

Análisis comparativo de la Escala de Usabilidad del Sistema (EUS) en dos versiones

Comparative analysis of the System Usability Scale (SUS) in two versions

Análise comparativa usabilidade do sistema Scale (SUE) em duas versões

María Isolde Hedlefs Aguilar

Universidad Autónoma de Nuevo León, México

lab.cognitive@gmail.com

Andrea Abigail Garza Villegas

Universidad Autónoma de Nuevo León, México

agarza.v@hotmail.com

Resumen

Se adaptó la EUS en sus dos versiones: original y positiva, en una población mexicana donde se habla el español, con el objetivo de que después puedan ser utilizadas en otras investigaciones. Para cada versión se utilizaron dos muestras, la versión original se aplicó a 243 participantes y la positiva se aplicó a 173 participantes. El coeficiente de Alpha de Cronbach para la versión original fue de .59 y para la positiva de .92, indicando que esta última tiene una muy buena confiabilidad. El AFE en la versión positiva mostró la presencia de dos factores, como menciona la literatura, que fueron Usabilidad y Facilidad de Aprendizaje, cumpliendo así con la validez de constructo. Se concluye que la versión positiva es mucho más confiable a la hora de su aplicación y que muestra una estructura factorial más acorde con la literatura.

Palabras clave: usabilidad, EUS, estandarización, confiabilidad, validez.

Abstract

We adapted the SUS in its two versions: original and positive, in a Mexican town where the Spanish is spoken, with the aim that can then be used in other research. Two samples were used for each version, the original version was applied to 243 participants and the positive applied 173 participants. Cronbach's Alpha coefficient for the original version was 0.59 and the positive of 0.92, indicating that the latter has a very good reliability. The AFE in the positive version showed the presence of two factors, as the literature mentions, that were Usability and Ease-of-Learning, thus fulfilling the construct validity. It is concluded that the positive version is much more reliable at the time of his application and showing a factorial structure more aligned to the literature.

Key Words: usability, EUS, standardization, reliability, validity.

Resumo

EUS foi adaptado em duas versões: original e positivo, em uma cidade mexicana, onde o espanhol é falado, com o objectivo de que mais tarde pode ser usado em outras investigações. Para cada versão foram utilizadas duas amostras, a versão original foi aplicado a 243 participantes e o positivo foi aplicado a 173 participantes. O coeficiente de Cronbach Alpha para a versão original foi 0,59 e 0,92 para o positivo, indicando que este último tem uma boa confiabilidade. A AFE na versão positiva mostrou a presença de dois fatores, como mencionado na literatura, que foram usabilidade e facilidade de aprendizagem, cumprindo assim a validade de construto. Conclui-se que a versão positiva é muito mais confiável quando a sua aplicação e que mostra uma estrutura fator mais consistente com a literatura.

Palavras-chave: usabilidade, EUS, padronização, confiabilidade, validade.

Fecha recepción: Enero 2016

Fecha aceptación: Julio 2016

Introduction

One of the first scales that emerged to assess the usability of an interface and where it was not necessary for the participant to perform lab tests, It was the Escala de Usabilidad del Sistema (EUS), whose English name is System Usability Scale (SUS) and it was elaborated by Brooke in 1986, which consisted of 10 items (5 items positive and 5 negative items).

The construction of the items from the SUS originated from the definition of usability based on the ISO 9241-11: “the degree in which a product can be used for specific users to achieve goals determined with effectiveness, efficiency and satisfaction in a particular context of use”. The effectiveness is defined as the degree of accuracy and completeness of user to achieve specific goals, while efficiency is defined as the degree in which resources are used so that user will achieve your targets with precision and completeness, and satisfaction is defined as the freedom of the user to be uncomfortable or show positive attitude using the product (ISO, 1998). In other words, effectiveness means that the user accomplishes the task, efficiency that the user perform the task as soon as possible, and satisfaction that the user feels when using the system or interface.

Therefore, the scale SUS has two major objectives. The first is that researchers can obtain a measure of the perception of the usability of a system, and the second that SUS scale does not require much time for its application (Brooke, 2013). Thus, this instrument provides an assessment of the perception about the usability of a system in a short time.

Unique measurements reported Brooke (1996) were that this scale showed high levels of correlation between the 10 items that comprise it. These ranged from $r = +0.7$ to $r = +0.9$, correlation which served to select items that would form the final scale, but here neither the reliability nor validity of the scale was reported.

One of the first reported the scale psychometric data were reliability measurements. With a participation of 77 people, their results showed a coefficient Alpha of 0.85 (Lucey, 1991 in Kirakowski, 1994). Then, in 2008, Bangor, Kortum and Miller applied this scale to a larger sample, which was formed by 324 2 and here was found to be the Alpha coefficient of 0.91, another very good reliability. Later, Lewis and Sauro (2009) were 324 applications and obtained a coefficient Alpha of 0.92. As we can see, the EUS showed from the start have very good Alpha coefficients, which means that it has very good reliability.

Another measurement that was needed to know if within the scale factors or missing. The first who performed these measurements were Bangor, Kortum and Miller (2008), on the same sample (2324) for reliability. They conducted a factor analysis and found one factor, they called Usability General (Lewis and Sauro, 2009; Borsci, Federici and Lauriola, 2009). Lewis and Sauro said in 2009 that it was likely that the group of Bangor did not see the possibility of this scale could have more than one factor.

Later, themselves (2009) made an investigation to know the factorial structure of the EUS, with a sample of 324 subjects, who then compared with those of Bangor et al (2008, Lewis and Sauro, 2009); that way they sought to obtain new data with both reliability and validity of EUS. In this investigation it was found that the scale there are two factors, not one group of Bangor he said. Hence the first factor corresponding to items 1, 2, 3, 5, 6, 7, 8, 9, called Usability; and the second factor to items 4 and 10, which called Ease of Learning (Lewis and Sauro, 2009).

Finally, Borsci, Federici and Lauriola (2009), with a sample of 196 Italian participants, showed that the two factors (usability and ease of learning) are independent; they presented a less restrictive model with correlated factors whose information was opposed to the work of Lewis and Sauro, who found no such correlation between factors. What I observed was that applying the original scale (with 5 positive and 5 negative items items) had some disadvantages, and some participants reported at the end that negative items answered strongly agree or misinterpreted the item. It also happened that the researchers coded bad negative items, which then conducted an investigation to compare the original scale (positive and negative version) with the positive version (only positive items), analyzing the following points: 1) whether the questionnaire Original bias acquiescence (tendency to answer yes to something, regardless of the content of the question), 2) know how big this bias and 3) whether developing a scale with positive items can be removed occurs bias acquiescence and extreme responses. The results obtained by this research were that there was no significant difference between the original scale and positive, so that the rewording (transform negative to positive items) did not affect the measurements of EUS and Cronbach's alpha for both versions (original and positive), which was high ($> .90$); It not found in data acquiescence or a strong bias in extreme responses. However, if the recurring problem of misinterpretation of the items by the participants and the researcher coding errors, causing bias in the responses of the scale was presented.

Because of this, Lewis and Sauro (2011) pointed out that you can use the positive version of EUS with confidence as the positive scales the user is less likely to make mistakes when answering and the researcher will not make a coding error but most important is that scores of positive version are similar to the standards of the original version.

On the other hand, Finstad (2006) conducted an investigation which compared two samples of 18 participants each; in the first sample she had native English speakers and native shows the second foreign language (Russian, German, Chinese, Filipino, Spanish and Hebrew). It was found that non-English speaking natives struggled to understand the item 8 ("I found the system very cumbersome to use"), especially the word Cumbersome, so asked for help to understand it. In this sense, Finstad (2006) and Lewis and Sauro (2009) requested that the word Cumbersome by Awkward, because if the participants did not come to understand well the item could affect the results should be replaced. Finally, Finstad (2006) mentioned that the SUS scale should not apply in English and electronically to people whose native language is other, because some terms induce wrong answers.

Our goal is to adapt the Spanish both the original scale (5 positive and 5 negative items items) as the positive version; and investigate the reliability of the versions both positive and negative items.

Method

Participants

The original version of EUS (positive and negative items) was applied to 243 participants, of whom 129 were female and 114 were male, with an average age of 21 years. Participants studied four races: 121 Engineering Management Systems (IAS), 4 Software Engineering Technology (ITS), 33 Administrative Mechanical Engineering (IMA) and 85 Degree in Psychology. The positive version to 173 participants, 99 female and 74 male was applied, with an average age of 21 years. They studied two races: 103 Engineering Management Systems (IAS) and 70 Degree in Psychology.

Instruments

In the first phase the original EUS, consisting of 5 positive items and 5 negative items (Brooke, 1996), presenting a .91 reliability and construct validity of two factors (usability and ease of learning) (Lewis and applied Sauro , 2009). In the second phase the EUS positive version, which consists of 10 positive items, with a reliability of .96 (Lewis and Sauro, 2012) and the same factorial structure mentioned by them was applied.

the reverse translation procedure was used, with the help of three bilingual usability experts to perform the adaptation of these two scales; two of them translated into Spanish the two scales (original and positive version) and then cross-checked translations. The third expert translated the two scales translated into English. Finally, the three experts checked against their versions to reach an agreement, to obtain the final versions of the two scales.

The two translations (the original version and positive) were online at Google Forms platform, in order to obtain the greatest number of participants. a Likert scale of 5 levels of response, ranging from strongly disagree (1) to strongly agree (5) was used.

some questions were added in the section of participant data, which had to do with age, experience time they were using the platform and, finally, the race they were studying (see both versions in Appendices A and B).

Process

Both the original version and for positive students were invited to participate voluntarily in the research, informing them that their personal information would be treated confidentially. The invitation was made personally by the social network Facebook as well as email. Through the electronic media they were sent a web address from which they could access the Google Forms platform where the two scales were evaluated. Participants took less than five minutes to respond to both the original version and the positive version.

Results

In the original version of EUS coefficient Alpha Cronbach of .59 was obtained, while the Bartlett test of sphericity came significant $p < .001$ and index Kaiser-Meyer-Olkin (KMO) was .84, so we proceeded to carry out the exploratory factor analysis (AFE). This showed that three factors and that the percentage of variance explained was 62.80%. In factor 1 were items 1, 3, 5, 7 and 9, ie,

the positive items, while the factor two items 8, 4 and 10 were, and factor 3 items 2 and 6 . That is, the negative items were distributed among the factors 2 and 3.

That way we proceeded to make an AFE, which resulted in two factors, as the literature points. percentage of explained variance of 52.78, lower than the first AFE The following results were found. In this analysis factor 1 all positive items were placed in factor 2 and all negative items.

In the positive version of EUS a Coefficient of Cronbach Alpha of .92, while the Bartlett test of sphericity out significant $p < .001$ and KMO index was .90, which means that it is feasible to perform the analysis was obtained factorial. Then he proceeded to make the AFE, being one factor and a percentage of explained variance of 58.07%. Figure 1 shows the sedimentation of factors and Table I shows the factor loadings.

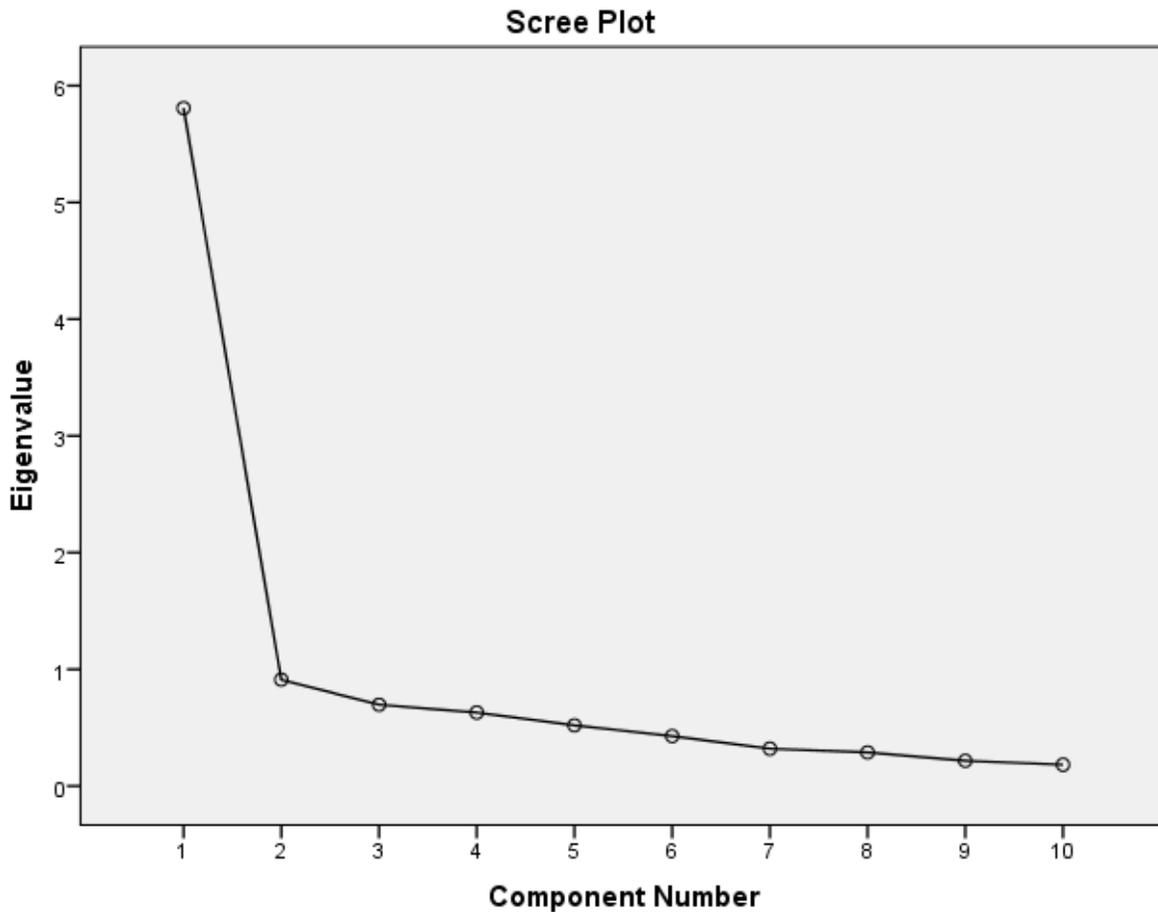


Figure 1. Graph of settling positive version of EUS.

Table I. Matrix components EUS positive version of their respective factor loadings.

Matriz de Componentes	
	<u>Componentes</u>
	<u>1</u>
1	.668
2	.832
3	.818
4	.660
5	.838
6	.694
7	.820
8	.805
9	.796
10	.653

Finally the AFE was done, forcing the analysis to two factors as points literature. We attempted to verify its existence. The AFE found that if the above factors (Figure 1), with an explained variance of 67.18%, is still more justified forced the AFE to two factors have. the two factors with their respective items (Table II) were also found.

Table II. Rotated component matrix positive version of EUS.

Matriz de Componentes Rotados		
	Componentes	
	1	2
1	.821	-.007
2	.736	.398
3	.675	.462
4	.234	.817
5	.730	.419
6	.660	.268
7	.723	.297
8	.760	.317
9	.594	.539
10	.216	.829

In factor 1 (usability), the Cronbach Alpha coefficient was .92, while factor 2 (learnability) the Cronbach Alpha coefficient was .70.

Thus, it appears that the positive version of EUS shows best results in the explained variance and placing the items with their respective factors, as mentioned in the literature.

Discussion and conclusions

With regard to the positive version of EUS, this was very good internal reliability (.92) and also showed that the AFE has good percentage of variance explained and the two factors mentioned literature, which are usability (items 1, 2, 3, 5, 6, 7, 8 and 9) and ease of learning (4 and 10). As it noted in the results, in the original version of EUS was no good coefficient Alpha, but good scores on the test Bartlett sphericity and the KMO index for the AFE, throwing three factors. Still it has not done any research work on this scale report that three factors (Lewis and Sauro, 2009; Borsci, Federici and Lauriola, 2009) are presented. One of the reasons why the coefficient Cronbach Alpha was not the same in such work, may be the inclusion of negative items or that participants found it difficult to understand. For example, Lewis and Sauro (2011) mentioned in his research that these negative items may be errors in responding and coding errors. We believe

that especially generates errors in answering, as there were participants mentioned having committed such error.

Another point to consider is designated by Lewis and Sauro (2011), who argue that both versions of the EUS can be used because both have high levels of internal consistency, but also that it is better to use the positive version of EUS to avoid problems bad coding errors in responding or has the positive version similar to the original scores. However, in our study we could not get that effect on the original version; rather supports data and Frye Stewart (2004), who found that the inclusion of negative and positive items has low internal consistency. Meanwhile, Pilotte and Gable (1990) and Schmitt and Stuits (1985) showed that using mixed items distorts the factorial structure, something that happened in the original version applied.

An important fact is that Lewis and Sauro (2009) reported Alpha coefficients for each of the factors, Factor .91 usability and ease of learning .70. Our study yielded about the same results (usability and ease of learning .92 .70), consistent with the literature, with the only difference that Lewis and Sauro (2009) used the original questionnaire and EUS us the positive version.

In short, the study showed that the positive version of EUS is better than the original version, to be more reliable in statistical terms and have adequate construct validity. Therefore, we make an invitation to the positive version instead of negative when evaluating the different web platforms used.

Bibliography

- Bangor, A., Kortum, P., & Miller, J. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24, 574-594. doi: 10.1080/10447310802205776.
- 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. doi: 10.1007/s10339-009-0268-9.
- Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. In: P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability Evaluation in Industry*. London: Taylor & Francis.
- Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, 8(2), 29-40.
- Finstad, K. (2006). The System Usability Scale and Non-Native English Speakers. *Journal of Usability Studies*, 1(8), 185-188.
- ISO 9241-11 (1998). Ergonomic requirements for office work with visual display terminals (VDTs). Recuperado el 12 de abril de 2016, de <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en>
- Kirakowski, J. (1994). The use of questionnaire methods for usability assessment. Unpublished manuscript. Recuperado el 12 de abril de 2016, de <http://sumi.ucc.ie/sumipapp.html>
- Lewis, J., & Sauro, J. R. (2009). The Factor Structure of the System Usability Scale. *In Human Centered Design*. doi: 10.1007/978-3-642-02806-9_12.
- Lewis, J., & Sauro, J. R. (2012). Quantifying the user experience: Practical statistics for user research. Elsevier, USA. doi:10.1016/B978-0-12-384968-7.00001-1.
- Lewis, J., & Sauro, J. R. (2011). When Designing Usability Questionnaires, Does It Hurt to Be Positive? *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. doi: 10.1145/1978942.1979266.
- Pilotte, W. J., & Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, 50(3), 603-610. doi: 10.1177/0013164490503016.
- Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9(4), 367-373. doi: 10.1177/014662168500900405.

Stewart, T. J., & Frye, A. W. (2004). Investigating the use of negatively phrased survey items in medical education settings: common wisdom or common mistake? *Academic Medicine*, 79(10), 18-20. doi: 10.1097/00001888-200410001-00006.

Apéndice A

Escala de Usabilidad del Sistema (versión original)

	Totalmente en desacuerdo			Totalmente de acuerdo		
1. Creo que me gustaría utilizar frecuentemente este sitio web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Encontré el sitio web innecesariamente complejo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Pienso que el sitio web es fácil de usar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Creo que necesitaré el apoyo de personal técnico para poder utilizar este sitio web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Encontré que varias de las funciones en el sitio web estaban bien integradas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Pensé que había demasiada inconsistencia en este sitio.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Me imagino que la mayoría de las personas podrían aprender a usar este sitio web muy rápido.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Encontré el sitio web muy difícil de usar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Me sentí muy confiado (seguro) al utilizar el sitio web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Necesité aprender muchas cosas antes de poder usar este sitio web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Apéndice B

Escala de Usabilidad del Sistema (versión positiva)

	Totalmente en desacuerdo			Totalmente de acuerdo		
1. Creo que me gustaría utilizar frecuentemente este sitio web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Encontré el sitio web sencillo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Pienso que el sitio web es fácil de usar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Pienso que podré utilizar este sitio web sin el apoyo de personal técnico.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Encontré que varias de las funciones en el sitio web estaban bien integradas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Pensé que había demasiada consistencia en el sitio web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Me imagino que la mayoría de las personas podrían aprender a usar este sitio web muy rápido.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Encontré el sitio web muy intuitivo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Me sentí muy confiado (seguro) al utilizar el sitio web.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Pude utilizar el sitio web sin tener que aprender nada nuevo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>