

Received March 2, 2021, accepted March 19, 2021, date of publication March 29, 2021, date of current version April 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069559

Design and Evaluation of a User Experience Questionnaire for Remote Labs

JORDI CUADROS¹, VANESSA SERRANO¹, JAVIER GARCÍA-ZUBÍA², (Senior Member, IEEE), AND UNAI HERNANDEZ-JAYO², (Member, IEEE)

¹IQS, Universitat Ramon Llull, 08017 Barcelona, Spain

²Faculty of Engineering, University of Deusto, 48007 Bilbao, Spain

Corresponding authors: Jordi Cuadros (jordi.cuadros@iqs.url.edu) and Vanessa Serrano (vanessa.serrano@iqs.url.edu)

This work was supported in part by the Aristos Campus Mundus, and in part by the Obra Social La Caixa.

ABSTRACT Remote laboratories have been in use for 25 years now. Whereas several learning oriented meta-analyses exist, validated and agreed-upon tools for assessing the user experience are not readily available. The present paper fills this gap by designing and evaluating a questionnaire focused on the needs of remote labs developers and educators using them. Building from pre-existing tools, a first version of the User eXperience Questionnaire (UXQ version 20190308) was designed to contain four scales, usability, utility, satisfaction and immersion. A total of 180 completed responses were collected from two different remote labs (VISIR and FPGA), in different campuses and in different courses to evaluate the questionnaire. The questionnaire was analyzed in terms of reliability, using Cronbach's alpha and McDonalds' omega, and validity of construct, through factor analysis. The reliability of the questionnaire and its four subscales is acceptable, but its validity should be improved. Accordingly, the questionnaire was redesigned to obtain the User eXperience Questionnaire (UXQ version 20191126), which includes three scales: usability, utility and immersion, and three questions per scale. This questionnaire was assessed using the same data. Reliability coefficients are above 0.7 and construct validity is satisfactory. A new questionnaire to evaluate the user experience in remote laboratories has been designed and validated. The questionnaire, now renamed as UXQ4RL v. 1.0, is presented and made available in this paper.

INDEX TERMS Engineering education, questionnaire, reliability, remote laboratory, student experience, validity.

I. INTRODUCTION

It has been 25 years since the first remote laboratories (RLs) were made known in [1], along with the deployment of the World Wide Web. Reference [2] identified remote experimentation as one of the most promising areas in engineering education. Nowadays, this is an active field of development and research and numerous remote experiments (REs) have already been designed in many disciplines, e.g. physics, chemistry, electronics, robotics or biology.

An RE is a real experiment that is accessed through the Internet; the student sits in a different location than the laboratory/experiment. An RL is designed to provide a similar experience to a hands-on lab and has several advantages. First, lab management is simplified because it is available 24/7. In RLs that are professionally offered and shared, teachers

are freed from having to update and maintain the system. RLs can also be shared among different institutions. In this case, activities and other educational resources can be enriched by the contributions of the different instructors. Among the disadvantages, there is a certain lack of trust and support from the faculty that has not been part of the RL design ([3] and [4]).

Nowadays, REs require designing, implementing, deploying and assessing. Whereas the first three parts mostly belong to the engineering field [5], the literature on assessing REs is limited. The assessment of learning ([6], [7], [8], and [9]) is thorough and concludes that RLs have positive effects on the teaching and learning process with an effect size equal or larger than a hands-on laboratory. Nevertheless, the data collected from the student experience, although common, are not systematically obtained and analyzed. Whereas [6] and [8] published two meta-analyses based on the results of other papers about the learning effects of the

RL, this work approaches the problem of user experience with RLs.

After an RE session or experimentation period, the use of a questionnaire to explore students satisfaction with the RL is a common practice. Questionnaires usually include subjective questions that inquire about the student experience in the RE ([10]–[12], and [13]). These surveys are intended to know whether using the RL was easy or not and if the learning activity was a satisfactory experience. Many of these questionnaires have been used but a thorough analysis of them is still missing, and the different groups in the RL community have never agreed on a shared assessment tool. In the authors' opinion, a unified questionnaire will benefit the research on RE by providing a common metric to assess the RLs and better measurement tools that will simplify the improvement processes of these educational resources.

Out of the many existing questionnaires (e.g. those cited in [6] and [8]), five are considered of the utmost importance in the RL community, given their popularity and the number of responses collected throughout the years. The first questionnaire was designed for the UNILabs project [14] by UNED, in all likelihood the strongest research group on RE [11]. Second and third are the questionnaires developed within the eMERGE [13] and the VISIR+ projects [9]. Both were constructed with the contribution of education experts. The fourth questionnaire was created for NCSLab [12] and it is the most recent example. Lastly, the questionnaire elaborated in [15] is included in this list because of its breadth. This questionnaire contains an immersion scale which has acquired considerable relevance in the RL community since then.

Table 1 summarizes the features (number of questions and scales) of these five questionnaires. It shows the use of user experience questionnaires in RLs has been inconsistent so far. Different questionnaires have different scales and suitable validation analyses are still missing.

Prior to this research, the authors used a questionnaire that built on similar scales: usefulness, sense of reality/immersion and usability [16]. Whereas the statistical analysis included in this reference is limited, it encompasses 15 years of experience of Web-Lab Deusto in designing, using and evaluating RLs.

However, in the general context of computer applications, numerous efforts have been made to develop a unique and validated questionnaire to assess usability and user experience. In this regard, the System Usability Scale (SUS; [17]) and the Usability Metric for User Experience (UMUX; [18]) questionnaires stand out. SUS is a ten-item questionnaire with a five-point Likert scale. This questionnaire was validated as a unidimensional scale [19], associated with a general concept of usability, but some researchers pointed out that its items represent two constructs, usability and learnability [20]. On the other hand, UMUX is a four-item questionnaire with a seven-point Likert scale, two of them positive and the other two negative. Later on, a variant of UMUX called UMUX-LITE, a two-item questionnaire with a seven-point Likert scale, was developed. UMUX-LITE is based on the two positive items of UMUX [21], which assess utility

and usability. UMUX-LITE items correlate with the standard and positive versions of SUS [22]. These findings indicated concurrent validity of UMUX-LITE. All these questionnaires have been standardized.

In an educational context, [23] developed and validated the E-learning Usability Questionnaire. Its third version is a 49-item questionnaire, structured as five-point Likert scales. It includes seven usability subscales: content, learning and support, visual design, navigation, accessibility, interactivity, and self-assessment and learnability. The questionnaire was thoroughly tested and attained large levels of reliability. Validity is also assessed in terms of content, criterion and construct validity, although the results for this last aspect are left open to further investigation.

TABLE 1. Features of the user satisfaction questionnaires in the RL community.

Reference	Number of questions	Scales	Questionnaire evaluation
[9]	20	3 - Perceived learning; Satisfaction with VISIR; Satisfaction with support	Face validity; Exploratory factor analysis (EFA); Scales reliability
[12]	20	3 - Technical issues; Educational issues; Added value	None
[13]	28	5 - Acceptance; English language skills; Usability; Usefulness; Self-assessed learning effect	Item difficulty; Scales reliability
[14]	15	4 - Learning Value, Value Added, Design Usability, Technology Function	None
[15]	11	3 - Effectiveness; Satisfaction; Immersion	None

However the previous questionnaires do not include all the remote lab community expectations. As shown in Table 1, [15] suggested that students' immersion may be an important scale when assessing remote labs. Immersion can be understood as "a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences" [24, p. 299]. A reference questionnaire for measuring the sense of presence, understood as "the subjective experience of being in one place or environment, even when one is physically situated in another", is the Presence Questionnaire (PQ), focused on virtual environments [25]. PQ is a 32-item questionnaire with a seven-point Likert scale. The authors inferred that PQ was an internally consistent measure with high reliability and that there was a weak but consistent positive relation between presence and task performance in virtual environments. A revisited version of this questionnaire was reported in [24].

Some other existing approaches, like those based on the technology acceptance model (TAM), have not been included because of its lack of fit to an educational context where students are asked to use a specific tool, an RL, for a limited time.

Hence, the effort to develop a better measure for assessing the user experience in RL is warranted, although any new questionnaire will need to be grounded on the questionnaires used in the RL community and on the existing efforts in assessing usability and presence in other computer applications.

The rest of this paper is organized as follows. First, we frame the scope and aims of this research. We detail next the methodology and results of the research. Finally, conclusions are reported and future lines of work are highlighted. References are listed at the very end.

II. AIM OF THE PAPER

According to the agreed need to have a common user experience evaluation tool for remote labs that builds on the previous efforts of the different groups, this paper aims to *design* and *validate* a multi-scale user experience questionnaire for assessing the use of remote labs in engineering education.

III. METHODOLOGY

To pursue the above described goals a new questionnaire was designed, and data were collected from users of two different remote labs in different courses. The data were then analyzed to evaluate the questionnaire, and a redesigned version of it was proposed.

A. QUESTIONNAIRE PREPARATION

The first attempt at creating the questionnaire object of this research, i.e. a questionnaire to accurately measure the experience of remote lab users was based on the following principles:

- The questionnaire should be easy to answer. Whenever possible, it should fit on a single page and the final version should have no more than ten questions.

- All questions should be of the same kind, a seven-point Likert scale.

- The questions and scales used should be based on pre-existing questionnaires, especially those cited in the introduction of this paper.

According to these criteria a first selection of questions on four scales, usability, usefulness, satisfaction and immersion, was agreed on. Four questions were adapted for each scale in the first iteration (UXQ version 20190308). This questionnaire is initially longer than desired to allow selecting the better performing ones for the next iteration.

B. DATA COLLECTION

Data were collected during the academic year 2018-19 from five different courses (two groups of Electronics and three groups of Physics). They used two different remote laboratories (VISIR and WebLab-Deusto-FPGA). User experience

TABLE 2. Description of the courses where UXQ version 20190308 was used.

Subject	Instructor	RL	Number of RL users	Use
Electronics	A	VISIR	46	Intensive, 4 wks
Electronics	B	VISIR	29	Intensive, 4 wks
Electronics	A + B	FPGA	39	Occasional
Physics	C	VISIR	31	Intensive, 4 wks
Physics	D	VISIR	31	Intensive, 4 wks
Physics	B	VISIR	15	Intensive, 4 wks

questionnaires were completed for each remote laboratory used, and a total of six groups of responses were obtained. These are detailed in Table 2.

VISIR was used as an integral part of the teaching methodology; all the students made use of it in class, at home and/or in the course examinations. WebLab-Deusto-FPGA was provided to the students of both Electronics courses as an additional resource and the students used it mainly at home.

Questionnaire responses were anonymously collected after having used the RL. Data were exported to a single data set for their processing and analysis.

Data collection and processing followed the current guidelines of Universidad de Deusto Ethics Committee. Given that data collection is anonymous and does not include personal information informed consent and formal approval were not required.

C. QUESTIONNAIRE ANALYSIS

From the collected data, the questionnaire is assessed in terms of its reliability and construct validity. Content validity is assumed since it is built from preexisting questionnaires in the field. The methodology described here will be used both for the initial version of the UXQ questionnaire (version 20190308) and its redesign (UXQ version 20191126). Both versions are included in the appendixes.

1) RELIABILITY

Reliability refers to the degree to which the results obtained by a measurement and procedure can be replicated. Reliability contributes to the validity of a questionnaire, but it is not a sufficient condition for its validity. One of the main aspects of reliability is the internal consistency, which measures whether different items produce similar results or not.

To measure the questionnaire's internal consistency, Cronbach's alpha coefficient is used and it is obtained from function `alpha` from `psych` R package [26]. This coefficient represents a measure of scale reliability, and analyzes how closely related the items of the questionnaire are as a group. Cronbach's alpha coefficient usually takes values from 0 to 1. The higher the value, the more reliable the questionnaire

is. Values of 0.7 or higher are considered acceptable [27]. A confidence interval for Cronbach's alpha coefficient at a 0.95 confidence level [28] is also included in the analysis.

A second way to understand the internal consistency of the questionnaire is as an indicator that all items are measuring the same latent variable. Two different measures can be used, McDonald's omega (ω) and hierarchical omega (ω_h) coefficients [29], which are recommended because of their applicability to a broader variety of factor models [30]. However, debate regarding which coefficient fits better to describe the questionnaires' consistency seems to be still open, as shown in the literature ([30]–[32], and [33]).

Coefficient omega ranges from 0 to 1 as well and is equivalent to Cronbach's alpha coefficient for instruments for which a single factor model is appropriate and where each item is associated with the common construct to the same degree [30]. The construct is a latent variable that is not directly observable but inferred from the observable variables that are responses to items. If items are associated with the common construct to different degrees, then omega is preferred. The same reference suggests that the hierarchical omega coefficient may be more appropriate where data are used to calculate a weighted composite score. Hierarchical omega also ranges from 0 to 1 and, like the omega coefficient, it is equivalent to alpha coefficient. Both omega and hierarchical omega coefficients can be obtained from function `omega` from `psych` R package as `omega total` in the reliability coefficient output. Confidence intervals for both coefficients, at a 0.95 confidence level, are calculated with a non-parametric bias-corrected bootstrapping procedure [34] using 500 replicates.

Because of the open debate related to the best coefficient to describe the questionnaire's internal consistency described previously earlier, in this study all these coefficients are calculated using `psych` R package according to the functions implemented. The calculations are done both for the whole questionnaire and for each subscale (usability, satisfaction, immersion, and utility).

2) VALIDITY

Validity expresses the degree to which a measurement measures what it purports to measure. The construct validity is analyzed using a factorial analysis, which, in its turn, is used for multi scales questionnaires if, based on the correlation between items, the initial structure of the questionnaire is reproduced.

To measure the questionnaire's construct validity, several complementary analyses are implemented. Firstly, a correlogram is displayed to visually analyze the correlation between items and identify if the theoretical structure of the questionnaire can be reproduced. The correlogram is obtained from function `corrplot` from `corrplot` R package.

To examine the validity more in-depth, a factor analysis, which is an analytic technique to identify latent factors (questionnaire constructs) is carried out defining and measuring some indicators (questionnaire items). There are

two basic approaches to the investigation of underlying constructs, respectively, the exploratory and the confirmatory approach. The exploratory factor analysis (EFA) detects the constructs that underlie a questionnaire based on the correlations between its items. In contrast, when a specific theoretical model is known, a confirmatory factor analysis (CFA) is used. Both approaches are used in this paper.

For EFA, the `omega` function from `psych` R package is used. It generates, like an omega chart, the representation of a minimum residual factor analysis, with an oblimin rotation [35] and a Schmid-Leiman transformation [26]. This analysis renders a set of loadings that quantify the relationship between the factors and the items and can be interpreted as regression coefficients. Results are shown graphically. On the left side, it presents the loadings of each question on the general factor. On the right side, it shows the loadings on the extracted factors. Only loadings with values over 0.2 are considered relevant and are included in the chart.

For CFA, the theoretical model is based on the initial distribution of the items per scales, specifically, the addition of the items on each scale. To identify if the theoretical model per scale and its coefficients are significant, the `cfa` function from `lavaan` R package is used. Using this function, the comparative fit index (CFI) is also obtained. This coefficient ranges between 0 and 1 and is used to compare the sample covariance matrix to the theoretical model. Values close to 1 indicate a good fitting and a reference of 0.9 is usually expected [36].

D. REDESIGN

Based on the methodology described before, the initial questionnaire structure was redesigned to simplify it and improve its consistency. The redesigned questionnaire was then analyzed using the same set of data, although only the answers to the selected items are considered.

IV. RESULTS

In accordance with the methodology discussed in the previous section, in the next subsection the results are presented: first, we discuss the design of the first version of the User eXperience Questionnaire (UXQ); second, we describe the data collected; third, we evaluate the psychometric features (reliability and validity) of this questionnaire; fourth, we discuss the improvement of the questionnaire; and last, we analyze the improved version of the UXQ.

A. COLLECTION OF QUESTIONS AND SCALES AND SELECTION OF ITEMS FOR THE USER EXPERIENCE QUESTIONNAIRE

The first version of the User eXperience Questionnaire (UXQ version 20190308) was designed based on the experiences collected from the usability and user perception assessment in the remote lab community and the examples collected from the presence and usability scales described above.

It was designed to include four scales and four items per scale, with the explicit intention of shortening the questionnaire in a second version by selecting the best questions for

TABLE 3. Example of a question of UXQ version 20190308.

	Strongly disagree <i>Muy en desacuerdo</i>		Neutral <i>Neutral</i>			Strongly agree <i>Muy de acuerdo</i>	
The remote lab is easy to use. <i>El laboratorio remoto es fácil de usar.</i>	1	2	3	4	5	6	7

each scale. To facilitate the distribution of the questionnaire, it was drawn up both in English and in the students’ language, Spanish. All the questions were formatted as Likert scales with seven levels of response. An example is shown in Table 3.

The four scales are usability, utility, immersion and satisfaction. Usability and utility are understood, like in the UMUX scales, as ease of use and fitness to purpose respectively. Immersion follows the definition found in [24, p. 299]

“*Involvement* is a psychological state experienced as a consequence of focusing one’s mental energy and attention on a coherent set of stimuli or meaningfully related activities or events [...] *Immersion* is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences”

The last scale, satisfaction, includes more general user perception questions that cannot be related to the previous scales.

Table 4 summarizes the sixteen questions, four per scale, included in this version of the questionnaire and their justifications.

These 16 questions were randomly distributed to construct the first version of the UXQ questionnaire (UXQ version 20190308). The questionnaire is included in Appendix A. Table 5 connects the questionnaire items to each scale.

B. DATA COLLECTION WITH UXQ VERSION 20190308

This version of the questionnaire was used in five courses in academic year 2018-19. Responses were collected from 191 remote lab users. The experience involved two different remote labs (VISIR and FPGA), on different campuses and different instructors took part.

A total of 191 questionnaire responses were collected. Out of these, 180 responses were complete. Only these responses were used to assess the psychometric features of the questionnaire (i.e reliability and construct validity). Figure 1 shows the distribution of the responses to the items of the questionnaire for the 180 complete responses.

C. RELIABILITY AND VALIDITY OF THE QUESTIONNAIRE

The reliability and validity of the questionnaire was analyzed using these 180 responses. The reliability, like Cronbach’s alpha, of the complete questionnaire is 0.91 ± 0.02 (95 % confidence level). Assuming a four factor model,

TABLE 4. Questions in UXQ version 20190308. The questionnaire includes the text of each question in English and Spanish. Only the English version is shown here.

Scale	Question	Justification
Usability	The remote lab is easy to use.	Adapted from UMUX [18] and UMUX-LITE [21]. Also included in [16].
Usability	The remote lab interface is easy to read.	Adapted from the E-learning Usability Questionnaire v.3 [23].
Usability	Users can predict the result of using each element in the interface.	Adapted from the E-learning Usability Questionnaire v.3 [23].
Usability	The application layout is sufficiently apparent so that help is not needed to use it.	Adapted from the E-learning Usability Questionnaire v.3 [23].
Utility	The application meets my requirements.	Adapted from UMUX [18] and UMUX-LITE [21].
Utility	The remote lab helps me learn.	Based on several questions on learning value appearing in [14] and [12].
Utility	The remote lab will help me pass the course.	Added following the experience of the authors in assessing other types of educational activities.
Utility	The application provides opportunities to practice knowledge.	Adapted from the E-learning Usability Questionnaire v.3 [23].
Immersion	Using the application feels like using a real lab.	Adapted from [16].
Immersion	My interactions with the application seem real.	Adapted from the Presence Questionnaire (PQ) [24].
Immersion	When working on the remote lab I concentrate on the assigned tasks.	Adapted from the Presence Questionnaire (PQ) [24].
Immersion	I was engaged in using the remote lab..	Adapted from the Presence the Questionnaire (PQ) [24].
Satisfaction	The application is enjoyable to use.	Adapted from the E-learning Usability Questionnaire v.3 [23].
Satisfaction	I am satisfied with the remote lab.	Adapted from a question used in [16].
Satisfaction	Using the application stimulates learning.	Adapted from the E-learning Usability Questionnaire v.3 [23].
Satisfaction	The remote lab assists learners to have positive feelings about the course..	Adapted from the E-learning Usability Questionnaire v.3 [23].

omega hierarchical gives 0.73 ± 0.11 and omega total equals 0.94 ± 0.02 .

For each of the four scales that were included in UXQ version 20190308: usability, utility, immersion and satisfaction, reliability was also assessed. Results are shown in Table 6.

TABLE 5. Distribution of the items per scale. Items are encoded according to their order in the UXQ version 20190308.

Scale	Items
Usability	Q01, Q08, Q11, Q12
Utility	Q04, Q06, Q14, Q16
Immersion	Q03, Q05, Q10, Q13
Satisfaction	Q02, Q07, Q09, Q15

TABLE 6. Reliability measures (95 % confidence level) for the scales of UXQ version 20190308.

Scale	Alpha	Omega total
Usability	0.74 ± 0.06	0.77 ± 0.08
Utility	0.82 ± 0.04	0.83 ± 0.06
Immersion	0.72 ± 0.07	0.75 ± 0.07
Satisfaction	0.83 ± 0.04	0.83 ± 0.06

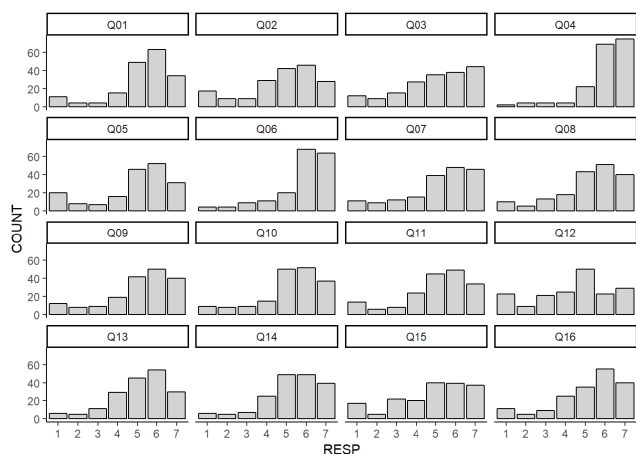


FIGURE 1. Distribution of the responses to the questionnaire.

According to the commonly used threshold ($\alpha \geq 0.7$), the questionnaire provides a reliable measure of the user experience both in general and with regard to the four scales.

In terms of validity, content validity is assumed within the questionnaire process as it includes all aspects considered in previous surveys. Construct validity is assessed by analyzing the correlation matrix and a confirmatory factor analysis (CFA).

Figure 2 shows the correlogram of the items in the questionnaire. The items are ordered according to similarity and four clusters are extracted by hierarchical clustering. The questions clusters formed from the responses do not match the designed scales.

A CFA of these data, indicates that the model and all its coefficients are significant (95 % confidence level). However, the comparative fit index (CFI) is 0.83 when a reference of at least 0.9 is usually expected as explained earlier. Hence trying

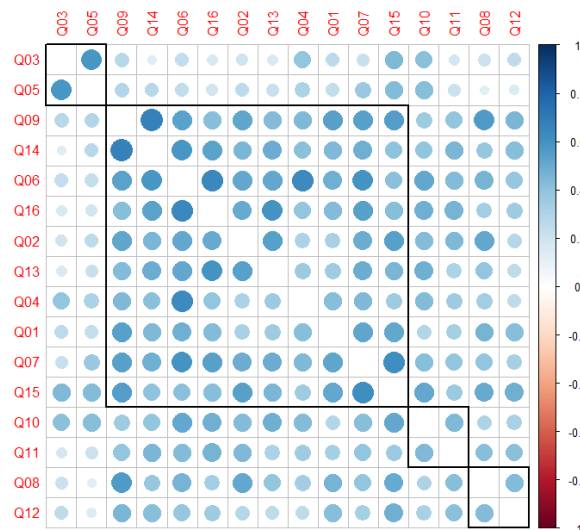


FIGURE 2. Correlogram of the questionnaire items.

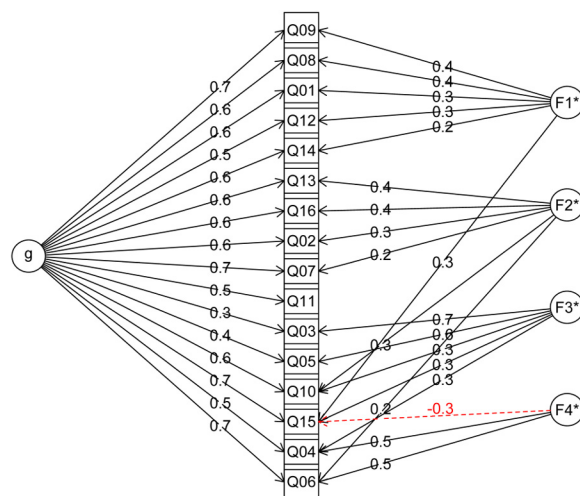


FIGURE 3. EFA for four subscales and a global factor.

to improve the questionnaire may be convenient according to this index.

To better understand the structure behind the responses to the questionnaire, exploratory factor analyses (EFAs) are run for four and three factors (Figures 3 and 4 respectively).

Figure 3 shows that the fourth factor of the model is mixed with the other factors of the questionnaire and suggests that a three factor model may be appropriate.

The analysis assuming three subscales shows that the three resulting factors can be distinguished. The first factor (F1 in Figure 4) is related to the questions of the utility subscale (Q04, Q06, Q14, Q16 and some other questions coming from other scales). The second factor (F2) varies in most of the questions in the usability scale (Q01, Q08 and Q12). Last, the third factor (F3) is related to questions on the immersion scale (Q03, Q05 and Q10). According to these observations, it may be worth attempting to redesign the questionnaire using only these three scales and selecting the questions that

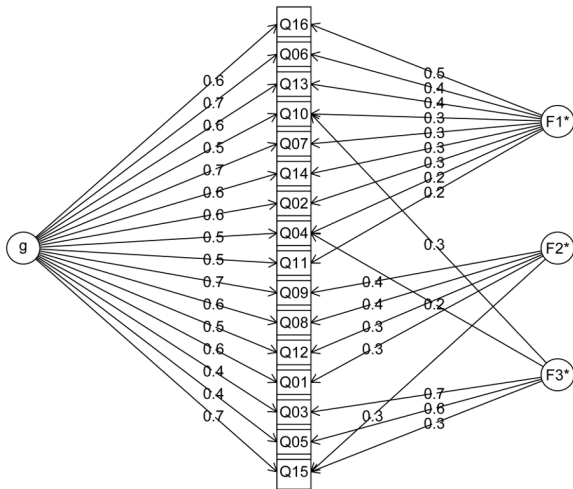


FIGURE 4. EFA for three subscales and a global factor.

TABLE 7. Distribution of the items per scales in the improved questionnaire. Questions are identified as in UXQ version 20190308.

Scale	Items
Usability	Q01, Q08, Q12
Utility	Q06, Q14, Q16
Immersion	Q03, Q05, Q10

showed a better correlation to these empiric factors (F1, F2, F3) from the original scales.

D. REDESIGN OF THE QUESTIONNAIRE

In the next step of the questionnaire development, a subset of questions should be selected to obtain a simpler questionnaire that could potentially have an even better performance in reliability and validity. Following the previous results, only three scales will be considered: usability, utility and immersion. The satisfaction scale is eliminated.

Each scale is completed to three items to produce a questionnaire with three scales and nine questions that maps to these scales, three questions per scale (Table 7).

This second version of the UXQ questionnaire (UXQ version 20191126) is included in Appendix B.

E. RELIABILITY AND VALIDITY OF THE 9-ITEM USER EXPERIENCE QUESTIONNAIRE

The psychometric features of this second version of the questionnaire are assessed from the 180 complete responses collected in the initial questionnaire. The reliability is 0.83 ± 0.04 (95 % confidence level) like Cronbach’s alpha. For a three factor model, omega hierarchical is 0.63 ± 0.21 and omega total equals 0.88 ± 0.04 .

Table 8 summarizes the reliability for each of the three scales included in this shortened version of the questionnaire (UXQ version 20191126).

According to the commonly used threshold ($\alpha \geq 0.7$), the questionnaire provides a reliable measure of the user

TABLE 8. Reliability measures (95 % confidence level) for the scales of UXQ version 20191126.

Scale	Alpha	Omega total
Usability	0.70 ± 0.08	0.74 ± 0.09
Utility	0.81 ± 0.05	0.81 ± 0.08
Immersion	0.73 ± 0.07	0.75 ± 0.07

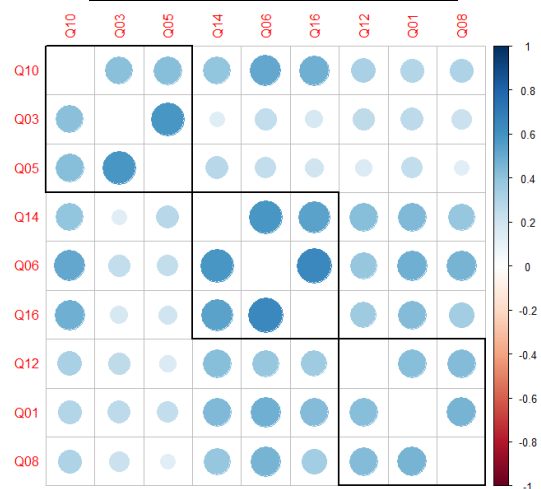


FIGURE 5. Correlogram of the redesigned questionnaire.

experience both in general and with respect to the three scales. These results must be read attentively since they assume independence of the items and their features are not affected by their location in the questionnaire.

In terms of validity, construct validity is assessed by analyzing the correlation matrix and a confirmatory factor analysis (CFA) using the same criteria as above.

Figure 5 shows the correlogram of the items in the questionnaire. The items are ordered according to similarity and three clusters are extracted by hierarchical clustering. The obtained clusters match the designed scales of usability, utility and immersion.

Likewise, the CFA gives significant coefficients at the 95 % confidence level. The comparative fit index (CFI) is 0.93 which exceeds the common reference of 0.9. Hence this questionnaire appears to be valid in terms of construct validity.

V. CONCLUSION

A first version of the User eXperience Questionnaire (UXQ version 20190308) was designed. It contained four scales, usability, utility, satisfaction and immersion, and four items per scale. All items were supported by previous experiences and existing literature, and were formatted as Likert scales with seven levels of response.

To evaluate this first version of the questionnaire a total of 180 completed responses were collected from two different remote labs (VISIR and FPGA), on different campuses and in different subjects. Reliability of the questionnaire was

assessed, both globally and per scale, using Cronbach's alpha and McDonald's omega coefficients. All results are above 0.7 and hence satisfactory. Construct validity was assessed exploring the correlation matrix and by EFA and CFA. According to these analyses, questionnaire responses are not in agreement with the designed questionnaire structure. EFA suggests that a three factor model may fit the responses to the questionnaire better.

Therefore, the questionnaire was redesigned to obtain a new version of the User eXperience Questionnaire (UXQ version 20191126) with three scales, usability, utility and immersion. It was completed to obtain three questions per scale.

This second version of the questionnaire was assessed using the same data. Reliability coefficients for the simplified questionnaire (alpha and omega total) are above 0.7 and construct validity is satisfactory.

This research has been able to provide a short and validated questionnaire that should be useful to assess the user experience in remote labs. This questionnaire, now renamed to UXQ4RL v. 1.0, is available in Appendix C and on the authors' website, <http://asistembe2.iqs.url.edu>.

VI. FUTURE LINES OF RESEARCH

Further investigations will be needed to validate this last version of the questionnaire in a real application and within different contexts of use in remote experimentation. Additionally, it could be worth exploring the possibility of summarizing the results of the questionnaires in a set of indices that allows comparing different remote labs user experiences.

APPENDIX

Questionnaires UXQ version 20190308 (Appendix A), UXQ version 20191126 (Appendix B), and UXQ4RL v. 1.0 (Appendix C) are included as supplementary material to this paper to preserve their format as designed.

REFERENCES

- [1] C. Bohus, L. A. Crowl, B. Aktan, and M. H. Shor, "Running control engineering experiments over the Internet," *IFAC Proc. Volumes*, vol. 29, no. 1, pp. 2919–2927, Jun./Jul. 1996, doi: [10.1016/S1474-6670\(17\)58121-5](https://doi.org/10.1016/S1474-6670(17)58121-5).
- [2] J. E. Froyd, P. C. Wankat, and K. A. Smith, "Five major shifts in 100 years of engineering education," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1344–1360, May 2012, doi: [10.1109/JPROC.2012.2190167](https://doi.org/10.1109/JPROC.2012.2190167).
- [3] H. A. Lahoud and J. P. Krichen, "Networking labs in the online environment: Indicators for success," *J. Technol. Stud.*, vol. 36, no. 2, pp. 31–40, May 2010, doi: [10.21061/jots.v36i2.a.4](https://doi.org/10.21061/jots.v36i2.a.4).
- [4] P. Orduña, L. Rodríguez-Gil, J. García-Zubia, I. Angulo, U. Hernandez, and E. Azcuenaga, "LabsLand: A sharing economy platform to promote educational remote laboratories maintainability, sustainability and adoption," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Erie, PA, USA, Oct. 2016, pp. 1–6, doi: [10.1109/FIE.2016.7757579](https://doi.org/10.1109/FIE.2016.7757579).
- [5] J. García-Zubia and G. R. Alves, Eds., *Using Remote Labs in Education: Two Little Ducks in Remote Experimentation*. Bilbao, Spain: Universidad de Deusto, 2011. [Online]. Available: <http://www.deusto-publicaciones.es/deusto/pdfs/otraspub/otraspub01.pdf>
- [6] J. R. Brinson, "Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research," *Comput. Educ.*, vol. 87, pp. 218–237, Sep. 2015, doi: [10.1016/j.compedu.2015.07.003](https://doi.org/10.1016/j.compedu.2015.07.003).
- [7] J. García-Zubia, J. Cuadros, S. Romero, U. Hernandez-Jayo, P. Orduña, M. Guenaga, L. Gonzalez-Sabate, and I. Gustavsson, "Empirical analysis of the use of the VISIR remote lab in teaching analog electronics," *IEEE Trans. Educ.*, vol. 60, no. 2, pp. 149–156, May 2017, doi: [10.1109/TE.2016.2608790](https://doi.org/10.1109/TE.2016.2608790).
- [8] J. Ma and J. V. Nickerson, "Hands-on, simulated, and remote laboratories: A comparative literature review," *ACM Comput. Surv.*, vol. 38, no. 3, p. 7, Sep. 2006, doi: [10.1145/1132960.1132961](https://doi.org/10.1145/1132960.1132961).
- [9] C. Viegas, A. Pavani, N. Lima, A. Marques, I. Pozzo, E. Dobbolletta, V. Atencia, D. Barreto, F. Calliari, A. Fidalgo, D. Lima, G. Temporão, and G. Alves, "Impact of a remote lab on teaching practices and student learning," *Comput. Educ.*, vol. 126, pp. 201–216, Nov. 2018, doi: [10.1016/j.compedu.2018.07.012](https://doi.org/10.1016/j.compedu.2018.07.012).
- [10] J. García-Zubia, U. Hernandez, I. Angulo, P. Orduña, and J. Irurzun, "Acceptance, usability and usefulness of WebLab-Deusto from the students point of view," *Int. J. Online Biomed. Eng.*, vol. 5, no. 1, pp. 9–15, 2009, doi: [10.3991/ijoe.v5i1.817](https://doi.org/10.3991/ijoe.v5i1.817).
- [11] R. Heradio, L. de la Torre, D. Galan, F. J. Cabrerizo, E. Herrera-Viedma, and S. Dormido, "Virtual and remote labs in education: A bibliometric analysis," *Comput. Educ.*, vol. 98, pp. 14–38, Jul. 2016, doi: [10.1016/j.compedu.2016.03.010](https://doi.org/10.1016/j.compedu.2016.03.010).
- [12] M. Hu, W. Hu, and H. Zhou, "NCSLab remote experimentation for control engineering education in Wuhan University," in *Proc. 35th Chin. Control Conf. (CCC)*, Chengdu, China, Jul. 2016, pp. 7416–7421, doi: [10.1109/ChiCC.2016.7554532](https://doi.org/10.1109/ChiCC.2016.7554532).
- [13] D. Lang, C. Mengelkamp, R. S. Jäger, D. Geoffroy, M. Billaud, and T. Zimmer, "Pedagogical evaluation of remote laboratories in eMerge project," *Eur. J. Eng. Educ.*, vol. 32, no. 1, pp. 57–72, Mar. 2007, doi: [10.1080/03043790601055626](https://doi.org/10.1080/03043790601055626).
- [14] S. Dormido, J. Sánchez, H. Vargas, L. De la Torre, and R. Heradio, "UNED labs: A network of virtual and remote laboratories," in *Using Remote Labs in Education: Two Little Ducks in Remote Experimentation*. Bilbao, Spain: Universidad de Deusto, 2011, ch. 2, pp. 253–270.
- [15] J. E. Corter, J. V. Nickerson, S. K. Esche, and C. Chassapis, "Remote versus hands-on labs: A comparative study," in *Proc. 34th Annu. Frontiers Educ. (FIE)*, Savannah, GA, USA, Oct. 2004, pp. FIG-17, doi: [10.1109/FIE.2004.1408586](https://doi.org/10.1109/FIE.2004.1408586).
- [16] J. García-Zubia, P. Orduña, I. Angulo, U. Hernandez, O. Dziabenko, D. Lopez-Ipina, and L. Rodríguez-Gil, "Application and user perceptions of using the WebLab-Deusto-PLD in technical education," in *Proc. 1st Global Online Lab. Consortium Remote Lab. Workshop*, Rapid City, SD, USA, Oct. 2011, pp. 1–6, doi: [10.1109/GOLC.2011.6086780](https://doi.org/10.1109/GOLC.2011.6086780).
- [17] J. Brooke, "SUS—A 'quick and dirty' usability scale," in *Usability Evaluation in Industry*. London, U.K.: CRC Press, 1996, ch. 21.
- [18] K. Finstad, "The usability metric for user experience," *Interacting Comput.*, vol. 22, no. 5, pp. 323–327, Sep. 2010, doi: [10.1016/j.intcom.2010.04.004](https://doi.org/10.1016/j.intcom.2010.04.004).
- [19] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *Int. J. Hum.-Comput. Interact.*, vol. 24, no. 6, pp. 574–594, Jul. 2008, doi: [10.1080/10447310802205776](https://doi.org/10.1080/10447310802205776).
- [20] J. R. Lewis and J. Sauro, "The factor structure of the system usability scale," in *Human Centered Design*. Berlin, Germany: Springer-Verlag, 2009, pp. 94–103, doi: [10.1007/978-3-642-02806-9_12](https://doi.org/10.1007/978-3-642-02806-9_12).
- [21] J. R. Lewis, B. S. Utesch, and D. E. Maher, "UMUX-LITE: When there's no time for the SUS," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, Apr. 2013, pp. 2099–2102, doi: [10.1145/2470654.2481287](https://doi.org/10.1145/2470654.2481287).
- [22] M. I. Berkman and D. Karahoca, "Re-assessing the usability metric for user experience (UMUX) scale," *J. Usability Stud.*, vol. 11, no. 3, pp. 89–109, May 2016.
- [23] P. Zaharias and A. Poylymenakou, "Developing a usability evaluation method for e-learning applications: Beyond functional usability," *Int. J. Hum.-Comput. Interact.*, vol. 25, no. 1, pp. 75–98, Jan. 2009, doi: [10.1080/10447310802546716](https://doi.org/10.1080/10447310802546716).
- [24] B. G. Witmer, C. J. Jerome, and M. J. Singer, "The factor structure of the presence questionnaire," *Presence*, vol. 14, no. 3, pp. 298–312, Jun. 2005, doi: [10.1162/105474605323384654](https://doi.org/10.1162/105474605323384654).
- [25] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, Jun. 1998, doi: [10.1162/105474698565686](https://doi.org/10.1162/105474698565686).
- [26] W. Revelle. (Oct. 5, 2020). *Psych: Procedures for Psychological, Psychometric, and Personality Research R Package Version 2.0.9*. [Online]. Available: <https://CRAN.R-project.org/package=psych>

- [27] J. R. A. Santos, "Cronbach's alpha: A tool for assessing the reliability of scales," *J. Extension*, vol. 37, no. 2, pp. 1–5, Apr. 1999. [Online]. Available: <https://www.joe.org/joe/1999april/tt3.php>
- [28] L. S. Feldt, D. J. Woodruff, and F. A. Salih, "Statistical inference for coefficient alpha," *Appl. Psychol. Meas.*, vol. 11, no. 1, pp. 93–103, Mar. 1987, doi: [10.1177/014662168701100107](https://doi.org/10.1177/014662168701100107).
- [29] R. P. McDonald, "Generalizability in factorable domains: 'Domain validity and generalizability' 1," *Educ. Psychol. Meas.*, vol. 38, no. 1, pp. 75–79, Apr. 1978, doi: [10.1177/001316447803800111](https://doi.org/10.1177/001316447803800111).
- [30] R. Komperda, T. C. Pentecost, and J. Barbera, "Moving beyond alpha: A primer on alternative sources of single-administration reliability evidence for quantitative chemistry education research," *J. Chem. Educ.*, vol. 95, no. 9, pp. 1477–1491, Jul. 2018, doi: [10.1021/acs.jchemed.8b00220](https://doi.org/10.1021/acs.jchemed.8b00220).
- [31] W. Revelle and D. M. Condon, "Reliability from α to ω : A tutorial," *Psychol. Assessment*, vol. 31, no. 12, pp. 1395–1411, Dec. 2019, doi: [10.1037/pas0000754](https://doi.org/10.1037/pas0000754).
- [32] C. Viladrich, A. Angulo-Brunet, and E. Doval, "A journey around alpha and omega to estimate internal consistency reliability," *Anales de Psicología/Ann. Psychol.*, vol. 33, no. 3, pp. 755–782, Jul. 2017, doi: [10.6018/analesps.33.3.268401](https://doi.org/10.6018/analesps.33.3.268401).
- [33] R. E. Zinbarg, W. Revelle, I. Yovel, and W. Li, "Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternative conceptualizations of reliability," *Psychometrika*, vol. 70, no. 1, pp. 123–133, Apr. 2005, doi: [10.1007/s11336-003-0974-7](https://doi.org/10.1007/s11336-003-0974-7).
- [34] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statist. Sci.*, vol. 1, no. 1, pp. 54–75, 1986, doi: [10.1214/ss/1177013815](https://doi.org/10.1214/ss/1177013815).
- [35] C. A. Bernaards and R. I. Jennrich, "Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis," *Educ. Psychol. Meas.*, vol. 65, no. 5, pp. 676–696, Oct. 2005, doi: [10.1177/0013164404272507](https://doi.org/10.1177/0013164404272507).
- [36] D. Hooper, J. Coughlan, and M. Mullen, "Structural equation modelling: Guidelines for determining model fit," *Electron. J. Bus. Res. Methods*, vol. 6, no. 1, pp. 53–60, Jan. 2008, doi: [10.21427/D7CF7R](https://doi.org/10.21427/D7CF7R).



VANESSA SERRANO received the degree in mathematics and the Diploma degree in statistics from the Universitat Autònoma de Barcelona, Spain, in 2005, and the Ph.D. degree in economics and business from IQS, Universitat Ramon Llull (URL), Barcelona, in 2015.

She is currently an Assistant Professor with the Quantitative Methods Department, URL. Her main research interest includes the development and the analysis of the use of interactive applications for learning in STEM through the collection of traces by data mining techniques. She is also a member of the ASISTEMBE Research Group.



JAVIER GARCÍA-ZUBÍA (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Deusto, Spain.

He is currently a Full Professor with the Faculty of Engineering, University of Deusto. He is also the Leader of the WebLab-Deusto Research Group. His research interests include remote laboratory design, implementation, and evaluation.



JORDI CUADROS received the B.Sc. and Ph.D. degrees in chemistry from IQS, Universitat Ramon Llull (URL), Barcelona, Spain, in 1997 and 2003, respectively, and the degree in education from Universidad Nacional a Distancia, Madrid, Spain, in 2010.

He spent two years as a Postdoctoral Fellow with the Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Associate Professor with the Quantitative

Methods Department, URL, where he leads the ASISTEMBE Research Group. His current research interests include analyzing the use of simulations for teaching and learning physical sciences at the university level by applying learning analytics.

Dr. Cuadros is also a member of the Catalan Chemical Society, the American Chemical Society, and its Division of Chemical Education.



UNAI HERNANDEZ-JAYO (Member, IEEE) was born in Barakaldo, Spain, in 1978. He received the M.S. and Ph.D. degrees in telecommunications engineering from the University of Deusto, Bilbao, Spain, in 2001 and 2012, respectively.

In 2004, he joined the University of Deusto, where he is currently an Assistant Professor with the Telecommunications Department, teaching classes in electronic design and communications electronics. His current research interests include the use of information and communication technologies in the educational process and the application of information and communication technologies in cooperative vehicular systems.

...