

A general conclusion should not be made from the slight improvement of the Cronbach's Alpha coefficients (see Fig. 3), since on the one hand the increases are only small and on the other hand only the measurement of many different products could lead to a valid statement. These differences might be also be attributed to the fact that the new items were added at the end, which might have influenced the way in which the users responded, an in-between subjects test would be required to rule this out. Additionally, since the Cronbach's Alpha is quite sensitive in a scale with only 4 items, one might expect a significant change if 50% of the items are replaced. A good description of different effects with Cronbach's Alpha can be found in the work of Schrepp [16].

Another quality for the evaluation of questionnaire results is the Precision (deviation between true scale mean in the population and the estimated scale mean from the sample) and the Error-Probability, which can be calculated with the help of the standard distribution. These values can be taken from the Excel tool for the UEQ, as can all the values mentioned above (see www.ueq-online.org). In both cases, the new UEQ variant with the Costa Rican items and the original UEQ, have the same corresponding values for Precision=0.25 and Error-Probability=0.01 (related to N=161). Although this shows that the study with 161 participants led to a trustworthy result, an improvement through the new items cannot be read from this either.

Furthermore, factor analyses were carried out and the loading of the items to the factors was considered (un-rotated, promax rotation, varimax rotation). Here, too, no noteworthy difference between the new UEQ variant with the Costa Rican items and the original UEQ could be detected, which is mainly due to the fact that when only one test item is used (in this case "Netflix"), all items primarily load on one or at most two factors. This was also expected in advance and simply means that (almost) all items fit the test object. Only the measurement of many different products would provide a higher significance here.

The main result is: translated UEQ scales are very stable against deviations (replacement of individual items by items with at least very similar meaning).

V. Conclusions and Further Work

In this study, items from the Spanish language version of the UEQ were adapted to the language culture in Costa Rica and evaluated in a study with 161 participants using the subject "Netflix". The aim of the study was to obtain an improved UEQ version for language use in Costa Rica. For this purpose, 7 item pairs were changed from the original UEQ and added to the original UEQ, so that the UEQ used in this study consisted of 33 item pairs.

Due to the widespread use of the UEQ in the Spanish-speaking community and the desire for a more culturally appropriate language version of the UEQ, this study is of great interest. But even beyond Spanish language differences, the results are interesting for all researchers and practitioners who would like to change individual items of the UEQ, as it provides the procedure to modify, add, and test new items.

We have demonstrated a procedure in which the items are not simply changed, but are appended to the original UEQ. In this way, a direct comparison is made with the same participants by conducting the evaluation with the original items on the one hand and with the changed items on the other. Thus, the effects of the changes can be directly compared with the results of the original UEQ.

In this study, we were able to show that changes to the items can lead to changed results. However, it is not possible to determine whether a modified questionnaire has a higher validity or reliability if only one product (here "Netflix") is evaluated.

It could be established that the UEQ behaves very robustly in the face of carefully implemented changes. Contrary to original expectations, the changes did not have as strong an effect as originally expected. This means that both the items of the original UEQ and the items in the new Costa Rica version were understood by the subjects. Thus, the new Costa Rican version, which uses words that are more familiar in the region, can also be used in further studies, reducing the risk of item misinterpretation by the users, although a cross-national comparison is then not possible.

It was also found that when only one product is evaluated, it is not possible to obtain statements about a clear improvement through statistical analyses. Additional studies are needed to get a clearer picture here.

In future studies, the translations of the newer UEQ+ [17] can be tested for this kind of robustness. The UEQ+ is a framework and currently provides 19 different scales, e.g. clarity [18]. From these 19 scales, a questionnaire is created that fits the product [19].

References

[1] ISO 9241-210, "Ergonomics of human-system interaction - part 210: Human-centred design for interactive systems," International Organization for Standardization, Geneva, Switzerland, 2019.

[2] J. Lazar, J.H. Feng, and H. Hochheiser, *Research methods in human-computer interaction,* 2nd ed., Glasgow, United Kingdom: Bell & Brain, 2010.

[3] M. Minge, and L. Riedel, "meCUE – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens [meCue – A modular questionnaire for capturing the user experience]," in S. *Boll, S. Maaß and R. Malaka (Ed.): Mensch und Computer [Humans and computers] 2013: Interaktive Vielfalt [Interactive diversity]*, Oldenbourg Verlag, München, pp. 89-98, 2013.

[4] J. Sauro, "SUPR-Q: A comprehensive measure of the quality of the website user experience," *Journal of Usability Studies,* vol. 10, no. 2, pp. 68–86, 2015, doi: 10.5555/2817315.2817317.

[5] B. Laugwitz, M. Schrepp, and T. Held, "Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten [Construction of a questionnaire for the measurement of user experience of software products]," in *A.M. Heinecke and H. Paul (Eds.): Mensch & Computer [Humans & computers] 2006*, Oldenbourg Verlag, pp. 125 – 134, 2006.

[6] B. Laugwitz, T. Held, and M. Schrepp , "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and Usability Engineering Group*, Heidelberg, Germany, pp. 63-76, 2008, doi: 10.1007/978-3-540-89350-9_6.

[7] M. Thielsch and M. Mooshgen, "Erfassung visueller Ästhetik mit dem VISAWI," *Usability Professionals 2011*, Stuttgart, Germany, pp. 260-265, 2011.

[8] M. Thielsch and G. Hirschfeld, "Facets of website content," *Human–Computer Interaction*, vol. 34, no.4, pp. 279–327, 2019.

[9] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, "Standards for educational and psychological testing," *American Educational Research Association*, 1999.

[10] M. Rauschenberger, M. Schrepp, S. Olschner, J. Thomaschewski, and MP. Cota, "Measurement of user experience: A Spanish Language Version of the User Experience Questionnaire (UEQ)," In: Á.J.A. Rocha, L.P.R. Calvo-Manzano, M. Pérez Cota (editors), *Information Systems and Technologies (CISTI),* Madrid, Spain, pp. 471-476, 2012.

[11] M. Rauschenberger, M. Schrepp, MP. Cota, S. Olschner, and J. Thomaschewski, "Efficient measurement of the user experience of interactive products - How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, no. 1, pp. 39-45, 2013, doi: 10.9781/ijimai.2013.215.

[12] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Construction of a Benchmark for the User Experience Questionnaire (UEQ)," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 4, pp. 40-44, 2017, doi: 10.9781/ijimai.2017.445.

[13] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, and J. Thomaschewski, "Developing a UX KPI based on the user experience questionnaire," *Computer Standards & Interfaces*, no. 65, pp. 38–44, 2019,

doi: 10.1016/j.csi.2019.01.007.

[14] International Test Commission, "The ITC Guidelines for Translating and Adapting Tests (Second edition)," 2017. Accessed: June 14 2021. [Online]. Available:www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

[15] D.B. Bonett, "Sample Size Requirements for Testing and Estimating Coefficient Alpha," Journal of Educational and Behavioral Statistics, vol. 27, no. 4, pp. 335-340, 2002, doi: 10.3102/10769986027004335.

[16] M. Schrepp, "On the Usage of Cronbach's Alpha to Measure Reliability of UX Scales," *Journal of Usability Studies*, vol. 15, no. 4, pp. 247–258, 2020.

[17] M. Schrepp, and J. Thomaschewski, "Design and Validation of a Framework for the Creation of User Experience Questionnaires," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 7, 2019, pp. 88-95, doi: 10.9781/ijimai.2019.06.006.

[18] M. Schrepp, R. Otten, K. Blum, and J. Thomaschewski, "What Causes the Dependency between Perceived Aesthetics and Perceived Usability?," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 6, pp. 78-85, 2021, doi: 10.9781/ijimai.2020.12.005.

[19] A.-L. Meiners, J. Kollmorgen, M. Schrepp, and J. Thomaschewski, "Which UX Aspects Are Important for a Software Product?," In: S. Schneegass, B. Pfleging, and D. Kern (editors): Mensch und Computer (MuC), Ingolstadt, Germany, pp. 136–139, 2021.

Mónica Hernández-Campos

MSc. Mónica Hernández-Campos obtained her Master's Degree in Cognitive Sciences at the University of Costa Rica. She is an active student of the Doctorate "Formación en la Sociedad del Conocimiento" of the University of Salamanca. She is a psychology teacher and academic advisor at the Costa Rica Institute of Technology. Her research interests are cognition, learning, and educational innovation.

Jörg Thomaschewski

Dr. 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 human-computer interaction, e-learning, and software engineering. Dr. Thomaschewski is founder of the research group "Agile Software Development and User Experience".

Yuen C. Law

Dr. Yuen C. Law obtained his PhD in computer Science from the RWTH Aachen University in 2016 and currently works at the Computer Science department at the Costa Rica Institute of Technology as docent and researcher. His interests include virtual and augmented reality applications, visualization, and human-computer interaction.