

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 67 (2015) 293 – 300

Procedia
Computer Science

6th International Conference on Software Development and Technologies for Enhancing
Accessibility and Fighting Infoexclusion (DSAI 2015)

European Portuguese validation of the System Usability Scale (SUS)

Ana Isabel Martins^{a*}, Ana Filipa Rosa^a, Alexandra Queirós^{a,b}, Anabela Silva^{b,c},
Nelson Pacheco Rocha^{a,d}

^a *Institute of Electronics and Telematics Engineering of Aveiro (IEETA), Campus Universitário, 3810 Aveiro, Portugal*

^b *Health Sciences School, University of Aveiro (ESSUA), Campus Universitário, 3810 Aveiro, Portugal*

^c *Center for Health and Technology Research (CINTESIS) University of Aveiro, Campus Universitário, 3810 Aveiro, Portugal*

^d *Health Sciences Department, University of Aveiro (SACS), Campus Universitário, 3810 Aveiro, Portugal*

Abstract

The System Usability Scale (SUS) is a widely used self-administered instrument for the evaluation of usability of a wide range of products and user interfaces. The principal value of the SUS is that it provides a single reference score for participants' view of the usability of a product or service. This paper presents the translation, cultural adaptation and a contribution to the validation of the European Portuguese version of SUS. The conducted work comprised two phases, the scale translation, and the scale validation. The first phase resulted in a European Portuguese version equivalent to the original in terms of semantic and content. The second phase involved the assessment of the validity and reliability of the scale. The instrument has construct validity as it presents a high and significant correlation with other two usability metrics, the Post-Study System Usability Questionnaire (PSSUQ) ($r = 0.70$) and a general usability question ($r = 0.48$). The reliability results show less than satisfactory ICC values (ICC = 0.36), however the percentage of agreement is satisfactory (76.67%). Further studies are needed to investigate the reliability of the Portuguese version.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2015)

Keywords: System Usability Scale; European Portuguese validation; usability evaluation; user tests

1. Introduction

Usability assessment is an important part of the overall design and development of a product or service, which consists of iterative cycles of prototyping, design and validation [1]. Ideally, the usability evaluation must be present at all stages of the design and development process, and must be iterative in order to allow a continuous improvement of the results.

* Corresponding author. Tel.: +351 234 370 500; fax: +351 234 370 545.

E-mail address: anaisabelmartins@gmail.com

The literature describes a large number of methodologies and tools used to ensure the usability quality of a product or service [2, 3]. There are usability evaluation methods for all design and development phases, from initial definition to the final changes of the product or service [4]. Furthermore, some of these methods are only suitable for a specific stage of the development process [5].

Within the usability evaluation methods described in the literature [2, 3], questionnaires assume a significant importance for qualitative self-reported data collection about the characteristics, thoughts, feelings, perceptions, behaviors or attitudes of users [4]. Questionnaires have the advantage of being low budget techniques, they do not require test equipment, and their results reflect the users' opinions. They also provide useful information about what are the strengths and weaknesses of a product or service.

In the 80s researchers felt the need to develop and evaluate products in a systematic and methodical way, considering questionnaire's psychometric properties [6]. Since then several questionnaires were developed and validated, and have long been used in the evaluation of products and services. However, a research through the major scientific databases related to usability assessment revealed the absence of questionnaires validated for the European Portuguese language. Whereby the evaluation of usability is a key aspect in the development of products and services, it is urgent to have validated questionnaires able to be applied within the Portuguese context.

The fact that the System Usability Scale (SUS) is one of the usability evaluation questionnaires most widely used [7] justifies the importance of having a European Portuguese version. However, in order to guarantee the quality of the resulting questionnaire, it is not enough to perform a literal translation of the SUS original version. A linguistic and cultural adaptation is crucial to ensure that the constructs of each item of the translated questionnaire have the same meaning as the respective original item, and an observational study must be implemented to verify the internal consistence of the resulting questionnaire, as well as its validity (i.e. accuracy of the results) and reliability (i.e. consistency of measurements, namely if they are reproducible).

This paper presents the translation, cultural adaptation and contribution to the validation of the European Portuguese version of SUS and is in line with current efforts of promoting an active involvement of end users in the design and development of new products and services [8]. In addition to this introductory section, the paper comprises four more sections: Related Work, Methodology, Results, and Discussion and Conclusions.

2. Related Work

The System Usability Scale (SUS) was developed by John Brooke more than 25 years ago as part of a usability engineering program (1986) as a "quick and dirty" survey scale that would allow the usability practitioner to quickly and easily assess the usability of a given product or service. The SUS development reflected a strong need in the usability community for a tool that could collect a user's subjective rating of the usability of a product or service [9].

About a decade later the author published a chapter describing SUS in a book about usability engineering in industry [10]. Since then, its use in a variety of research projects and industrial evaluations has increased as SUS has been made freely available for use in usability assessment [10]. SUS is an inexpensive, yet effective tool, for assessing the usability of a product, as well as a wide range of user interfaces, including standard operative system based software interfaces, Web pages and Web applications, cell phones, landline phones, modem and networking equipment, pagers, interactive voice response systems (IVR), speech systems, or video delivery hardware and software. The principal value of SUS is that it provides a single reference score for participants' view of a product's usability [9].

SUS is robust and its use is well established and generalized, with more than 1200 publications [11]. Since the beginning of data collection using SUS in 1996, Bangor and colleagues [9] reported 2324 surveys over the course of 206 studies. For this reason SUS is now considered an "industry standard" [11].

The SUS has several attributes that make it a good choice for general usability practitioners [9]:

- The survey is technology free, making it flexible enough to assess a wide range of technologies.
- It is quick and easy to fill.

- It provides a single score on a scale that is easily understood by the wide range of people (from project managers to computer programmers) who are typically involved in the development of products and services and who may have little or no experience in human factors and usability.
- It is a nonproprietary survey, making it a cost effective tool.

SUS is composed of 10 statements that are scored on a 5 point Likert scale of strength of agreement. Its final score can range from 0 to 100, where higher scores indicate better usability. Because the statements alternate between the positive and negative, care must be taken when scoring the survey [9, 10]. The total score of SUS is calculated through the formula defined by Brooke [10]:

- For odd items: subtract one from the user response.
- For even-numbered items: subtract the user responses from 5.
- Add up the converted responses for each user and multiply that total by 2.5. This converts the range of possible values from 0 to 100 instead of from 0 to 40.

SUS scores are not percentages, despite returning a value between 0 and 100. To understand how a product compares to others, it's necessary to look at its percentile ranking. A SUS score above 68 would be considered above average and anything below 68 is below average [11].

Originally, the SUS was used in one-time, isolated tests to determine a single usability and satisfaction score for a given product or service. As its use within human factors groups became more frequent, more uses became evident and were adopted. Over the last 10 years it has been used to:

- Provide a point estimate measure of usability and customer satisfaction.
- Compare different tasks within the same interface.
- Compare iterative versions of the same system.
- Compare competing implementations of a system.
- Competitive assessment of comparable user interfaces.
- Compare different interface technologies.

While SUS was only intended to measure perceived ease-of-use (a single dimension), recent research shows that it provides a global measure of system satisfaction [12]. Furthermore, it can be divided into two sub-scales of usability and learnability: Usable (8 items) and Learnable (2 items - specifically, Items 4 and 10). These subscales correlate highly with the overall SUS ($r = .99$ and $.78$, respectively) and correlate significantly with one another ($r = .66$) [12].

This scale proved to be applicable to different languages, and has been translated into Spanish, French, and Dutch [13]. The translated versions were shown to have similar reliability to the original English version [13]. Therefore, the objective of this study is to translate, adapt to the Portuguese culture and contribute to the validation of the European Portuguese version of SUS.

3. Methodology

For a measuring instrument to be applied in different cultures, it is important to ensure that the translations and adaptations are equivalent. The cross-cultural adaptation of an instrument involves two main steps: i) Scale translation - assessment of the conceptual and linguistic equivalence, and ii) Scale Validation - evaluation of the psychometric properties [14].

The translation, cultural and linguistic adaptation of the SUS to Portuguese followed the internationally established guidelines [15], in order to ensure the quality of the resulting translation and the semantic equivalence (i.e. the consistency of the meaning of concepts between this version and the original) [14].

3.1. Phase I - Scale Translation

The translation process of the original version of SUS was performed in accordance with the internationally established guidelines [16] and involved the following steps:

- Step 1 (Translation): The original version of SUS was translated to European Portuguese by two independent translators, a health professional and a technology professional, whose native language is European Portuguese.
- Step 2 (Reconciliation version): Three researchers compared the two translations and built a reconciliation version between them and the original version of SUS.
- Step 3 (Retroversion): The reconciliation version was translated from European Portuguese into English by a translator whose native language is English, without training in health and without knowledge of the original versions of SUS. The retroversion was made to confront the original version with the translated one and analyze if both are equivalent.
- Step 4 (Pre-final version): A committee of three researchers developed the pre-final versions of SUS based on the back-translation and on the original instrument.
- Step 5 (Pilot Test): The pre-final version was submitted to a pilot test with 4 individuals of the general population to assess the easiness / difficulty of understanding of the questions, according to the methodology proposed by Foddy [16]. The information collected was used to improve the tool and build the final version.
- Step 6: The back-translation and the description of the methodology used in the translation process were sent, as a courtesy, to the author of the original version.

3.2. Phase II - Scale Validation

For the scale validation, an observational study was performed in a church-sponsored charity institution, the *Santa Casa da Misericórdia de Oliveira do Bairro*. The reliability and validity of the SUS was based on real data collection. This process consisted of a usability assessment of a web based Agenda. An agenda application was selected because it is an easy concept that virtually everyone understands and are familiar with.

The usability was evaluated with SUS and another usability scale, the Post-Study System Usability Questionnaire (PSSUQ) that was validated for European Portuguese, together with a general usability question. The PSSUQ is a usability evaluation questionnaire based on scenarios that was developed by IBM. It consists of 19 items aimed at addressing five usability characteristics of a system: rapid completion of the task, ease of learning, high quality documentation and online information, functional adequacy and rapid acquisition of productivity [17].

The general usability question evaluates global usability using just one question was based on the After-Scenario Questionnaire (ASQ) developed by Lewis [18]: overall, how would you rate the application on a scale from 0 to 10? This question is accompanied by a numeric rating scale from 0 to 10, in order to facilitate the completion of the question.

The sample was selected according to the following inclusion criteria: age over 18 years, ability to read, understand and sign the informed consent. The only exclusion criterion was the presence of limitation on the thick upper limb movements.

The observational study took place between November, 2013 and February, 2014, and comprised two sessions separated by 2 to 4 weeks for each subject. The sessions consisted of four parts:

- Introduction - The evaluator applied a social demographic questionnaire and then delivered the session script, explaining orally all information contained therein.
- Test - The subject performed the tasks described in the session script.
- Usability Assessment Instruments - The evaluator assisted the instruments filling, namely the SUS, PSSUQ and a general usability question.
- Summary - The evaluator thanked the participation of the subject and, if necessary, scheduled the next evaluation session.

All adults able to fill in the instruments used in this usability assessment were eligible to participate if they gave written informed consent. The written informed consent was obtained prior to data collection.

In order to address the ethical considerations, a request of approval was directed to the administration of the *Santa Casa da Misericórdia de Oliveira do Bairro*.

Prior to the translation and validation of the SUS we requested the authors of the original version for permission.

Among that, all data collection was anonymized. The involved subjects received all the information regarding the study and their participation before completing the informed consent.

The user was asked to use the web based agenda in order to check the different tasks assigned for each day of the week, interacting freely with the application. Each user explored the Agenda for about 10 minutes.

The results and interpretation of the SUS were conducted within the defined parameters for the original scale.

To describe and characterize the subjects who constitute the population sample, central tendency and dispersion measures were used, including mean, range and standard deviation. The inter-rater reliability was assessed using Intraclass Correlation Coefficient (ICC) and percentage of agreement. The ICC varies between 0 and 1, and is considered "weak" ($ICC < 0.40$); "satisfactory" ($0.40 \leq ICC < 0.75$); and "very good" ($ICC \geq 0.75$). A percentage of agreement above 80% is considered acceptable. To calculate the percentage of agreement the qualitative score (usable vs. non-usable) of each SUS evaluation (first and second evaluation moment) was compared. In other words, we counted the number of participants that classified the web application as usable and non-usable (based on the SUS cut-off point defined by the author, i.e., 68 points) in both evaluation moments, and calculated the percentage of agreement. The correlation between the SUS and the two other scales was assessed by a Spearman Correlation Coefficient. The level of significance was set at $p < 0.05$.

Statistical analyzes were performed with SPSS - Statistical Package for Social Sciences (SPSS Inc, Chicago).

4. Results

4.1. Phase 1 - Scale Translation

The retroversion and the original scale were compared and considered equivalent in semantic and content meaning. The pilot study was conducted with 4 participants from the general community, 2 male and 2 female, aged between 26 and 58 years. The participants understood the semantic and the content of each item. However they found that the SUS alternative items (positive/negative) may turn the scale slightly more difficult to fill. Even so the authors decided to maintain the alternative items identical to the original scale. In general, the European Portuguese version of SUS was considered easy to understand.

The translation process resulted in 10 items that were considered equivalent to the corresponding items of the original SUS (Table 1).

Original Item	Corresponding item in Portuguese
I think that I would like to use this system frequently.	Acho que gostaria de utilizar este produto com frequência.
I found the system unnecessarily complex.	Considereei o produto mais complexo do que necessário.
I thought the system was easy to use.	Achei o produto fácil de utilizar.
I think that I would need the support of a technical person to be able to use this system.	Acho que necessitaria de ajuda de um técnico para conseguir utilizar este produto.
I found the various functions in this system were well integrated.	Considereei que as várias funcionalidades deste produto estavam bem integradas.
I thought there was too much inconsistency in this system.	Achei que este produto tinha muitas inconsistências.
I would imagine that most people would learn to use this system very quickly.	Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este produto.
I found the system very cumbersome to use.	Considereei o produto muito complicado de utilizar.
I felt very confident using the system.	Senti-me muito confiante a utilizar este produto.
I needed to learn a lot of things before I could get going with this system.	Tive que aprender muito antes de conseguir lidar com este produto.

Table 1 - Original item vs. corresponding item in European Portuguese

4.2. Phase II - Scale Validation

The sample consisted of 32 participants with an average age of 47 years (SD = 14.56). The sample was mostly female 87.50%. The characterization of the sample is described in Table 2.

Gender n (%)	Female	28	(87.50)
	Male	4	(12.50)
Age (years)	Mean (SD)	47	(14.56)
	Min - Max	20 – 84	

Table 2 - Sample characterization

Regarding the reliability results, the ICC value is 0.36 which indicates weak inter-rater reliability (CI 95% - 0.01; 0.63), while the percentage of agreement is 76.67%, 23 of the 30 participants that completed the SUS twice agreed with themselves in relation to the usable/non-usable usability of the web application.

Regarding the scale validation phase, results show that the SUS and PSSUQ presented a correlation of $r = 0.70$ ($p < 0.05$) which suggests that both evaluate the same construct that indicates construct validity. Similarly, the correlation between SUS and the general usability question ($r = 0.48$, $p < 0.05$) also indicates that the Portuguese version of SUS has construct validity. This results show that the correlation between the SUS and PSSUQ and between the SUS and the general usability question is high and statistically significant.

5. Discussion and Conclusions

This study aimed to translate and adapt the original version of SUS into the European Portuguese language and culture and contribute to its validation. The results of the scale translation phase indicate that the items were easy to

understand and there were no semantic or content problems. The translated items were considered equivalent to the original version.

This translation, cultural adaptation and contribution to the validation of the European Portuguese version of SUS is in line with the authors work that has been focused in involving the end user in the development and validation of new products and services [8, 19-21]. In this sense, this validation represents a step forward to the systematic and methodical data collection from users' self-reported opinions.

The reliability results show weak ICC values ($ICC = 0.36$), while the percentage of agreement is satisfactory (76.67%). A possible explanation for the low ICC values is the fact that SUS has inverse items, one is in the positive sense and the next one is in the negative, alternately. The reason why the author constructed the scale with alternated items was to avoid response biases, especially as the questionnaire invites rapid responses by being short. By alternating positive and negative statements, the goals of the author were to have respondents read each statement and make an effort to think whether they agreed or disagreed with it [11]. However, the insights produced reveal that in the case of this study, this may have been a disadvantage, as it may have induced filling errors when using this scale. The sample used in this study is not, probably, used to fill this kind of instruments, and the fact that the participants need to have to decide if they agree or disagree with a positive sentence, followed by a negative one, and may effectively induce mistakes as it is complex logic reasoning. In fact it is not the first time that this problem is mentioned in the literature. Sauro and Lewis [22] found that there were some problems both on the part of users in responding to questions and of researchers in scoring the questionnaires. They developed an "all-positive" version of SUS that they suggested addressed these problems without affecting the validity of SUS.

Considering the percentage of agreement (76.67%) the Portuguese version of SUS can be used to distinguish between usable and non-usable applications. However, having in consideration the low reliability values ($ICC = 0.36$), the authors intend to validate the "positive" version of SUS in this population and repeat this study with a larger sample size.

Funding

This study was partially funded by Project QREN N° 13852 I&D «AAL4ALL - Ambient Assisted Living for All», UI 127/94 - IEETA.

References

- [1] M. Ivory and M. Hears. The State of the Art in Automating Usability Evaluation of User Interfaces. *J. ACM Comput. Surv.*, 2001, 33 (4), pp. 470-516.
- [2] A. I. Martins, A. Queirós, N. P. Rocha and B. S. Santos. Avaliação de Usabilidade: Uma Revisão Sistemática da Literatura. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 2013, 11(1), pp. 31-44.
- [3] A. Martins, A. Queirós and N. Rocha. Usability evaluation of products and services: a systematic review. *IADIS International Conference e-Society 2013*, 2013.
- [4] B. Hanington and B. Martin, *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Beverly, MA: Rockport Publishers, 2012.
- [5] I. L. Nunes. *Usability of Interfaces, Ergonomics - A Systems Approach*. InTech, 2012.
- [6] M. Sweeney and A. Dilton. Methodologies employed in the psychological evaluation of HCI. *Human-Computer Interaction - INTERACT '87*, 1987.
- [7] J. R. Lewis. *Usability Testing*. IBM Software Group, p. 71, 2006.
- [8] A. Teixeira et al.. A new living lab for usability evaluation of ICT and next generation networks for elderly@home. 1st International Living Usability Lab Workshop on AAL Latest Solutions, Trends and Applications, AAL 2011, in Conjunction with BIOSTEC 2011, 2011.

- [9] A. Bangor, P. T. Kortum and J. T. Miller. An empirical evaluation of the System Usability Scale. *Int. J. Hum. Comput. Interact*, 2008, 24 (6), pp. 574-594.
- [10] J. Brooke. SUS - A quick and dirty usability scale. *Usability Eval. Ind.*, 1996, 189 (194), pp. 4-7.
- [11] W. From. SUS: a retrospective. *J. Usability Stud.*, 2013, 8(2), pp. 29-40.
- [12] J. R. Lewis and J. Sauro. The Factor Structure of the System Usability Scale. *Human Centered Design, Human Centered Design*. Springer Berlin Heidelberg, 2009, pp. 94-103.
- [13] J. Sauro, *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Measuring Usability LLC, 2011.
- [14] P. Ferreira and F. Marques. *Avaliação psicométrica e adaptação cultural e linguística de instrumentos de medição em saúde: princípios metodológicos gerais*. 1998.
- [15] D. Beaton, C. Bombardier, F. Guillemin and M. Ferraz. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 2000, 25 (24), pp. 3186-3191.
- [16] W. Foddy and W. Foddy. *Constructing questions for interviews and questionnaires: theory and practice in social research*. Cambridge university press, 1994.
- [17] J. R. Lewis, Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies, *Int. J. Human-Computer Interac*, 2002, 14 (3&4) pp. 463-488.
- [18] J. R. Lewis. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. *ACM SIGCHI Bull.*, 1990, 23 (1), pp. 78-81.
- [19] A. Queirós, A. Silva, J. Alvarelhão, N.P. Rocha and A. Teixeira. Usability, accessibility and ambient-assisted living: A systematic literature review. *Universal Access in the Information Society*, 2015, 14 (1), pp. 57-66.
- [20] A. Teixeira et al.. New telerehabilitation services for the elderly. *Handbook of Research on ICTs for Healthcare and Social Services: Developments and Applications*. IGI Global, 2013.
- [21] A. Queirós et al.. ICF inspired personas to improve development for usability and accessibility in ambient assisted living. *Procedia Computer Science*, 2014, 27, pp. 409-418.
- [22] J. Sauro and J. R. Lewis. When designing usability questionnaires, does it hurt to be positive? 2011 Annual Conference on Human Factors in Computing Systems - CHI'11, 2011.