



Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale

Guillaume Gronier & Alexandre Baudet

To cite this article: Guillaume Gronier & Alexandre Baudet (2021) Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale, International Journal of Human-Computer Interaction, 37:16, 1571-1582, DOI: [10.1080/10447318.2021.1898828](https://doi.org/10.1080/10447318.2021.1898828)

To link to this article: <https://doi.org/10.1080/10447318.2021.1898828>



© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 17 Mar 2021.



Submit your article to this journal [↗](#)



Article views: 1017



View related articles [↗](#)



View Crossmark data [↗](#)

Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale

Guillaume Gronier and Alexandre Baudet

Information Technology for Innovative Sciences (ITIS), Luxembourg Institute of Science and Technology (LIST), Esch-sur-Alzette, Luxembourg

ABSTRACT

While the System Usability Scale (SUS) is probably one of the most widely used questionnaires to measure the perceived ease of use of interactive systems, there is currently no scientific valid translation in French. This article describes the translation and statistical validation of the French version of the SUS, called F-SUS. On the basis of two translations carried out by a committee of bilingual experts, the various psychometric analyses made it possible to select only one translation. Fidelity measurement, factor analysis and sensitivity measurement obtained results very close or similar to the original version of the SUS. Thus, the F-SUS can be used with confidence by French-speaking usability researchers and practitioners.

1. Introduction

Among the methods that allow the evaluation of the user experience (UX), “standardized” questionnaires are often used to capture the user’s satisfaction of a product (Maguire, 2001). The “standardized” character of these questionnaires sets them apart from “home-made” questionnaires (Hornbaek, 2006; Sauro & Lewis, 2012), due to the scientific validation that has been applied to them. This validation is mainly based on two psychometric measurements: validity and reliability (Drost, 2011; Nunnally, 1978a; Nunnally, 1978b; Peterson, 2013). Validity refers to the meaning of the research components (Drost, 2011). For example, when researchers measure behaviors, they focus on knowing if they are measuring what they intended to measure. Does an IQ test measure intelligence? Does such a usability measurement questionnaire allow the perceived ease of use of a product to be measured? These are validity questions. As for reliability, this validates the repeatability of the measurements taken with the help of a questionnaire. In this way, the questionnaire should allow us to obtain the same scores when different people take the measurements, on different occasions, in different conditions, with presumably alternative instruments that measure the same thing (Drost, 2011). In other words, reliability is the consistency with the measurements, or the stability of the measurements in different conditions in which we should essentially obtain the same results (Nunnally, 1978b).

Consequently, a standardized questionnaire is a questionnaire devised for repeated use, presenting satisfactory psychometric scores for validity and reliability, generally organized as a specific set of questions presented in a precise order according to a pre-defined format, with specific rules to

produce measurements based on the responses of the participants (Sauro & Lewis, 2012).

In the user experience (UX) domain, standardized questionnaires are part of a general framework of human-centered iterative design processes (ISO 9241-210:2010, 2010), which aim to improve the user experience of a product by making it correspond to the needs, expectations and specificities of the final target users. The ISO 13407 1999 norm (ISO 13407:1999, 1999) references questionnaires among the 12 most frequently used design methods. The norm defines them as indirect evaluation methods, which collect, by means of predefined questionnaires, the opinions of users on the interface. Hornbaek (2006) also drew up a list of questionnaires to record user satisfaction and measure the perceived usability of a system. Maguire (2001) adds that satisfaction questionnaires capture the subjective impressions of users regarding their experience with a product.

Today, we are faced with numerous standardized questionnaires measuring the UX. Most of them have been subject to a scientific validation which has been published. One of the most popular standardized questionnaires measuring perceived usability is probably the System Usability Scale (SUS), created originally as a “quick and dirty” scale (Brooke, 1996). It is this questionnaire that we have chosen to translate into French and we will come back to its history and presentation in more detail in the following section of this article.

The SUS has been the inspiration for several other questionnaires measuring the UX. We can cite the Usability Metric for User Experience (UMUX) (Finstad, 2010), which was

designed to be a more concise version of the SUS, using 4 items from it that were reformulated and for which the response methods were slightly changed. In this way, the UMUX uses every other inverted item, but modifies the forms of response to go from a 5-point Likert scale to a 7-point scale. Finstad (2010) finds a correlation between the SUS and the UMUX of above 0.80, while in a confirmatory study, Berkman and Karahoca (2016) measure a correlation of 0.74. These two values remain very high and indicate that the psychometric properties of the UMUX are very close to those of the SUS. In a similar approach, other authors have created the UMUX-LITE questionnaire (Lewis et al., 2013), which is also based on the SUS but which seeks to reduce the number of items even more than the UMUX. Thus, the UMUX-LITE uses the two odd UMUX items, that is, the non-inverted items. Its correlation with the SUS, calculated by the authors, is 0.81.

As with the SUS, other UX questionnaires were created during the 1990s to measure the instrumental quality of a system. We propose a brief overview of them in order to position the SUS against other questionnaires measuring user experience. This is particularly the case with the Computer System Usability Questionnaire (CSUQ) (Lewis, 1995), a questionnaire containing 19 items to measure the usefulness of a system, the quality of its information and the quality of its interface. The CSUQ is based on the Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1995), from which it takes the items but formulates them differently: the PSSUQ was designed to be used after a user test in order to measure user satisfaction with the system he/she was confronted with, while the CSUQ wants to appear more generic and adapted to less controlled situations. Thus, for example, the PSSUQ uses the formulation “this system was easy to use” while the CSUQ uses “this system is easy to use”. These two questionnaires complement the After-Scenario Questionnaire (ASQ) (Lewis, 1995), which applies after the user has carried out a predefined task with one system as part of the user tests.

During roughly the same period, the Questionnaire for User Satisfaction (QUIS) (Chin et al., 1988) was designed to find out the satisfaction of users concerning several aspects that make up the man-machine interface. Five main dimensions were selected by the authors: the general consensus of users, screen quality, terminology and information system, how easy it is to learn and system performance. On a structure that is similar on several dimensions, the Software Usability Measurement Inventory (SUMI) (Kirakowski & Corbett, 1993; Van Veenendaal, 1998) is built around five subscales: the perceived efficiency of the system, measured impact, perceived usefulness, the feeling of controlling the system and how easy it is to learn. More recently, the Design-Oriented Evaluation of Perceived Usability (DEEP) (Yang et al., 2012) was developed, combining certain items from the following questionnaires: Purdue Usability Testing Questionnaire (PUTQ) (Lin et al., 1997), Web Analysis and Measurement Inventory (WAMMI) (Kirakowski & Cierlik, 1998) and Usefulness, Satisfaction, and Ease of use (USE) (Lund, 2001). Other items are inspired by the heuristic methods put forward by Perlman (1997) (Practical Heuristics for Usability Evaluation – PHUE) and those of Bolchini and Garzotto (2008) (MiLE+). Specifically dedicated to website

assessment, DEEP consists of 19 items spread out over the following dimensions: perceived content, perceived structure and information architecture, perceived navigation, perceived cognitive effort, coherence between the perceived layout and perceived visual guidance.

Certain questionnaires differentiate themselves from those mentioned above by the fact that certain items are oriented toward collecting users’ emotions, something which constitutes an important component of the UX. The modular evaluation of key Components of User Experience (meCUE) (Minge et al., 2016) relies directly on the CUE (Components of User Experience) model and is based on four main dimensions, designated modules by the authors. These dimensions encompass the perception of the product by its instrumental qualities (usefulness and usability) and non-instrumental qualities (visual esthetics, status and commitment), positive and negative emotions experienced by the user, the consequences of the interaction with the product in terms of loyalty and intention of use and the global assessment of the product. The User Experience Questionnaire (UEQ) (Laugwitz et al., 2008; Schrepp et al., 2014) is also based on the differentiation of the pragmatic and hedonic qualities of a product. In this way, in this questionnaire, the attractiveness of a product is based on measuring the insight, efficiency and loyalty on the pragmatic side and on measuring the stimulation and novelty on the hedonic side.

In conclusion, the AttrakDiff questionnaire (Hassenzahl et al., 2003), presented in the form of semantic differentiators, also distinguishes the perceived pragmatic qualities of a product, which represents a full dimension, the hedonic-stimulation qualities and the hedonic-identity qualities, two other dimensions centered on the user emotions. A fourth dimension, measuring the global attractiveness of the product, is based on the perception of the pragmatic and hedonic qualities of the product.

Today, very few questionnaires have been the subject of a scientifically validated French translation. We can cite for example, the AttrakDiff translated by Lallemand et al. (2015). However, the scientific community and UX professionals often use questionnaires to measure the perceived quality of a product (Hornbaek, 2006). It therefore seemed necessary for us to continue working on a translation of questionnaires measuring UX in order to focus on one of the very first questionnaires that is still very often used today: the System Usability Scale (SUS).

2. The System Usability Scale (SUS)

2.1. Presentation of the SUS

Created in 1986 by Brooke as part of an engineering programme on the usability of the systems, the SUS questionnaire was distributed quickly to assess the usability of the interactive systems (Brooke, 1996, 2013). The SUS was first designed as a “quick and dirty” questionnaire, in order to transfer user tests to the laboratory via a subjective perceived usability measurement, while guaranteeing a fast and non-binding handover for the users questioned.

The SUS comprises ten items, formulated as affirmative statements, for each of which the user states his/her position, by expressing his/her agreement or disagreement on a 5-point Likert scale (1 = do not agree at all; 5 = completely agree). If the user does not know what to answer for an item, he/she is requested to answer nevertheless by ticking the middle of the scale (score 3).

The 10 items of the definitive version of the SUS were selected from a preliminary list of 50 items, designed to cover the three main usability concepts according to the ISO norm 9241-11: effectiveness, efficiency and satisfaction (ISO 9241-11:2018, 2018). The 50 items were first submitted to a panel of 20 users to evaluate two interactive systems, one considered to be really easy to use (a linguistic tool), the other almost impossible to use, even for highly technically skilled users (a tool for system programmers). Then, only the ten items with the most extreme answers were retained, i.e. the items that best discriminate between the two systems. (Brooke, 1996).

2.2. SUS score calculating method

The global SUS score is calculated to take into account inverted items (even pairs) and to obtain a total score of between 1 and 100. For this, it is calculated in three steps:

- (1) First of all, one point is subtracted from the score ticked by the user for items 1, 3, 5, 7 and 9 (odd numbers, not inverted).
- (2) Then, for items 2, 4, 6, 8 and 10 (even numbers, inverted), the calculation is 5 minus the score ticked by the user.
- (3) The 10 new scores recalculated in this way are added together and multiplied by 2.5.

Brooke (2013) explains that this calculation method was defined in order to respond more to marketing requirements than scientific requirements. When the SUS was created, Brooke and his team considered that the project managers, product managers and engineers were more likely to understand a scale that went from 0 to 100 than a scale from 10 to 50 (with 50 being the maximum mark that could have been obtained using the usual calculation method for scales with inverted items). Brooke also indicated that obtaining a mark out of 100 makes it easier to understand the score and the comparison between different systems, since the differences between several scores are perceived as bigger than if the marks were out of 50.

Bangor et al. (2008) sought to give meaning to the SUS scores calculated out of 100. The authors relate that the meaning of the score was always a problem when they had to report the result of a study to a project manager or design team. Therefore, a pilot programme was launched to determine whether adjectives could be linked to SUS scores in order to come up with a more absolute mark. An evaluation scale with 7 adjectives was then used to complement the SUS. 212 participants were invited to complete the SUS, then to answer the question: "Overall, I would rate the user-friendliness of this product as".

The participants had to choose one of the following 7 adjectives: Worst imaginable, Horrible, Poor, OK, Good, Excellent, Best imaginable. A little later, the authors replicated the same study with a larger sample of participants (959 usable results) (Bangor et al. 2009). All adjectives obtained significantly different scores, except "Worst imaginable" and "Horrible". As a result, Bangor et al. (2009) selected 6 adjectives: Worst imaginable, Horrible, OK, Good, Excellent, Best imaginable.

2.3. The SUS, a unidimensional scale?

Some authors have been interested in the statistical validity of the SUS, and in particular in its factor structure. In a first large-scale study on the statistical properties of SUS, including 2324 surveys, Bangor et al. (2008) confirmed the unidimensional nature of SUS, i.e. that the ten items converge toward a single overall score, in line with the way SUS was designed by Brooke (1996). In 2009, Lewis and Sauro challenged the analyses of Bangor et al. (2008) and identified two factors, one comprising eight items relating to usability and the other comprising two items relating to learning. The authors then invited practitioners to use these two dimensions of SUS (usability and learnability) to refine their results, in addition to the overall score.

Borsci et al. (2009) sought to clarify these two divergent results by testing 3 alternative models of SUS: a one-factor model according to the results of Bangor et al. (2008), a two uncorrelated factors model according to the results of Lewis and Sauro (2009), and a two correlated factors including usability and learnability with the same items (1, 2, 3, 5, 6, 7, 8, 9 for the usability; 4 and 10 for the learnability). Statistical analyses revealed that the third model, with the two correlated factors, was the most satisfactory. The authors therefore recommended that the Lewis and Sauro (2009) guidelines of differentiating between usability (factor 1) and learning (factor 2) scores be followed, while indicating that certain conditions (such as the population studied, or the system being assessed) could lead to a high correlation between the two factors.

Later, Lewis and Sauro (2017) revisited the factor structure of SUS by compiling a database of 9156 completed questionnaires. A confirmatory factor analysis (CFA) approach was applied to 3 new models, the first testing the unidimensional SUS, the second testing the bidimensional SUS including usability and learnability factors, the third testing the bidimensional SUS including the positive-tone (odd-numbered) and negative-tone (even-numbered). The new analyses by Lewis and Sauro showed that the two-factor factor structure based on positive and negative tones was the most satisfactory, even though it was not interesting for use by usability practitioners as it did not allow the distinction of particular dimensions in relation to the system being evaluated.

Thus, to this day and until further contradictory research on the factor structure of SUS, SUS should be considered and analyzed as a unidimensional scale.

2.4. Scope of use

Designed to be applied to all types of system as for example, websites (Hussain et al., 2015), mobile applications (Adinda & Suzianti, 2018; Beul-Leusmann et al., 2014), expert systems (South et al., 2017; Yang et al., 2015), serious games (Tolentino et al., 2011) or e-learning systems (Peruri et al., 2017), the SUS has been used to assess a large number of systems, from the creation of the questionnaire up to today. Without being able to be exhaustive, we propose in Appendix A an inventory of SUS uses, only considering studies published after 2010. This classification allows us one hand to take into account a very large scope of application for SUS, and on the other to highlight the capability of SUS to be adapted to modern technologies despite the criticisms that Brooks may have received (Brooke, 2013).

The SUS also acts as a benchmark against which to validate the reliability of new questionnaires measuring usability. For example, Santos et al. (2015) compared the results obtained with the SUS and the HARUS (Handheld Augmented Reality Usability Scale) questionnaire, a scale for measuring the usability of portable augmented reality systems. Karlin and Ford (2013), for their part, created a perceived usability measurement scale dedicated to an ecological feedback scale, the UPscale, inspired in particular by the SUS items.

2.5. SUS translations

Following numerous assessments that the SUS was used for, several translations were done. Thus, in a scientific approach, the SUS has been translated into Indonesian (Sharfina & Santoso, 2016), Portuguese (Martins et al., 2015), Polish (Borkowska & Jach, 2017), Arabic (AlGhannam et al., 2018), Slovenian (Blažica & Lewis, 2015), Greek (Katsanos et al., 2012) and Persian (Dianat et al., 2014). There is also a German version (Rummel, 2015), although this has not been published. While the SUS is used to assess systems in France (see, for example, Larue, 2009), there has never been an official translation and scientific validation.

Finally, we note that the SUS has been transcribed into American Sign Language (Berke et al., 2017), as well as pictorial language (Baumgartner et al., 2019).

3. Methodology for the French translation of the SUS

3.1. Translation of the SUS

To undertake the translation of the French-System Usability Scale (F-SUS), we drew on the method proposed by Vallerand (1989). This author proposes a transcultural validation methodology for psychological questionnaires. This methodology is made up of seven steps, enabling the translation and validation of English-language questionnaires into French. For our translation needs, we selected three distinct steps, described below.

3.1.1. Step 1: preparation of an experimental version

The first step consisted of preparing an experimental version of the original questionnaire in the target language (in this case, French). Several methods are possible, including traditional translation where the researcher does the translation on his/her own, the committee method and the back translation method. For our study, and in order to minimize the bias of traditional translation as cited in Vallerand (1989), we have opted for a committee translation. In this context, three bilingual researchers, all French citizens, were requested to propose a translation of the 10 items from the SUS. First, each translator did his/her own individual translation. Secondly, the researchers presented their translations and had a group discussion on the content of the translations.

At the end of the committee session, two versions of the SUS were selected: a first “word-for-word” translation that kept the original formulations from the beginning of the sentences in the SUS (“I think that”, “I found that ...”, “I thought that ...”) and a second, more literary version, which is closer to a natural formulation in French. Nevertheless, the two versions agree on the sense of the translations; the translators’ discussions focused on the wording of the sentences. The two versions are presented in Table 1. Note that items 9 and 10 have the same translation for both the word-for-word and literary version.

Following the committee translation, a back translation was done on the two translations by five other French-English bilingual researchers who were French native speakers. The back translation consists of doing a “blind” translation (that is,

Table 1. The two translations of the SUS, word-for-word and literary. Items 9 and 10 were translated in the same way. N.B. The literary translation will be the official version used for the F-SUS.

Original items	Word-for-word French translation	Literary French translation
1. I think that I would like to use this system frequently.	Je pense que je voudrais utiliser ce système fréquemment.	Je voudrais utiliser ce système fréquemment.
2. I found this system unnecessarily complex.	Je trouve que ce système est inutilement complexe.	Ce système est inutilement complexe.
3. I thought the system was easy to use.	J'ai trouvé que ce système est facile à utiliser.	Ce système est facile à utiliser.
4. I think that I would need the support of a technical person to be able to use this system.	Je pense que j'aurais besoin du soutien d'un technicien pour être capable d'utiliser ce système.	J'aurais besoin du soutien d'un technicien pour être capable d'utiliser ce système.
5. I found the various functions in this system were well integrated.	J'ai trouvé que les différentes fonctionnalités de ce système sont bien intégrées.	Les différentes fonctionnalités de ce système sont bien intégrées.
6. I thought there was too much inconsistency in this system.	J'ai pensé qu'il y a trop d'incohérences dans ce système.	Il y a trop d'incohérences dans ce système.
7. I would imagine that most people would learn to use this system very quickly.	J'imagine que la plupart des gens apprendront à utiliser ce système très rapidement.	La plupart des gens apprendront à utiliser ce système très rapidement.
8. I found this system very cumbersome to use.	J'ai trouvé ce système très lourd à utiliser.	Ce système est très lourd à utiliser.
9. I felt very confident using the system.	Je me suis senti-e très en confiance en utilisant ce système.	Je me suis senti-e très en confiance en utilisant ce système.
10. I needed to learn a lot of things before I could get going with this system.	J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce système.	J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce système.

without seeing the original version), from the target language translation back to the source language. For our study, the aim was to verify that the translations into French allowed us to rediscover the sense of the original items in English. All the back translations allowed us to validate the committee translations, in particular, for the word-for-word translation. For the literary translation, the back translation allowed us to check that the sense of the items was kept, but the formulations “I think that”, “I find that” were obviously not found.

3.1.2. Step 2: pretest of the experimental versions

The second step consisted of pretesting the two experimental versions of the F-SUS (word-for-word and literary), in order to determine whether the items that made up these two versions were clear, written without ambiguity and in language that is relatable to the target population (Vallerand, 1989). To do this, we submitted each of the experimental versions to a different panel of UX practitioners. We asked them to indicate the level of understanding of each of the items on a scale of 1 (not at all understandable) to 5 (completely understandable). 15 practitioners contributed to the assessment of the word-for-word version and 17 practitioners to the literary version. The means and standard deviation for each item of the SUS are presented in Table 2.

The results of the pretest show that, overall, neither of the two translations stood out as being different to the other. Furthermore, no item seems to pose any comprehension problems, with the least well-rated item having a score of 3.24 (item 5 for the literary translation), which we considered acceptable since this score is above 3 on a scale of 5.

In agreement with these results, we chose to preserve both versions for the statistical evaluations and to keep the translations as they were written during the experimental version phase.

3.2. Step 3: psychometric evaluation of the F-SUS

Even with a meticulous translation of the items, it cannot be guaranteed that the F-SUS has the same psychometric characteristics as the original English version (Van De Vijver & Leung, 2001). The following step consists therefore of carrying out a psychometric assessment of the F-SUS. Several statistical analyses were undertaken in order to measure the accuracy, factor structure, sensitivity and validity of the content.

Table 2. Scores on the level of understanding of the SUS items for the word-for-word and literary translations on a scale of 1 (not at all understandable) to 5 (completely understandable).

SUS item	Word-for-word translation (n = 15)		Literary translation (n = 17)	
	Mean	Standard deviation	Mean	Standard deviation
1	4.40	0.88	4.29	0.96
2	4.53	0.62	4.12	1.13
3	4.60	0.71	4.94	0.24
4	4.27	1.00	4.82	0.51
5	3.93	0.93	3.24	1.39
6	3.93	0.93	4.35	0.76
7	4.60	0.61	4.71	0.75
8	3.80	0.91	4.00	1.14
9	4.40	0.80	4.35	0.76
10	4.53	0.88	4.76	0.42

The minimum recommended sample size for psychometric tests, and more specifically for factor analysis, is to have at least five participants per item, which for the SUS with 10 items, corresponds to a minimum of 50 participants (Nunnally, 1978b).

For the psychometric assessment, we gave the F-SUS to 215 users. Either the word-for-word or the literary translations were distributed randomly. The volunteer users were students in human and social sciences, who were asked to think of a system that they knew well or to display a system to assess on a screen, for example, on their computer or smartphone. We deliberately did not take an interest in the system assessed. Indeed, we considered that the system was of little importance, but it was essential that it was always the same system that was assessed for all the SUS items. The users were made aware of this particular point.

In the end, after having selected the questionnaires that were fully and correctly completed, we retained 79 users for the word-for-word translations (22 women, 57 men; average age = 21.59 years old, standard deviation = 2.79; minimum age = 18; maximum age = 40) and 88 users for the literary translation (78 women, 10 men, average age = 19.66 years old, standard deviation = 3.85; minimum age = 17; maximum age = 42).

4. Results

4.1. Measure of reliability

The Cronbach alpha coefficients of our two translations are positive and confirm sufficient reliability, with a threshold above .70, as recommended by Landauer (1997) or Kline (2005). Therefore, for the word-for-word translation of the SUS, we obtain an alpha of .904; and an alpha of .899 for the literary translation. These two reliability scores are among the highest measured during the different translations of the SUS (Lewis, 2018). They are close to those observed for example, by Lewis and Sauro (2009) (alpha of .91) or Bangor et al. (2008) (alpha of .911).

No item deletion, in either version, brings any significant advantage.

4.2. Factor analyses

A principal component factor analysis (PCA) was carried out to test the construct validity of the two translations of the SUS. The interest of a PCA is to check whether the factor structure of our translations is similar to the structure of the original questionnaire (Brooke, 1996).

Cattell's scree plot (Cattell, 1966) (Eigenvalues) suggests a 2-factor structure for the two translations (Figure 1). If this result does not converge toward a structure with a single factor, as in the SUS designed by Brooke (1996) and confirmed by Bangor et al. (2008), it joins that obtained by Lewis and Sauro (2009), then verified by Borsci et al. (2009).

Table 3 presents the 2-factor Varimax rotation for the 10 SUS items and for each of the translations.

The 2-factor PCA shows that for the two translations, items 1, 2, 3, 5, 6, 7, 8 and 9 are aligned with the first factor and items 4 and 10 are aligned with the second factor. These

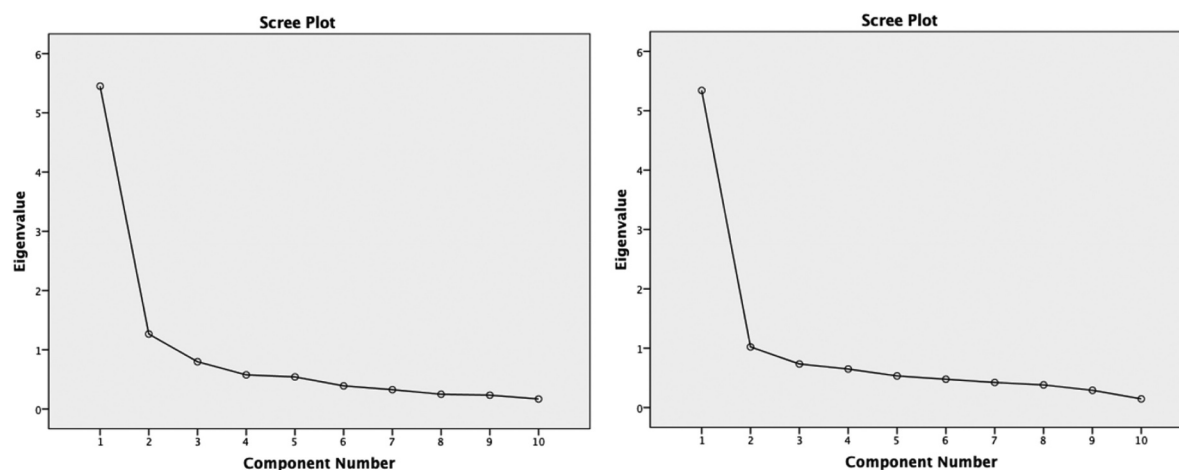


Figure 1. Scree plot for word-for-word translation (left) and literary translation (right).

Table 3. 2-Factor Varimax rotation for the 10 SUS items for word-for-word translation and literary translation.

Item	Word-for-word translation		Literary translation	
	Factor 1	Factor 2	Factor 1	Factor 2
1	.849	.096	.728	.098
2	.720	.114	.757	.280
3	.851	.261	.675	.275
4	.208	.812	.344	.682
5	.809	.199	.678	.400
6	.780	.188	.852	.143
7	.602	.538	.710	.303
8	.855	.210	.847	.250
9	.569	.247	.735	.049
10	.112	.893	.054	.878

results are in accordance with the factor analyses led by Lewis and Sauro (2009). The authors entitled the scale identified by the first factor “Usability” and the second scale identified by the second factor (with items 4 and 10) “Learning”. Nevertheless, it should be remembered that the authors later went back on these conclusions, and showed that the SUS has a two factors structure, but this structure was not really interesting to take into account because it was in fact based on the SUS’s positive-tone (odd-numbered) and negative-tone (even-numbered).

We also note that the factor structure of the literary translation seems to be the most coherent with regard to its better distinction in terms of saturation between the two factors on all items, and particularly for items 7 and 9 where the saturation difference is well over +0.3.

4.3. Sensitivity measurement

4.3.1. Gender effect

In order to check whether the translations are gender-sensitive, we undertook T-tests for independent samples. Whether for the word-for-word translation or for the literary translation, we observed that the variances were not significantly different for all items. We can therefore conclude that no gender difference or effect is observed for both translations.

These results are consistent with most of the results obtained in the studies that investigated the gender question on SUS ratings. In a general overview of the SUS, Lewis (2018) identified six studies that have focused on the influence of gender. Of these, 5 found no significant effect (Bangor et al., 2008; Berkman & Karahoca, 2016; Kortum & Sorber, 2015; Kortum & Bangor, 2013; Tossell et al., 2012). Only Kortum and Oswald (2018) study found a significant gender-related difference in responses to SUS, but this difference is rather due to a personality difference between the participants.

4.3.2. Trends toward extreme answers

In order to check possible trends toward extreme answers related to our translations or to possible cultural variables, the 5% trimmed means were investigated: no significant difference was observed.

Shown graphically, the different scatter plots related to the items do not imply any trends toward extreme answers.

5. Discussion

5.1. Choice of translation of the SUS

Consequently, it was a question of choosing between the two translations that we produced, word-for-word or literary, to select only one. In view of the statistical results, the literary version seems to us to obtain the best results, in particular, concerning factor analyses. Indeed, the Varimax rotation obtains more contrasted values between the two factors than for the word-for-word version, in particular, for item 7 (“J’imagine que la plupart des gens apprendront à utiliser ce système très rapidement”). This item seems to cause comprehension problems for the word-for-word version, which uses the beginning of the sentence of the original version of the SUS: “I would imagine that”. We therefore think that the formulation “j’imagine que” introduces ambiguity in the understanding of the item. Indeed, contrary to the other items of the SUS that encourage the user to express a personal opinion by means of introductory formulations like “je pense que” (“I think that”) or “je

trouve que” (“I find that”), the formulation “j’imagine que” requires the user to guess something, to make an assumption without being able to check it. For the user therefore, it is a question of giving a speculative opinion, which can generate SUS answers with little contrast.

We also observe that the translation committee did not select the conditional form of the formulation “j’imaginer-ais” (“I would imagine”), (the committee translated “j’imagine que la plupart des gens apprendront à utiliser ce système très rapidement”), which would usually apply with the auxiliary verb “would”, i.e. “I would imagine that most people would learn to use this system very quickly”. The translation committee justified this by the rare and redundant use of the conditional, in French, with verbs such as “imaginer” (to imagine), which already expresses a possibility, hypothesis or projection in the future.

Furthermore, item understanding scores (Table 2) very slightly favor the literary version ($M = .299$ for the word-for-word version, $M = .358$ for the literary version), even if the difference could be considered insignificant. The standard deviation is also weaker for the literary version ($ES = .827$ for the word-for-word version; $ES = .806$ for the literary version).

Lastly, and less statistically, the French language generally favors direct turns of phrase, with the least possible ambiguity, for formulating questionnaire items. Therefore, the literary version displays an editorial structure better adapted to francophone culture. The complete version of the F-SUS selected by our study is presented in Annex A.

5.2. A two-factor scale?

The question of the single factor of the SUS has already been the subject of several research studies (Borsci et al., 2009; Lewis & Sauro, 2009). Our factor analyses revealed a two-factor scale for the SUS, regardless of whether the translation is literary or word-for-word. This is for items 4 (J’aurais besoin du soutien d’un technicien pour être capable d’utiliser ce système) and 10 (J’ai eu besoin d’apprendre beaucoup de choses avant de pouvoir utiliser ce système), which differ from the other items. As we have highlighted, these results correspond to those found by Lewis and Sauro (2009), which named “Learning” as the dimension formed by items 4 and 10 and “Usability” as the second dimension. In 2009, Lewis and Sauro were surprised that the “Learning” dimension did not include item 7 (“La plupart des gens apprendront à utiliser ce système très rapidement”), which also clearly addresses the concept of learning in its formulation. They justified the exclusion of item 7 by the fact that this item highlights the awareness of others rather than the skills of the user him/herself.

Later, Lewis and Sauro (2017) carried out new factor analyses and called into question the two-dimension structure (Usability – Learning) of the SUS. According to the authors, their new results suggest (but do not prove) that there could be certain research contexts in which items 4 and 10 appear as an independent factor. Thus, still according to Lewis and Sauro (2017), the SUS is above all unidimensional, as other authors have verified (Bangor et al., 2008; Borsci et al., 2009),

and its two-dimensional structure is only validated for its mixed tonality structure, that is, for its alternance between positive and negative items (remember that one in two items is inverted).

Our research may thus rekindles the debate concerning a two-dimensional scale, since our results are in the line with those of Lewis and Sauro in 2009. We find interesting to note that we see the same factor analyses for both translations that we produced (word-for-word and literary), something that cannot be a coincidence. Remember that each of these translations was submitted to an independent group of users, that is, it was not the same users that completed the word-for-word translation and the literary translation. Therefore, there cannot be a “subject” effect. On the other hand, since the users had the same instructions, there may be a “context” effect, as indicated by Lewis and Sauro (2017).

Nevertheless, we should point out that these results were obtained from a small sample size ($N = 167$). However, when the samples are large, the two-factor structure of the SUS tends to disappear. This is what happened with the Lewis and Sauro (2017) and Bangor et al. (2008) complementary studies. Therefore, we are considering a complementary study on a larger sample, before making a more categorical statement on the factor structure of the French SUS.

6. Conclusion and research perspectives

This research aims to propose a French translation for the System Usability Scale, one of the questionnaires measuring usability of the most popular systems in the area of user experience. By relying on Vallerand’s transcultural validation methodology (Vallerand, 1989), we first selected two French translations of the SUS, one that resumed all the items word-for-word and another that came closer to the customary linguistic formulations in French with a more literary translation. The psychometric assessments for each of these two scales, and in particular the factor analyses, encouraged us to select the literary version of the F-SUS.

Complementary studies, some of which are already underway, could be carried out in order to continue the French-language distribution of the SUS.

6.1. Studies on other systems and other populations

Our statistical analyses were based on the questionnaires completed by students in human and social sciences, by asking them to think of a system that they knew well or to display it on a screen. By doing this, we were not affected by the characteristics of the system to be assessed. We considered that the coherence of the individual answers prevailed over the coherence of the collective answers. This methodological approach has proven to be valid since we noticed a very strong internal consistency for each of the translations.

To go further, we currently envisage questioning different populations (students, active persons and UX professionals) in order to assess predefined systems. To date, we have gathered numerous assessments from a university digital workspace by students, from the website of a large European bank by the workforce, and from a bibliographic

document management system by UX professionals. We have chosen these different systems so that the users find out about them, and are then able to give their opinion by using the F-SUS. We think that these complementary studies will enable us to continue with the validation of the literary version of the F-SUS, and we hope to validate our first factor analyses, especially concerning the two-dimensionality of the SUS.

6.2. Comparison with other scales

We also consider it important to compare the results of the F-SUS with other francophone UX or usability measurement scales. Unfortunately, to date there are few translated scales that have been scientifically validated. We can mention the AttrakDiff scale (Hassenzahl et al., 2003), which was translated in 2015 by Lallemand et al. With AttrakDiff presenting four distinct dimensions relating to the user experience (hedonic quality identity, hedonic quality stimulation, pragmatic quality and global attractiveness), of which only one can be clearly related to the usability measurement (pragmatic quality), we are considering comparing the results of the F-SUS with those obtained by the pragmatic dimension of AttrakDiff.

Sauro (2014) also proposes a translation of the SUPR-Q (Sauro, 2015) that could be exploited.

6.3. Validation of the French version of the SUS in a “positive” version

The research of Lewis and Sauro (Sauro & Lewis, 2011; Lewis & Sauro, 2017) showed that a positive version of the SUS, that is, not including inverted items, could avoid careless mistakes and misunderstandings among users. Furthermore, a positive version could stop researchers making coding mistakes. In addition, Kortum et al. (2020) demonstrated that the positive version of SUS had the same statistical properties as the standard SUS. These two versions could therefore be used interchangeably, although it was always preferable to compare scores obtained using the same version of SUS.

It therefore appears interesting to continue these studies by focusing on validating a positive version of the F-SUS and comparing the results we obtain with those of Lewis and Sauro.

6.4. Validation of the French version of the UMUX and UMUX-LITE

Finally, as a streamlined version of the SUS, the UMUX (Finstad, 2010) and UMUX-LITE scales (Lewis et al., 2013) will also be able to be translated and validated in a French version. These two questionnaires appear interesting to us since they allow us to keep a global score out of 100, just like the SUS, from a reduced number of items (4 for the UMUX, 2 for the UMUX-LITE) and without any inverted items for the UMUX-LITE. These questionnaires thus correspond to the new forms of remote handovers, which require streamlined protocols with the fewest possible sources of completion errors for users.

Acknowledgments

The authors would like to thank Laurence Johannsen, Pierre-Jean Barlatier and Anne-Laure Mention who were involved in the translation process of the F-SUS.

References

- Adinda, P. P., & Suzianti, A. (2018). Redesign of user interface for E-government application using usability testing method. In *ICCIP* (pp. 145–149). Qingdao, China. <https://doi.org/10.1145/3290420.3290433>
- Aguilar-Reyes, F., Poblete, K. L., & González, V. M. (2015). UX-AdChat: Semi-automatic, predictable usability evaluator of written dialogue systems. In *CLIHIC*. Córdoba, Argentina.
- AlGhannam, B. A., Albustan, S. A., Al-Hassan, A. A., & Albustan, L. A. (2018). Towards a standard arabic system usability scale: psychometric evaluation using communication disorder app. *International Journal of Human-Computer Interaction*, 34(9), 799–804. <https://doi.org/10.1080/10447318.2017.1388099>
- Ašeriškis, D., & Damaševičius, R. (2014). Gamification of a project management system. In *ACHI 2014-7th International Conference on Advances in Computer-Human Interactions* (pp. 200–207). Barcelona, Spain.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114–123. <https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Baumgartner, J., Frei, N., Kleinke, M., Sauer, J., & Sonderegger, A. (2019). Pictorial system usability scale (P-SUS): Developing an instrument for measuring perceived usability. In *CHI 2019, May 4–9, 2019, Glasgow, Scotland, UK* (pp. 1–11). <https://doi.org/10.1145/3290605.3300299>
- Berke, L., Huenerfauth, M., & Patel, K. (2017). Design and psychometric evaluation of American sign language translations of usability questionnaires. In *ASSETS'17, Oct. 29–Nov. 1, 2017, Baltimore, MD, USA* (pp. 175–184). <https://doi.org/10.1145/3314205>
- Berkman, M. I., & Karahoca, D. (2016). Re-Assessing the usability metric for user experience (UMUX) scale. *Journal of Usability Studies*, 11(3), 89–109. <https://uxpajournal.org/assessing-usability-metric-umux-scale/>
- Beul-Leusmann, S., Samsel, C., Wiederhold, M., Krempels, K. H., Jakobs, E. M., & Ziefle, M. (2014). Usability evaluation of mobile passenger information systems. In Marcus A. (Eds.), *Design, user experience, and usability. Theories, methods, and tools for designing the user experience. DUXU 2014. Lecture notes in computer science* (Vol. 8517). Springer, Cham. https://doi.org/10.1007/978-3-319-07668-3_22
- Blažica, B., & Lewis, J. R. (2015). A Slovene translation of the system usability scale: The SUS-SI. *International Journal of Human-Computer Interaction*, 31(2), 112–117. <https://doi.org/10.1080/10447318.2014.986634>
- Blecken, A., Brüggemann, D., & Marx, W. (2010). Usability evaluation of a learning management system. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 1–9). Honolulu, HI, USA. <https://doi.org/10.1109/HICSS.2010.422>
- Bolchini, D., & Garzotto, F. (2008). Quality and potential for adoption of usability evaluation methods: An empirical study on MiLE+. *Journal of Web Engineering*, 7(4), 299–317. <http://dl.acm.org/citation.cfm?id=2011277.2011281>
- Borkowska, A., & Jach, K. (2017). Pre-testing of polish translation of system usability scale (SUS). In Borzemski L., Grzech A., Świątek J., Wilimowska Z. (Eds.), *Information systems architecture and technology: Proceedings of 37th international conference on information systems architecture and technology – ISAT 2016 – Part I. Advances in*

- intelligent systems and computing* (Vol. 521). Springer, Cham. https://doi.org/10.1007/978-3-319-46583-8_12
- Borsci, S., Federici, S., & Lauriola, M. (2009). On the dimensionality of the system usability scale: A test of alternative measurement models. *Cognitive Processing*, 10(3), 193–197. <https://doi.org/10.1007/s10339-009-0268-9>
- Botella, F., Moreno, J. P., & Peñalver, A. (2014). How efficient can be a user with a tablet versus a smartphone? In *Interaccion'14* (Vol. 10-12-Sept). Puerto de la Cruz, Tenerife, Spain. <https://doi.org/10.1145/2662253.2662317>
- Braunhofer, M., Elahi, M., & Ricci, F. (2014). Usability assessment of a context-aware and personality-based mobile recommender system. In Hepp M., Hoffner Y. (Eds.), *E-Commerce and web technologies. EC-web 2014. Lecture notes in business information processing* (Vol. 188). Springer, Cham. https://doi.org/10.1007/978-3-319-10491-1_9
- Brooke. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. McClelland Eds., *Usability evaluation in industry*, 189–194 Taylor & Francis. <http://hell.meiert.org/core/pdf/sus.pdf>
- Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, 8(2), 29–40. http://www.usabilityprofessionals.org/upa_publications/jus/2013february/brooke1.html%5Cnhttp://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10
- Chin, J. P., Diehl, V. A., & Norman, L. K. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '88*, 213–218. Washington, D.C., USA. <https://doi.org/10.1145/57167.57203>
- Çöltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, 36(1), 5–17. <https://doi.org/10.1559/152304009787340197>
- Dhillon, J. S., Wünsche, B. C., & Lutteroth, C. (2012). Evaluation of a web-based telehealth system: A preliminary investigation with seniors in New Zealand. In *CHINZ'12* (pp.17–24). Dunedin, New Zealand. <https://doi.org/10.1145/2379256.2379260>
- Dianat, I., Ghanbari, Z., & Asgharijafarabadi, M. (2014). Psychometric properties of the persian language version of the system usability scale. *Health Promotion Perspectives*, 4(1), 82–829. <https://doi.org/10.5681/hpp.2014.011>
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105–123. <https://www.erpjournal.net/wp-content/uploads/2020/02/ERP38-1-Drost-E.-2011.-Validity-and-Reliability-in-Social-Science-Research.pdf>
- Evans, J., Papadopoulos, A., Silvers, C. T., Charness, N., Boot, W. R., Schlachta-Fairchild, L., ... Ent, C. B. (2016). Remote health monitoring for older adults and those with heart failure: Adherence and system usability. *Telemedicine and E-Health*, 22(6), 480–488. <https://doi.org/10.1089/tmj.2015.0140>
- Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5), 323–327. <https://doi.org/10.1016/j.intcom.2010.04.004>
- Fritz, F., Balhorn, S., Riek, M., Breil, B., & Dugas, M. (2012). Qualitative and quantitative evaluation of EHR-integrated mobile patient questionnaires regarding usability and cost-efficiency. *International Journal of Medical Informatics*, 81(5), 303–313. <https://doi.org/10.1016/j.ijmedinf.2011.12.008>
- García-Peñalvo, F. J., García-Holgado, A., Vázquez-Ingelmo, A., & Seoane-Pardo, A. M. (2018). Usability test of WYRED platform. In P. Zaphiris & A. Ioannou, (Eds.), *Learning and collaboration technologies. Design, development and technological innovation. LCT 2018. Lecture notes in computer science* (Vol. 10924). Springer, Cham. https://doi.org/10.1007/978-3-319-91743-6_5
- Georgsson, M., & Staggers, N. (2016). Quantifying usability: An evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. *Journal of the American Medical Informatics Association*, 23(1), 5–11. <https://doi.org/10.1093/jamia/ocv099>
- Ghosh, D., Foong, P. S., Zhang, S., & Zhao, S. (2018). Assessing the utility of the system usability scale for evaluating voice-based user interfaces. In *ChineseCHI'18* (pp. 11–15). Montreal, QC, Canada. <https://doi.org/10.1145/3202667.3204844>
- Hafiz, P., Maharjan, R., & Kumar, D. (2018). Usability of a mood assessment smartphone prototype based on humor appreciation. In *MobileHCI 2018 - Beyond Mobile: The Next 20 Years - 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Conference Proceedings Adjunct* (pp. 151–157). Barcelona, Spain. <https://doi.org/10.1145/3236112.3236134>
- Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2016). Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Computers in Human Behavior*, 61(August), 463–471. <https://doi.org/10.1016/j.chb.2016.03.051>
- Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler, (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung* (pp. 187–196). Stuttgart: B. G. Teubner. https://doi.org/10.1007/978-3-322-80058-9_19
- Heinonen, R., Luoto, R., Lindfors, P., & Nygård, C. H. (2012). Usability and feasibility of mobile phone diaries in an experimental physical exercise study. *Telemedicine and E-Health*, 18(2), 115–119. <https://doi.org/10.1089/tmj.2011.0087>
- Held, J. P., Klaassen, B., Van Beijnum, B.-J. F., Luft, A. R., & Veltink, P. H. (2017). Usability evaluation of a vibrotactile feedback system in stroke subjects. *Frontiers in Bioengineering and Biotechnology*, 4(January), 1–7. <https://doi.org/10.3389/fbioe.2016.00098>
- Hookham, G., Nesbitt, K., & Kay-Lambkin, F. (2016). Comparing usability and engagement between a serious game and a traditional online program. *ACM International Conference Proceeding Series*, 01-05-Febr. Canberra, Australia. <https://doi.org/10.1145/2843043.2843365>
- Hornbaek, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79–102. <https://doi.org/10.1016/j.ijhcs.2005.06.002>
- Hussain, A., Mkpjojiogu, E. O. C., & Hussain, Z. (2015). Usability evaluation of a web-based health awareness portal on smartphone devices using ISO 9241-11 model. *Jurnal Teknologi*, 77(4), 1–5. <https://doi.org/10.11113/jt.v77.6035>
- ISO 13407:1999.(1999). *Human-centred design processes for interactive systems (ISO 13407:1999)*. International organization for standardization. [http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Human-centred+design+processes+for+interactive+systems+\(ISO+13407:1999\)#5](http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Human-centred+design+processes+for+interactive+systems+(ISO+13407:1999)#5)
- ISO 9241-11:2018. (2018). *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*.
- ISO 9241-210:2010. (2010). *Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems*.
- Karlin, B., & Ford, R. (2013). The usability perception scale (UPscale): A measure for evaluating feedback displays. In Marcus A. (Eds.), *Design, user experience, and usability. Design philosophy, methods, and tools. DUXU 2013. Lecture notes in computer science* (Vol. 8012). Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-39229-0>
- Katsanos, C., Tselios, N., & Xenos, M. (2012). Perceived usability evaluation of learning management systems: A first step towards standardization of the system usability scale in Greek. In *Proceedings of the 2012 16th Panhellenic Conference on Informatics, PCI 2012, Piraeus, Greece: IEEE CPS*, 5-7 Oct., 2012 (pp. 302–307). Piraeus, Greece. <https://doi.org/10.1109/PCI.2012.38>
- Kerzazi, N., & Lavallee, M. (2011). Inquiry on usability of two software process modeling systems using ISO/IEC 9241. In *Canadian Conference on Electrical and Computer Engineering* (pp. 000773–000776). Niagara Falls, ON, Canada. <https://doi.org/10.1109/CCECE.2011.6030560>
- Kim, M. S., Shapiro, J. S., Genes, N., Aguilar, M. V., Mohrer, D., Baumlin, K., & Belden, J. L. (2012). A pilot study on usability analysis of emergency department information system by nurses. *Applied Clinical Informatics*, 3(1), 135–153. <https://doi.org/10.4338/ACI-2011-11-RA-0065>

- Kirakowski, J., & Cierlik, B. (1998). Measuring the usability of web sites. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(4), 424–428. <https://doi.org/10.1177/154193129804200405>
- Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. *British Journal of Educational Technology*, 24(3), 210–212. <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.1993.tb00076.x/abstract>
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). The Guilfo.
- Kortum, P., Acemyan, C. Z., & Oswald, F. L. (2020). Is it time to go positive? Assessing the positively worded system usability scale (SUS). *Human Factors*. <https://doi.org/10.1177/0018720819881556>
- Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. *International Journal of Human-Computer Interaction*, 34(2), 177–186. <https://doi.org/10.1080/10447318.2017.1336317>
- Kortum, P., & Sorber, M. (2015). Measuring the Usability of mobile applications for phones and tablets. *International Journal of Human-Computer Interaction*, 31(8), 518–529. <https://doi.org/10.1080/10447318.2015.1064658>
- Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the system usability scale. *International Journal of Human-Computer Interaction*, 29(2), 67–76. <https://doi.org/10.1080/10447318.2012.681221>
- Lacerda, T. C., Von Wangenheim, C. G., Von Wangenheim, A., & Giuliano, I. (2014). Does the use of structured reporting improve usability? A comparative evaluation of the usability of two approaches for findings reporting in a large-scale telecardiology context. *Journal of Biomedical Informatics*, 52, 222–230. <https://doi.org/10.1016/j.jbi.2014.07.002>
- Lallemant, C., Koenig, V., Gronier, G., & Martin, R. (2015). Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. *Revue Européenne De Psychologie Appliquée/European Review of Applied Psychology*, 65(5), 239–252. <https://doi.org/10.1016/j.erap.2015.08.002>
- Landauer, T. K. (1997). Chapter 9 - behavioral research methods in human-computer interaction. In *Handbook of human-computer interaction* (Second ed., pp. 203–227). North-Holland.
- Larue, V. (2009). Apport du System Usability Scale à l'activité d'ergonomie d'évaluation. In *IHM 2009*, 13-16 Octobre 2009, Grenoble, France (pp. 155–161).
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *USAB 2008* (pp. 63–76). https://doi.org/10.1007/978-3-540-89350-9_6
- Lewis, J. (1995). IBM Computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57–78. <http://www.tandfonline.com/doi/abs/10.1080/10447319509526110>
- Lewis, J., Utesch, B., & Maher, D. (2013). UMUX-LITE: When there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Computer Interaction* (pp. 2099–2102). Paris, France. <http://dl.acm.org/citation.cfm?id=2481287>
- Lewis, J. R. (2018). The system usability scale: Past, present, and future. *International Journal of Human-Computer Interaction*, 34(7), 577–590. <https://doi.org/10.1080/10447318.2018.1455307>
- Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In *Human Centered Design, First International Conference, HCD 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24* (pp. 26–31). <https://doi.org/10.1007/978-3-642-02806-9>
- Lewis, J. R., & Sauro, J. (2017). Revisiting the factor structure of the system usability scale. *Journal of Usability Studies*, 12(4), 183–192. <https://uxpajournal.org/visit-factor-structure-system-usability-scale/>
- Liang, J., Xian, D., Liu, X., Fu, J., Zhang, X., Tang, B., & Lei, J. (2018). Usability study of mainstream wearable fitness devices: Feature analysis and system usability scale evaluation. *JMIR MHealth and UHealth*, 6(11), e11066. <https://doi.org/10.2196/11066>
- Lin, H., Choong, -Y.-Y., & Salvendy, G. (1997). A proposed index of usability: A method for comparing the relative usability of different software systems. *Behaviour & Information Technology*, 16(4), 267–277. <https://doi.org/10.1080/014492997119833>
- Lin, H. C. K., Chen, M. C., & Chang, C. K. (2015). Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system. *Interactive Learning Environments*, 23(6), 799–810. <https://doi.org/10.1080/10494820.2013.817435>
- Lin, H. C. K., Wu, C. H., & Hsueh, Y. P. (2014). The influence of using affective tutoring system in accounting remedial instruction on learning performance and usability. *Computers in Human Behavior*, 41 (December), 514–522. <https://doi.org/10.1016/j.chb.2014.09.052>
- Lin, H. K., Hsieh, M., Wang, C., Sie, Z., & Chang, S. (2011). Establishment and usability evaluation of an interactive ar learning system on conservation of fish. *TOJET: The Turkish Online Journal of Educational Technology*, 10(4), 181–187. <http://www.tojet.net/articles/v10i4/10418.pdf>
- Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, 8(2), 3–6. <https://doi.org/10.1177/1078087402250360>
- Lutes, K. D., Chang, K., & Baggili, I. M. (2006). Diabetic e-management system (DEMS). In *Proceedings - Third International Conference on Information Technology: New Generations, ITNG 2006* (Vol.2006, pp. 619–624). Las Vegas, NV, USA. <https://doi.org/10.1109/ITNG.2006.53>
- Maguire, M. (2001). Methods to support human-centred design. *International Journal of Human-Computer Studies*, 55(4), 587–634. <https://doi.org/10.1006/ijhc.2001.0503>
- Martin-Gonzalez, A., Chi-Poot, A., & Uc-Cetina, V. (2016). Usability evaluation of an augmented reality system for teaching euclidean vectors. *Innovations in Education and Teaching International*, 53(6), 627–636. <https://doi.org/10.1080/14703297.2015.1108856>
- Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., & Rocha, N. P. (2015). European Portuguese validation of the system usability scale (SUS). *Procedia Computer Science*, 67(Dsai), 293–300. <https://doi.org/10.1016/j.procs.2015.09.273>
- Meldrum, D., Glennon, A., Herdman, S., Murray, D., & McConn-Walsh, R. (2012). Virtual reality rehabilitation of balance: Assessment of the usability of the Nintendo Wii® Fit Plus. *Disability and Rehabilitation: Assistive Technology*, 7(3), 205–210. <https://doi.org/10.3109/17483107.2011.616922>
- Meulendijk, M. C., Spruit, M. R., Drenth-van Maanen, A. C., Numans, M. E., Brinkkemper, S., Jansen, P. A. F., & Knol, W. (2015). Computerized decision support improves medication review effectiveness: An experiment evaluating the STRIP assistant's usability. *Drugs and Aging*, 32(6), 495–503. <https://doi.org/10.1007/s40266-015-0270-0>
- Minge, M., Thüring, M., & Wagner, I. (2016). Developing and validating an English version of the meCUE questionnaire for measuring user experience. *Proceedings of the Human Factors and Ergonomics Society*, 60(1), 2056–2060. <https://doi.org/10.1177/1541931213601468>
- Munsinger, B., & Quarles, J. (2019). Augmented Reality for Children in a Confirmation Task: Time, Fatigue, and Usability, 1–5. <https://doi.org/10.1145/3359996.3364274>
- Nawaz, A., Skjæret, N., Ystmark, K., Helbostad, J. L., Vereijken, B., & Svanæs, D. (2014). Assessing seniors' user experience (UX) of exergames for balance training. In *Proceedings of the NordiCHI 2014: The 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (pp. 578–587). Helsinki, Finland. <https://doi.org/10.1145/2639189.2639235>
- Ng, A. W. Y., Lo, H. W. C., & Chan, A. H. S. (2011). Measuring the usability of safety signs: A use of system usability scale (SUS). In *IMECS 2011 - International MultiConference of Engineers and Computer Scientists 2011* (Vol.2, pp. 1296–1301). Kowloon, Hong Kong.
- Nunnally, J. C. (1978a). An overview of psychological measurement. In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders* (pp. 97–146). Plenum Pre. <https://doi.org/10.1007/978-3-540-92673-3>
- Nunnally, J. C. (1978b). *Psychometric theory* (2 ed.). McGraw-Hill.
- Orfanou, K., Tselios, N., & Katsanos, C. (2015). Perceived usability evaluation of learning management systems: Empirical evaluation of the system usability scale. *International Review of Research in Open and Distance Learning*, 16(2), 227–246. <https://doi.org/10.19173/irrodl.v16i2.1955>
- Perlman, G. (1997). Practical usability evaluation. In *CHI '97* (pp. 369–370). <https://doi.org/10.1145/223355.223726>

- Peruri, A., Borchert, O., Cox, K., Hokanson, G., & Slator, B. M. (2017). Using the System usability scale in a classification learning environment. In Auer M., Guralnick D., Uhomobhi J. (Eds.), *Interactive collaborative learning. ICL 2016. Advances in intelligent systems and computing* (vol. 544). Springer, Cham. https://doi.org/10.1007/978-3-319-50337-0_14
- Peterson, R. A. (2013). Meta-analysis of alpha cronbach ' s coefficient. *Journal of Consumer Research*, 21(2), 381–391. <https://doi.org/10.1093/bioinformatics/btr476>
- Raptis, D., Tselios, N., Kjeldskov, J., & Skov, M. B. (2013). Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. *MobileHCI 2013 - Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 127–136. Munich, Germany. <https://doi.org/10.1145/2493190.2493204>
- Rivythi, A., & Tselios, N. (2019). Extension of technology acceptance model by using system usability scale to assess behavioral intention to use e-learning. *Education and Information Technologies*, 24(4), 2341–2355. <https://doi.org/10.1007/s10639-019-09869-4>
- Rummel, B. (2015). System Usability Scale – Jetzt auch auf Deutsch. System Usability Scale – jetzt auch auf Deutsch
- Ruoti, S., Andersen, J., Zappala, D., & Seamons, K. (2015). *Why Johnny still, still can't encrypt: Evaluating the usability of a modern PGP client*. <http://arxiv.org/abs/1510.08555>
- Santos, M. E. C., Polvi, J., Taketomi, T., Yamamoto, G., Sandor, C., & Kato, H. (2015). Toward standard usability questionnaires for hand-held augmented reality. *Virtual Reality Software and Technology*, 35(5), 66–75. <https://doi.org/10.1109/MCG.2015.94>
- Sauro, J. (2014). *4 steps to translating a questionnaire*. Retrieved February 9, 2021, from <https://measuringu.com/translation/>
- Sauro, J. (2015). SUPR-Q: A comprehensive measure of the quality of the website user experience. *Journal of Usability Studies*, 10(2), 68–86. <https://uxpajournal.org/supr-q-a-comprehensive-measure-of-the-quality-of-the-website-user-experience/>
- Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? *Conference on Human Factors in Computing Systems - Proceedings*, 2215–2223. Vancouver, BC, Canada. <https://doi.org/10.1145/1978942.1979266>
- Sauro, J., & Lewis, J. R. (2012). *Quantifying the user experience*. Elsevier. <https://doi.org/10.1016/B978-0-12-384968-7.00013-8>
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In A. Marcus (Ed.), *Design, user experience, and usability. Theories, methods, and tools for designing the user experience* (pp. 383–392). Springer International Publishing.
- Sharfina, Z., & Santoso, H. B. (2016). An Indonesian adaptation of the system usability scale (SUS). In *2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016* (pp. 145–148). Malang, Indonesia. <https://doi.org/10.1109/ICACSIS.2016.7872776>
- South, H., Taylor, M., Dogan, H., & Jiang, N. (2017). Digitising a medical clerking system with multimodal interaction support. *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017-Janua, 238–242. Glasgow, UK. <https://doi.org/10.1145/3136755.3136758>
- Tassabehji, R., & Kamala, M. A. (2012). Evaluating biometrics for online banking: The case for usability. *International Journal of Information Management*, 32(5), 489–494. <https://doi.org/10.1016/j.ijinfomgt.2012.07.001>
- Tolentino, G. P., Battaglini, C., Pereira, A. C. V., De Oliveria, R. J., & De Paula, M. G. M. (2011). Usability of serious games for health. In *Proceedings - 2011 3rd International Conference on Games and Virtual Worlds for Serious Applications, VS-GAMES 2011* (pp. 172–175). Athens, Greece. <https://doi.org/10.1109/VS-GAMES.2011.33>
- Tossell, C. C., Kortum, P., Shepard, C., Rahmati, A., & Zhong, L. (2012). An empirical analysis of smartphone personalisation: Measurement and user variability. *Behaviour and Information Technology*, 31(10), 995–1010. <https://doi.org/10.1080/0144929X.2012.687773>
- Triantafyllidis, A. K., Koutkias, V. G., Chouvarda, I., & Maglaveras, N. (2014). Development and usability of a personalized sensor-based system for pervasive healthcare. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014* (pp. 6623–6626). Chicago, IL, USA. <https://doi.org/10.1109/EMBC.2014.6945146>
- Vallerand, R. J. (1989). Vers une méthodologie de validation transculturelle de questionnaires psychologiques : Implications pour la recherche en langue française. *Psychologie Canadienne*, 30(4), 662–680. <https://doi.org/10.1037/h0079856>
- Van De Vijver, F. J. R., & Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality*, 69(6), 1007–1031. <https://doi.org/10.1111/1467-6494.696173>
- Van Veenendaal, E. (1998). Questionnaire based usability testing. In *Conference Proceedings European Software Quality Week, Brussels*, November 1998 (pp. 1–9).
- Yang, D., Zhang, D., Chen, L., & Qu, B. (2015). NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications*, 55(August), 170–180. <https://doi.org/10.1016/j.jnca.2015.05.010>
- Yang, T., Linder, J., & Bolchini, D. (2012). DEEP: Design-oriented evaluation of perceived usability. *International Journal of Human-Computer Interaction*, 28(5), 308–346. <https://doi.org/10.1080/10447318.2011.586320>

About the Authors

Guillaume Gronier is senior researcher at Luxembourg Institute of Science and Technology (LIST). His research interests include cognitive ergonomics, human-centered design, user experience, and adaptation of usability assessment methods. He is the coauthor (with Carine Lallemand) of the French reference book *Méthodes de Design UX*.

Alexandre Baudet is researcher at Luxembourg Institute of Science and Technology (LIST). His research interests include competency and performance management and especially computer-based assessment.

Appendix A: Referencing of different SUS applications.

Category	Application	Reference	
Mobile application	Government Information Application	(Adinda & Suzianti, 2018)	
	Application for public transport	(Beul-Leusmann et al., 2014)	
	Mood assessment application	(Hafiz et al., 2018)	
	Smartphone screen sizes on perceived usability	(Raptis et al., 2013)	
	Mobile recommendation system based on context and personality	(Braunhofer et al., 2014)	
	Logbook dedicated to physical exercise	(Heinonen et al., 2012)	
Application in the field of Health	Online health record on mobile phones	(Fritz et al., 2012)	
	Mobile health system for diabetes	(Georgsson & Staggars, 2016)	
	Use of the Nintendo Wii Fit Plus® for the treatment of balance disorders in vestibular and neurological diseases	(Meldrum et al., 2012)	
	Remote monitoring of the health of older people and people suffering from heart failure	(Evans et al., 2016)	
	VibroTactile Feedback System for Stroke Survivors	(Held et al., 2017)	
	Health Awareness Web Portal	(Hussain et al., 2015)	
	Tool for Reducing Inappropriate Diagnosis (STRIP)	(Meulendijk et al., 2015)	
	Comparison of reporting in telemedicine	(Lacerda et al., 2014)	
	Emergency services information system by nurses	(Kim et al., 2012)	
	Customized sensor-based system for pervasive healthcare	(Triantafyllidis et al., 2014)	
	Online telehealth system	(Dhillon et al., 2012)	
	Electronic Diabetes Management System	(Lutes et al., 2006)	
	Gamification and Serious Game	Gamified project management system, Core	(Ašeriškis & Damaševičius, 2014)
		Serious Game for Health	(Tolentino et al., 2011)
Balance exercises for seniors		(Nawaz et al., 2014)	
Comparison of the Shadow serious game system with a traditional SHADE system		(Hookham et al., 2016)	
Learning platform (e-learning)	koala	(Blecken et al., 2010)	
	Measuring acceptance of the eClass learning management system	(Revythi & Tselios, 2019)	
	CIRCLE classification learning environment	(Peruri et al., 2017)	
	Augmented reality assisted solid geometry learning system	(Lin et al., 2015)	
	eClass and Moodle learning management systems	(Orfanou et al., 2015)	
	Affective tutoring system in remedial accounting education	(Lin et al., 2014)	
	e-Learning environment based on a predefined task model describing low-level interactions	(Harrati et al., 2016)	
	Interactive learning system, based on augmented reality, for fish conservation	(Lin et al., 2011)	
Expert systems	Augmented reality system for teaching Euclidean vectors	(Martin-Gonzalez et al., 2016)	
	Medical secretarial system	(South et al., 2017)	
	Banking biometric system	(Tassabehji & Kamala, 2012)	
	Software process modeling system	(Kerzazi & Lavallee, 2011)	
	Monitoring and visualization of collective behavior on a large scale (NationTelescope)	(Yang et al., 2015)	
Other	Online dialogue management system (AdChat)	(Aguilar-Reyes et al., 2015)	
	Platform for young people to explore their interests in the digital society (WYRED)	(García-Peñalvo et al., 2018)	
	Safety signs	(Ng et al., 2011)	
	GPG client	(Ruoti et al., 2015)	
	Interactive map design	(Çöltekin et al., 2009)	
	Voice-based user interfaces	(Ghosh et al., 2018)	
	Portable fitness equipment	(Liang et al., 2018)	
	Augmented reality for children in a confirmation task	(Munsinger & Quarles, 2019)	
Comparison of user efficiency for tablet and smartphone users	(Botella et al., 2014)		
Measuring the impact of smartphone screen size on perceived usability	(Raptis et al., 2013)		