



Calidad

y

Sistemas de Manufactura

Ingeniantes

Estudio comparativo de cuestionarios para la evaluación de la usabilidad en software



Colaboración

Manuel A. Barajas Bustillos; Rosa María Reyes M; Jorge de la Riva R., Instituto Tecnológico de Cd. Juárez; Aidé Maldonado Macías, Universidad Autónoma de Cd. Juárez

RESUMEN: La usabilidad es un atributo de calidad que cada vez es más importante en el desarrollo de software y que es importante que sea evaluada. La usabilidad puede ser evaluada de diversas maneras, los cuestionarios son una de las formas más populares de evaluar la usabilidad, por lo que se eligieron 4 de los cuestionarios de evaluación de la usabilidad para el software, disponibles en español: QUIS, SUMI, CSUQ y SUS. Se aplicaron en dos grupos diferentes en donde se evaluaron dos programas diferentes. En base a los datos recolectados, se comprobó por medio del alfa de Cronbach que los cuatro presentan un nivel de confiabilidad mayor de 0.7 sin importar el número de preguntas.

PALABRAS CLAVE: Ingeniería de software, Calidad de los programas, Usabilidad, Alfa de Cronbach, Cuestionario, Comparación.

ABSTRACT: Usability is a quality attribute that is important in software development and that it is important that it be evaluated. Usability can be evaluated in several ways, questionnaires are the most popular way to evaluate usability, so we chose four of the usability evaluation questionnaires for the software, available in Spanish: QUIS, SUMI, CSUQ and SUS. They were applied in two different groups in which two different programs were evaluated. Based on the collected data, it was verified by Cronbach's alpha that the four present a level of reliability greater than 0.7 regardless of the number of questions.

KEYWORDS: Software engineering, Program quality, Usability, Cronbach's alpha, Questionnaire, Comparison.

INTRODUCCIÓN

Aunque la definición de usabilidad varía dependiendo del autor, esta es una característica que es cada vez más importante en la Ingeniería de Software, de acuerdo con Pressman [1], la usabilidad es un atributo de calidad y es "el esfuerzo que se requiere para aprender, operar, preparar las entradas e interpretar las salidas de un programa", mientras que Nielsen [2] menciona que la usabilidad es una serie de atributos que tienen los programas que proporcionan facilidad de aprendizaje, eficiencia, recuerdo en el tiempo, bajo nivel de errores y satisfacción, por otra parte, en la definición proporcionada por el ISO/IEC [3] se menciona que la usabilidad es una característica que permite a un producto ser usado con efectividad, eficiencia y satisfacción.

Independientemente de la definición de usabilidad, esta es un atributo intangible del software que puede ser evaluada principalmente por dos métodos diferentes: métodos heurísticos y métodos empíricos [4]. En el caso de los métodos heurísticos es necesaria la participación de expertos en usabilidad que participan desde el principio del diseño del software [2] mientras que en los empíricos (entrevistas, cuestionarios, grupos focales, entre otros) se tiene que tomar en cuenta la participación del usuario final para poder tener resultados aceptables [5]. Dentro de los métodos empíricos, los cuestionarios son una de las formas de evaluación de la usabilidad más comunes [6], esto debido a la rapidez con la que se obtienen los resultados y su facilidad de implementación.

Si bien organizaciones como la International Standard Organization [7] y la European Telecommunications Standards Institute [8] reconocen la importancia de la usabilidad en el desarrollo del software no proporcionan un cuestionario de evaluación, por lo que el desarrollador debe de elegir entre diversos cuestionarios realizados.

Algunos de los principales cuestionarios utilizados en la evaluación de la usabilidad son: QUIS (Questionnaire for User Interface Satisfaction), desarrollado por Ben Schneiderman en 1987 [9]; el SUMI (Software Usability Measurement Inventory), desarrollado por Jurek Kirakowski en 1993 [10]; el CSUQ (Computer System Usability Questionnaire) el cual era utilizado en IBM y que fue evaluado por James R Lewis en 1995 [11], traducido al español en 2015 [12]; el SUS (System Usability Scales) utilizado por DEC evaluado en 1991 por N.M Lucey en 1991 [10] traducido al español en 2016 [13]. Lo anterior en parte es mostrado en la Tabla 1. En el caso de QUIS y SUMI son cuestionarios de uso comercial, si bien su uso para fines académicos es posible, en caso de que se desee aplicarlos en un entorno comercial es necesario el realizar un pago para tener el derecho de uso, por ejemplo: para el caso de QUIS en una aplicación comercial el pago es de US\$750 [14], mientras que en el caso de SUMI este es a partir de €650 [15]. Mientras que en el caso de los cuestionarios CSUQ y SUS, estos cuentan con una traducción al español que fue validada principalmente en páginas web, aunque es muy común el hacer uso de los cuestionarios para la evaluación de la usabilidad en ambos ambientes, solo se cambia las partes donde se habla del software por web [16]. Los cuestionarios presentados en la Tabla 1 presentan un número de preguntas diferentes, por ejemplo: QUIS presenta 27, SUMI tiene 50, CSUQ con 16 y SUS que solo presenta 10. Si bien cada cuestionario presenta diferentes secciones que nos pueden servir para mejorar el nivel de usabilidad del software, su calificación final nos indicará el nivel de usabilidad global del software evaluado.

Tabla 1. Algunos de los cuestionarios con versión en español.

Cuestionario	Año de desarrollo	Última versión	Referencias
QUIS	1988	2016	[9, 17]
SUMI	1993	N/D	[10]
CSUQ	1995	2012	[11, 12, 18]
SUS	1991	2011	[10, 13, 19]

N/D: No disponible

Un cuestionario, de cualquier tipo, antes de su aplicación tuvo que haber sido validado, esto con el fin de saber si este es confiable. La confiabilidad de un cuestionario es la precisión con que este mide lo que en realidad se evalúa, en condiciones normales de aplicación en una población determinada [19]. Esta confiabilidad se valida

principalmente por medio del coeficiente alfa de Cronbach [20]. Si se tiene un cuestionario de menos de 20 preguntas una confiabilidad mayor o igual de 0.70 en el coeficiente alfa de Cronbach, se considera que es un valor aceptable, mientras que un valor cercano a 1 es mejor, esto es siguiendo el criterio que se muestra en la Tabla 2 [21]. Si se tiene un cuestionario de más de 20 preguntas, en caso de que se presente un valor mayor de 0.90 en el alfa de Cronbach, se considera que las preguntas presentan cierta redundancia o duplicidad, por lo que es recomendable el eliminar una de ellas [20].

Tabla 2. Confiabilidad de acuerdo al alfa de Cronbach para cuestionarios de menos de 20 preguntas

Nivel de Confiabilidad	Alfa de Cronbach
Excelente	0.9 - 1
Bueno	0.8 - 0.9
Aceptable	0.7 - 0.8
Cuestionable	0.6 - 0.7
Pobre	0.5 - 0.6
Inaceptable	< 0.5

En la Tabla 3 se muestra el valor del alfa de Cronbach en los 4 cuestionarios analizados que se obtuvieron en su diseño y validación.

Tabla 3. Valores del alfa de Cronbach en los cuestionarios elegidos

Cuestionario	Alfa de Cronbach	Ref.
QUIS	0.94	[17]
SUMI	0.92	[10]
CSUQ	0.97	[12]
SUS	0.92	[13]

Los 4 cuestionarios fueron seleccionados por que tienen en común que evalúan los siguientes atributos de la usabilidad [12,13, 15, 22]:

- Eficiencia.
- Efectividad.
- Facilidad de aprendizaje.
- Satisfacción.

Objetivo de este trabajo.

Por medio de este trabajo se estudian 4 cuestionarios para evaluar la usabilidad del software y que están disponibles en español para así poder obtener información que permitirá a los desarrolladores de software el determinar cuál cuestionario utilizar en su trabajo en base al tiempo de aplicación y la confiabilidad. Se evaluaron las

versiones en español de los cuestionarios QUIS, SUMI, CSUQ (versión original) y SUS (en su versión positiva).

MATERIAL Y MÉTODOS

Metodología

En la Figura 1 se muestra una representación de la metodología seguida en este trabajo.

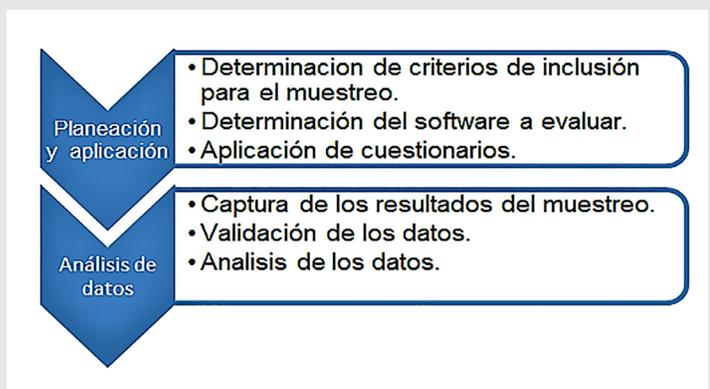


Figura 1. Metodología utilizada para realizar la comparación de cuestionarios.

La metodología fue dividida en dos etapas:

1. Planeación y aplicación.
2. Análisis de datos.

Planeación y aplicación.

En esta primera etapa se decidió que el criterio de inclusión de los participantes fuera el incluir en el muestreo a cualquier persona que manifestara el haber hecho uso, al menos una vez, del software a evaluar. Para el tipo de muestreo se decidió que, al no ser práctico realizar un muestreo probabilístico, se realizaría un muestreo por conveniencia [23] con un mínimo de 20 participantes, esto en base a que esta es la cantidad mínima que sugieren varios autores que se requiere en un estudio de usabilidad [2, 24]. Fueron aplicados 4 cuestionarios a la vez por cada participante, su orden de aplicación fue modificado en cada uno de ellos. Por último, se determinó que el software a evaluar fueran dos: Microsoft Excel 2010© y AutoCAD 2012©, esto debido a que son dos de los más populares y están disponibles en los centros de cómputo de dos instituciones de nivel superior en la localidad de Cd. Juárez, Chihuahua en donde se realizó el muestreo.

Análisis de datos

En esta etapa, la captura y validación de los resultados de los cuestionarios aplicados incluye la captura, depuración y validación de la base de datos. Los datos obtenidos de los cuestionarios se capturaron en el programa SPSS© v.24 para su análisis, en el caso de datos perdidos estos se sustituyeron por la mediana [25], gracias a que los datos en los cuatro cuestionarios se encuentran en una escala tipo Likert [26]. En el análisis de los datos se obtuvo la confiabilidad de los diferentes cuestionarios por medio del alfa de Cronbach. Los

resultados de los cuestionarios se convirtieron a una escala porcentual con fin de realizar una comparación de la calificación obtenida por cada uno de los cuestionarios los cuales presentan un número de escala en la respuesta diferente, QUIS presenta una escala de 10, SUMI de 3, CSUQ de 7 y SUS de 5. Normalmente se considera que a una mejor calificación global, será mayor el nivel de usabilidad.

Materiales

Los materiales utilizados para este trabajo son:

- Computadora con SPSS© v.24 y office© 2010 instalados.
- Impresora.
- Plumas.
- Copiadora.

RESULTADOS Y CONCLUSIONES

La descripción de la muestra recopilada es la siguiente: los cuestionarios se aplicaron en dos instituciones de nivel superior en la localidad de Cd. Juárez, Chihuahua, en su aplicación participaron 73 individuos divididos en dos grupos, 35 para la evaluación de AutoCAD y 38 en la evaluación de Excel; de los 78 participantes, 67.1% fueron hombres y 32.9% mujeres; el 11% de los entrevistados consideraban tener un nivel avanzado en su dominio del programa evaluado, 72.6% consideraban tener un nivel medio y 16.4% consideraban tener un nivel básico; el 20.5% de los usuarios expresó tener alguna experiencia con el uso de programas de similares características, mientras que 79.5% expresaron desconocerlos; el 95.9% de los entrevistados son estudiantes de nivel superior, mientras que el resto se clasificó en la categoría de otros; el 93,2% está en el rango de edad de 19-25 años, 2.8% está en el rango de 25-35 años y 1.4% está en el rango de más de 36 años.

La distribución de los datos son mostrados en la Figura 2, en donde por medio de los gráficos de caja se puede observar que los datos presentan una mayor agrupación en torno a la media para el caso de QUIS y CSUQ, mientras que en el caso de SUS y SUMI estos presentan una mayor dispersión.

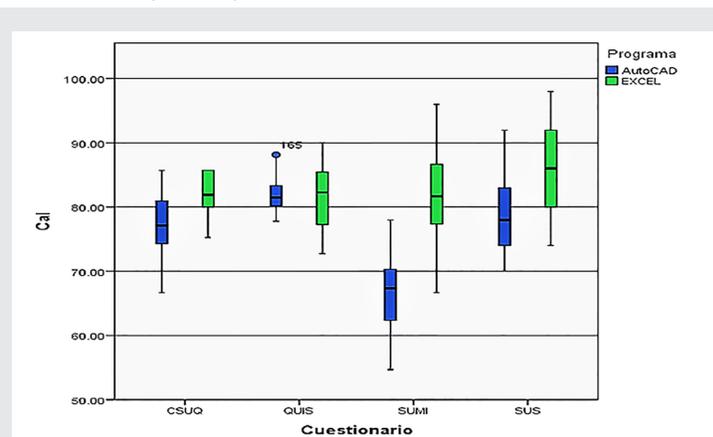


Figura 2. Gráficos de caja mostrando la dispersión de las medias de los resultados.

Como se puede observar en la Figura 3, en el caso de Excel, la calificación media en los 4 cuestionarios es muy parecida, en el caso de SUMI se presenta una mayor variación, mientras la calificación media obtenida para Excel es de 80.75 la media obtenida para AutoCAD es de 67.01, la más baja de los 4 cuestionarios, mientras SUMI, SUS y CSUQ muestran una menor diferencia en la calificaciones obtenidas entre Excel y AutoCAD.

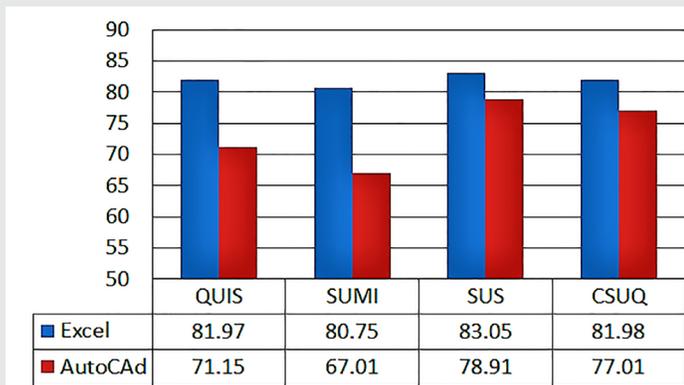


Figura 3. Calificación media para cada cuestionario.

En lo que respecta a la confiabilidad, esta se obtuvo por medio del alfa de Cronbach en cada cuestionario, en todos los casos se encontró que estos estaban en los niveles de aceptación (alfa de Cronbach mayor de 0.70), estos son mostrados en la Tabla 4.

Tabla 4. Confiabilidad de los cuestionarios evaluados

Cuestionario	Año de desarrollo	Última versión	Referencias
QUIS	1988	2016	[9, 17]
SUMI	1993	N/D	[10]
CSUQ	1995	2012	[11, 12, 18]
SUS	1991	2011	[10, 13, 19]

Podemos observar que en el caso de los 4 cuestionarios el alfa de Cronbach obtenido es menor al reportado por sus autores, excepto el valor del SUMI al evaluar Excel. Si bien SUMI presenta un valor de alfa de Cronbach de 0.924 se recomienda, en caso de que se desee aplicarlo, el tener en cuenta que muy posiblemente su diseño presente cierta redundancia.

Es interesante notar que uno de los cuestionarios más simples estudiados, el SUS (con sólo 10 preguntas), presenta un nivel de confiabilidad bueno, el cual está en el rango de cuestionarios de más preguntas.

Como recomendaciones de futuro, los autores de este artículo consideran que es importante realizar estudios tomando en cuenta otras muestras y evaluar programas diferentes a los aquí evaluados.

BIBLIOGRAFÍA

- [1] Pressman, R.S. (2010). *Ingeniería del software, Un enfoque práctico*, 7a ed., McGraw-Hill Interamericana, México D. F.
- [2] Nielsen, J. (1993). *Usability Engineering*, 1a ed., Morgan Kaufmann, San Diego.
- [3] ISO/IEC (2011). *Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models*. Int. Organ. Stand. Consultado el 12 de mayo de 2017, de la página electrónica: http://www.iso.org/iso/catalogue_detail.htm?csnumber=35733.
- [4] Mascheroni, M. & Greiner, C. (2012) *Calidad de software e ingeniería de usabilidad*. XIV Workshop de Investigadores en Ciencias de la Computación, 656-659.
- [5] Ferré, X., Juristo, N., Windl, H. & Constantine, L. (2001) *Usability basics for software developers* [versión electrónica]. *IEEE Software*, 18, 22-29.
- [6] Holzinger, A. (2005) *Usability engineering methods for software developers* [versión electrónica]. *Communications of the ACM*, 48, 71-74.
- [7] International Organization for Standardization (1998). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - part 11: guidance on usability*. 1998, 28. Consultado el 20 de abril de 2017, de la página electrónica: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=16883.
- [8] European Telecommunications Standards Institute (2000). *Human Factors (HF); Usability evaluation for the design of telecommunication systems, services and terminals*. Consultado el 26 de mayo de 2017, de la página electrónica: http://www.etsi.org/deliver/etsi_eg/201400_201499/201472/01.01.01_60/eg_201472v010101p.pdf.
- [9] Shneiderman, B. (1987). *Designing the user interface: strategies for effective human-computer-interaction*, 1a ed., Addison-Wesley Publishing Co., Reading, MA.
- [10] Kirakowski, J. (1999). *SUMI Background Reading. What is sumi?* Consultado el 31 de mayo de 2017, de la página electrónica: <http://sumi.uxp.ie/about/sumipapp.html>.
- [11] Lewis, J. (1995). *IBM Computer Usability Satisfaction Questionnaires: Psychometric Eva-*

- uation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7, 57-78.
- [12] Hedlefs Aguilar, M.I., la Garza González, A. de, Sánchez Miranda, M.P. & Garza Villegas, A.A. (2015) Adaptación al español del Cuestionario de Usabilidad de Sistemas Informáticos CSUQ. *Revista Iberoamericana de las Ciencias Computacionales e Informática* 4. Consultada el 26 de mayo de 2017, <http://www.reci.org.mx/index.php/reci/article/view/35>.
- [13] Hedlefs Aguilar, M.I. & Garza Villegas, A.A. (2016) Análisis comparativo de la Escala de Usabilidad del Sistema (EUS) en dos versiones. *Revista Iberoamericana de las Ciencias Computacionales e Informática*, 5. Consultada el 26 de mayo de 2017, <http://reci.org.mx/index.php/reci/article/view/48>.
- [14] QUIS Licensing. (N/d). Consultado el 1 de junio de 2017, de la página electrónica: <http://www.lap.umd.edu/QUIS/licensing.html>.
- [15] What is SUMI? (N/d). Consultado el 26 de mayo de 2017, de la página electrónica: <http://sumi.uxp.ie/about/whatis.html>.
- [16] Tullis, T.S. & Stetson, J.N. (2004). A Comparison of Questionnaires for Assessing Website Usability. *Usability Professional Association Conference*, 1-12.
- [17] Shneiderman, B., Jacobs, S., Cohen, M., Diakopoulos, N., Elmqvist, N. & Plaisant, C. (2016). *Designing the user interface : strategies for effective human-computer interaction*, 6a ed., Addison-Wesley, Reading, MA.
- [18] Sauro, J. & Lewis, J.R. (2012). *Quantifying the user experience : practical statistics for user research*, 1st ed., Morgan Kaufmann, Waltham.
- [19] Sauro, J. & Lewis, J.R. (2011). When designing usability questionnaires, does it hurt to be positive? *29th SIGCHI Conference on Human Factors in Computing System*, 2215-2223.
- [20] Celina, H. & Campo Arias, A (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*. 34(4), 572-580.
- [21] George, D. & Mallery, P. (2016). *IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference*, 14th ed., Routledge, New York.
- [22] Chin, J.P., Diehl, V.A. & Norman, K.L. (1988). *Development of an instrument measuring user satisfaction of the human-computer interface*. *ACM CHI'88 Proc*. 218, 213-218.
- [23] Polit, D.F. & Hungler, B.P. (2000). *Investigación Científica en Ciencias de la Salud*, 6a ed., McGraw-Hill Interamericana, México D. F.
- [24] Alroobaea, R. & Mayhew, P.J. (2014) How many participants are really enough for usability studies? *Proc. 2014 Sci. Inf. Conf. SAI 2014* 48-56.
- [25] Hair, J.F., Black, W.C., Babin, B.J. & Anderson, R.E. (2010) *Multivariate Data Analysis*, 7a ed., Prentice Hall.
- [26] Hernández, R., Fernández, C. & Baptista, M. del P. (2010) *Metodología de la Investigación*, 5a ed., McGraw-Hill Interamericana, México D. F.