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

Introducción

Una de las primeras escalas que surgieron para evaluar la usabilidad de una interfaz y donde no era necesario que el participante realizara pruebas de laboratorio, fue la Escala de Usabilidad del Sistema (EUS), cuyo nombre en inglés es System Usability Scale (SUS) y fue elaborada por Brooke en 1986, la cual constó de 10 ítems (5 ítems positivos y 5 ítems negativos).

La construcción de los ítems de la EUS se originó a partir de la definición de usabilidad basada en el ISO 9241-11: "el grado en el que un producto puede ser utilizado por usuarios específicos para lograr metas determinadas con efectividad, eficiencia y satisfacción en un contexto particular de uso". La eficacia es definida como el grado de precisión y totalidad del usuario para lograr objetivos específicos, mientras que la eficiencia se define como el grado en que los recursos son utilizados para que el usuario logre sus objetivos con precisión y totalidad, y la satisfacción se define como la libertad del usuario para mostrarse incómodo o mostrar actitud positiva utilizando el producto (ISO, 1998). Es decir, eficacia significa que el usuario logre la tarea, eficiencia que el usuario realice la tarea lo más rápido posible, y satisfacción la que siente el usuario al utilizar el sistema o la interfaz.

Por tanto, la escala EUS tiene dos grandes objetivos. El primero es que los investigadores puedan obtener una medición de la percepción de la usabilidad de un sistema, y el segundo que la escala EUS no requiera mucho tiempo para su aplicación (Brooke, 2013). Así, este instrumento proporciona una evaluación de la percepción sobre la usabilidad de un sistema en un tiempo reducido.

Las únicas mediciones que reportó Brooke (1996) fueron que esta escala presentaba niveles altos de correlación entre los 10 ítems que la conforman. Estos iban del _+0.7 a _+0.9, correlación que le sirvió para poder seleccionar los ítems que conformarían la escala final, pero aquí no se reportó ni la confiabilidad ni la validez de la escala.

Uno de los primeros datos psicométricos reportados de la escala fueron las mediciones de confiabilidad. Con una participación de 77 personas, sus resultados mostraron un coeficiente de Alpha de .85 (Lucey, 1991 en Kirakowski, 1994). Después, en 2008, Bangor, Kortum y Miller aplicaron esta escala a una muestra más grande, la cual estuvo conformada por 2 324 y aquí se encontró que el Coeficiente de Alpha fue de .91, otra muy buena confiabilidad. Posteriormente, Lewis y Sauro (2009) realizaron 324 aplicaciones y obtuvieron un Coeficiente de Alpha de .92.

Como se puede observar, la EUS demostró desde el inicio tener muy buenos coeficientes de Alpha, lo cual significa que presenta muy buena confiabilidad.

Otra de las mediciones que se necesitaba conocer era si dentro de la escala se encuentran factores o no. Los primeros que realizaron estas mediciones fueron Bangor, Kortum y Miller (2008), con la misma muestra (2 324) para obtener la confiabilidad. Realizaron un análisis factorial y encontraron un solo factor, que denominaron Usabilidad General (Lewis y Sauro, 2009; Borsci, Federici y Lauriola, 2009). Lewis y Sauro señalaron en 2009 que lo más probable era que el grupo de Bangor no vio la posibilidad de que esta escala pudiera tener más de un solo factor.

Más adelante, ellos mismos (2009) hicieron una investigación para poder conocer la estructura factorial del EUS, con una muestra de 324 sujetos, que después compararon con los de Bangor et al (2008, en Lewis y Sauro, 2009); de esa manera buscaron obtener con los nuevos datos tanto la confiabilidad como la validez del EUS. En dicha investigación se encontró que en la escala existen dos factores y no uno como dijo el grupo de Bangor. Ahí el primer factor correspondía a los ítems 1, 2, 3, 5, 6, 7, 8, 9, que denominaron Usabilidad; y el segundo factor a los ítems 4 y 10, que denominaron Facilidad de Aprendizaje (Lewis y Sauro, 2009).

Por último, Borsci, Federici y Lauriola (2009), con una muestra de 196 participantes italianos, mostraron que los dos factores (usabilidad y facilidad de aprendizaje) son independientes; ellos presentaron un modelo menos restrictivo con factores correlacionados cuya información era opuesta a la del trabajo de Lewis y Sauro, quienes no encontraron dicha correlación entre los factores. Lo que sí observaron fue que aplicar la escala original (con 5 ítems positivos y 5 ítems negativos) tenía algunas desventajas, ya algunos participantes reportaron al final que en los ítems negativos contestaron totalmente de acuerdo o que malinterpretaron el ítem. También sucedió que los investigadores codificaron mal los ítems negativos, por lo que después realizaron una investigación para comparar la escala original (versión positiva y negativa) con la versión positiva (solo ítems positivos), analizando los siguientes puntos: 1) si en el cuestionario original se presenta el sesgo de aquiescencia (tendencia a contestar afirmativamente a algo, sin importar el contenido de la pregunta), 2) conocer qué tan grande es este sesgo y 3) ver si elaborando una escala con ítems positivos se puede eliminar el sesgo de aquiescencia y las respuestas extremas. Los resultados obtenidos por esta investigación fueron que no había ninguna diferencia significativa entre la escala original y la positiva, por lo que el cambio de redacción (transformar los ítems negativos a positivos) no afectaba las mediciones del EUS y el Alfa de Cronbach para

ambas versiones (original y positivo), que fue alta (>.90); tampoco se encontró en los datos una fuerte aquiescencia o sesgo en las respuestas extremas. Sin embargo, sí se presentó el problema recurrente de la malinterpretación de los ítems por parte de los participantes y los errores de codificación del investigador, provocando sesgos en las respuestas de la escala.

Debido a ello, Lewis y Sauro (2011) señalaron que puede utilizarse la versión positiva del EUS con confianza, ya que en las escalas positivas el usuario es menos propenso a cometer errores a la hora de contestar y el investigador no realizará un error de codificación, pero lo más importante es que las puntuaciones de la versión positiva son similares a las normas de la versión original.

Por otro lado, Finstad (2006) realizó una investigación donde comparó dos muestras de 18 participantes cada una; en la primera muestra había nativos de habla inglesa y en la segunda muestra nativos de habla no inglesa (rusos, alemanes, chinos, filipinos, españoles y hebreos). Se encontró que los nativos de habla no inglesa tuvieron dificultades para poder comprender el ítem 8 ("I found the system very cumbersome to use"), sobre todo la palabra "cumbersome", por lo que solicitaron ayuda para poder comprenderla. En este sentido, Finstad (2006) y Lewis y Sauro (2009) pidieron que se sustituyera la palabra "cumbersome" por "awkward", ya que si los participantes no llegaban a comprender bien el ítem podía repercutir en los resultados. Por último, Finstad (2006) mencionó que la escala SUS no debería aplicarse en el idioma inglés y de forma electrónica a personas cuya lengua nativa sea otra, pues algunos términos inducen a respuestas equivocadas.

Nuestro objetivo es adaptar el español tanto a la escala original (5 ítems positivos y 5 ítems negativos) como a la versión positiva; así como indagar sobre la fiabilidad de las versiones tanto con ítems positivos como negativos.

Método

Participantes

La versión original del EUS (ítems positivos y negativos) fue aplicada a 243 participantes, de los cuales 129 pertenecían al sexo femenino y 114 al sexo masculino, con una media de edad de 21 años. Los participantes estudiaban cuatro carreras: 121 Ingeniería en Administración de Sistemas (IAS), 4 Ingeniería en Tecnología de Software (ITS), 33 Ingeniería Mecánica Administrativa (IMA) y 85 Licenciatura en Psicología. La versión positiva se aplicó a 173

participantes, 99 del sexo femenino y 74 del masculino, con una media de edad de 21 años. Estudiaban dos carreras: 103 Ingeniería en Administración de Sistemas (IAS) y 70 Licenciatura en Psicología.

Instrumentos

En la primera fase se aplicó el EUS original, que consta de 5 ítems positivos y 5 ítems negativos (Brooke, 1996), presentando una confiabilidad de .91 y validez de constructo de dos factores (usabilidad y facilidad de aprendizaje) (Lewis y Sauro, 2009). En la segunda fase se aplicó el EUS versión positiva, el cual consta de 10 ítems positivos, con una confiabilidad de .96 (Lewis y Sauro, 2012) y misma estructura factorial mencionada por ellos.

Para realizar la adaptación de estas dos escalas se utilizó el procedimiento de traducción inversa, con ayuda de tres expertos bilingües en usabilidad; dos de ellos tradujeron al español las dos escalas (versión original y versión positiva) y luego cotejaron sus traducciones. El tercer experto tradujo las dos escalas traducidas al inglés. Por último, los tres expertos cotejaron sus versiones hasta llegar a un acuerdo, para así obtener las versiones finales de las dos escalas.

Las dos traducciones (la versión original y la positiva) se hicieron en línea en la plataforma de Google Forms, con el fin de poder obtener la mayor cantidad de participantes. Se utilizó una escala Likert de 5 niveles de respuesta, que iban de totalmente en desacuerdo (1) a totalmente de acuerdo (5).

Se añadieron algunas preguntas en el apartado de datos del participante, las cuales tenían que ver con la edad, el tiempo de experiencia que tenían usando la plataforma y, por último, la carrera que estaban estudiando (véanse ambas versiones en los apéndices A y B).

Procedimiento

Tanto para la versión original como para la positiva se invitó a los estudiantes a participar de manera voluntaria en la investigación, haciéndoles saber que su información personal sería tratada de manera confidencial. La invitación se realizó de manera personal a través de la red social de Facebook, así como también por correo electrónico. A través de los medios electrónicos se les envió una dirección web desde la cual podían acceder a la plataforma Google Forms, donde se encontraban las dos escalas a evaluar. Los participantes tardaron menos de cinco minutos en responder tanto la versión original como la versión positiva.

Resultados

En la versión original del EUS se obtuvo un coeficiente de Alpha de Cronbach de .59, mientras que en la prueba de esfericidad de Bartlett salió significativo p < .001 y el índice de Kaiser-Meyer-Olkin (KMO) fue de .84, por lo que se procedió a realizar el análisis factorial exploratorio (AFE). Esto mostró que existen tres factores y que su porcentaje de varianza explicada fue de 62.80 %. En el factor 1 estaban los ítems 1, 3, 5, 7 y 9, es decir, los ítems positivos, mientras que en el factor 2 se encontraban los ítems 8, 4 y 10, y en el factor 3 los ítems 2 y 6. Es decir, los ítems negativos se distribuyeron entre los factores 2 y 3.

De esa manera se procedió a realizar un AFE, que resultó en dos factores, tal como señala la literatura. Se encontraron los siguientes resultados: porcentaje de varianza explicada de 52.78, menor al primer AFE. En este análisis en el factor 1 se colocaron todos los ítems positivos y en el factor 2 todos los ítems negativos.

En la versión positiva del EUS se obtuvo un Coeficiente de Alpha de Cronbach de .92, mientras que la prueba de esfericidad de Bartlett salió significativo p < .001 y el índice KMO fue de .90, lo cual significa que es factible realizar el análisis factorial. Luego se procedió a realizar el AFE, encontrándose un solo factor y un porcentaje de varianza explicada de 58.07 %. La figura 1 muestra la sedimentación de los factores y la tabla I las cargas factoriales.

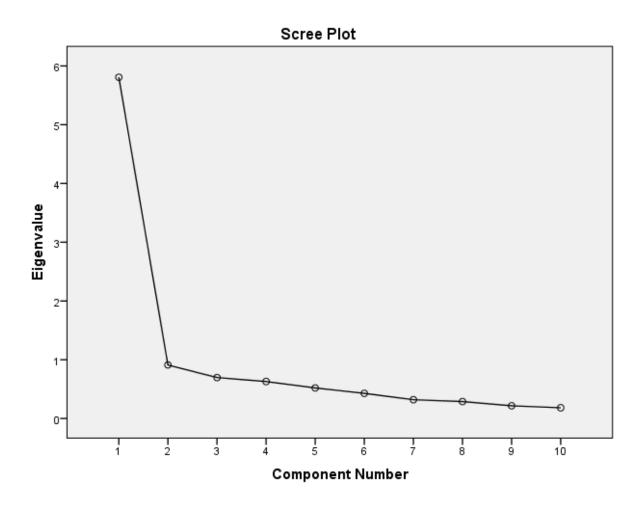


Figura 1. Gráfica de sedimentación versión positiva del EUS.

Tabla I. Matriz de componentes de la versión positiva del EUS con sus respectivas cargas factoriales.

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

Por último se realizó el AFE, forzando dicho análisis a dos factores tal como señala la literatura. Se intentó comprobar su existencia. El AFE arrojó que si se muestran los factores antes mencionados (figura 1), con una varianza explicada de 67.18 %, se justifica todavía más el haber forzado el AFE a dos factores. También se encontraron los dos factores con sus respectivos ítems (tabla II).

Tabla II. Matriz de componentes rotados de la EUS versión positiva.

Mat	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					

En el factor 1 (usabilidad), el Coeficiente de Alpha de Cronbach fue de .92, mientras que en el factor 2 (facilidad de aprendizaje) el Coeficiente de Alpha de Cronbach fue de .70.

Así, se observa que la versión positiva del EUS muestra mejores resultados tanto en la varianza explicada como en la colocación de los ítems con sus respectivos factores, tal como se menciona en la literatura.

Discusión y conclusiones

Con respecto a la versión positiva del EUS, esta mostró muy buena confiabilidad interna (.92) y también mostró que el AFE presenta buen porcentaje de varianza explicada así como los dos factores que menciona la literatura, los cuales son usabilidad (ítems 1, 2, 3, 5, 6, 7, 8 y 9) y facilidad de aprendizaje (4 y 10). Como se observó en los resultados, en la versión original del EUS no hubo un buen Coeficiente de Alpha, pero sí buenas puntuaciones en la prueba de esfericidad de Bartlett y el índice KMO para realizar el AFE, arrojando tres factores. Todavía no se ha hecho ningún trabajo de investigación que reporte que en esta escala se presentan tres factores (Lewis y Sauro, 2009; Borsci, Federici y Lauriola, 2009). Una de las razones por las que el Coeficiente de Alpha de Cronbach no fue igual en los trabajos mencionados, puede ser la inclusión de ítems negativos o que los participantes encontraron difícil comprenderlos. Por

ejemplo, Lewis y Sauro (2011) mencionan en su investigación que en estos ítems negativos puede haber errores al responder y también errores de codificación. Nosotros creemos que sobre todo genera errores al responder, ya que hubo participantes que mencionaron haber cometido ese tipo de error.

Otro punto a analizar es lo señalado por Lewis y Sauro (2011), quienes afirman que ambas versiones del EUS pueden ser utilizadas porque las dos tienen altos niveles de consistencia interna, pero también que es mejor utilizar la versión positiva del EUS para evitar problemas de mala codificación, errores al responder o que la versión positiva tenga puntuaciones similares a la versión original. Sin embargo, en nuestro estudio no se pudo obtener ese efecto en la versión original; más bien apoya los datos de Stewart y Frye (2004), quienes hallaron que la inclusión de ítems negativos y positivos presenta baja consistencia interna. Por su parte, Pilotte y Gable (1990), y Schmitt y Stuits (1985) demostraron que utilizar ítems mixtos distorsiona la estructura factorial, algo que sucedió en la versión original aplicada.

Un dato importante es que Lewis y Sauro (2009) reportaron Coeficientes de Alpha para cada uno de los factores, siendo en factor de usabilidad de .91 y en la facilidad de aprendizaje de .70. Nuestro estudio obtuvo resultados casi iguales (usabilidad .92 y facilidad de aprendizaje .70), coincidentes con la literatura, con la única diferencia de que Lewis y Sauro (2009) usaron el cuestionario original del EUS y nosotros la versión positiva.

En síntesis, el estudio demostró que la versión positiva del EUS es mejor que la versión original, al ser más confiable en términos estadísticos y tener una adecuada validez de constructo. Por lo tanto, hacemos una invitación para que se utilice la versión positiva en lugar de la negativa al momento de evaluar las diferentes plataformas web.

Bibliografía

- 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	
Creo que me gustaría utilizar frecuentemente este sitio web.	0	0	0	0	0
2. Encontré el sitio web innecesariamente complejo.	0	0	0	0	0
3. Pienso que el sitio web es fácil de usar.	0	Ο	0	0	0
4. Creo que necesitaré el apoyo de personal técnico para poder utilizar este sitio web.	0	0	0	0	0
5. Encontré que varias de las funciones en el sitio web estaban bien integradas.	0	0	0	0	0
6. Pensé que había demasiada inconsistencia en este sitio.	0	Ο	0	0	0
7. Me imagino que la mayoría de las personas podrían aprender a usar este sitio web muy rápido.	0	0	0	0	0
8. Encontré el sitio web muy difícil de usar.	0	Ο	0	0	0
9. Me sentí muy confiado (seguro) al utilizar el sitio web.	0	0	0	0	0
10. Necesité aprender muchas cosas antes de poder usar este sitio web.	0	0	0	0	0

Apéndice B Escala de Usabilidad del Sistema (versión positiva)

	Totalmente en desacuerdo		Totalmente de acuerdo		
Creo que me gustaría utilizar frecuentemente este sitio web.	0	0	0	0	0
2. Encontré el sitio web sencillo.	0	0	0	0	0
3. Pienso que el sitio web es fácil de usar.	0	0	0	0	0
4. Pienso que podré utilizar este sitio web sin el apoyo de personal técnico.	0	0	0	0	0
5. Encontré que varias de las funciones en el sitio web estaban bien integradas.	0	0	0	0	0
6. Pensé que había demasiada consistencia en el sitio web.	0	0	0	0	0
7. Me imagino que la mayoría de las personas podrían aprender a usar este sitio web muy rápido.	0	0	0	0	0
8. Encontré el sitio web muy intuitivo.	0	0	0	0	0
9. Me sentí muy confiado (seguro) al utilizar el sitio web.	0	0	0	0	0
10. Pude utilizar el sitio web sin tener que aprender nada nuevo.	0	0	0	0	0