Health Promotion Perspectives, 2014, 4(1), 82-89
doi: 10.5681/hpp.2014.011
http://journals.tbzmed.ac.ir/HPP

Psychometric Properties of the Persian Language Version of the System Usability Scale

Iman Dianat[1], Zahra Ghanbari[1], *Mohammad AsghariJafarabadi[2]

[1]Department of Ergonomics, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Road Traffic Injury Research Centre, Tabriz University of Medical Sciences, Tabriz, Iran

ARTICLE INFO

Article type:
Original Article

Article history:
Received: Mar 27 2014
Accepted: June 29 2014
e-published: July 12 2014

Keywords:
Ergonomics,
Validity,
Reliability,
System usability scale,
Iran

***Corresponding Author:**
Mohammad AsghariJafarabadi
Tel: +98 411 3357580;
e-mail:
m_asgharl862@yahoo.com

ABSTRACT

Background: The System Usability Scale (SUS) is a widely used self–administered instrument for evaluating the usability of a wide range of products and services. The aims of this descriptive-methodological study were to develop and investigate the psychometric properties of the Iranian version of the SUS.
Methods: The study was conducted among 202 university students from the Tabriz University of Medical Sciences, Iran. Content validity was evaluated by a panel of 10 experts. Construct validity was assessed by exploratory and confirmatory factor analyses. The internal consistency and test-retest reliability were assessed by Cronbach's alpha and Intraclass Correlation Coefficient (ICC), respectively. Additionally the feasibility of the measure was judged by ceiling and floor effect.
Results: Content validity of the short form of Iranian SUS was established. Factor analyses supported the conceptual uni-dimensionality, and thus confirmed the construct validity of the measure. The internal consistency (α = 0.79) and test retest reliability (ICC = 0.96) were both approved and there was also no ceiling nor floor effect.
Conclusions: The findings support the use of SUS for both practical and research settings in Iranian population.

Citation: Dianat I, Ghanbari Z, AsghariJafarabadi M. Psychometric Properties of the Persian Language Version of the System Usability Scale. Health Promot Perspect 2014; 4(1):82-89.

Introduction

The International Organization for Standardization (ISO) has defined the usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a in a specified context of use".[1] Effectiveness is the "extent to which a goal or task is achieved", efficiency is the "amount of effort required to accomplish a goal" and satisfaction is the "level of comfort that the users feel when using a product and how acceptable the product is to users as a means of achieving their goals".[2]

Although the usability has been established as an important consideration in product design, it is required to establish its importance when it comes to product choice.[3] Thus, usability studies are important for both companies and consumers. From a company point of view, it is essential to conduct market research to understand the exact consumer needs and therefore improve sales and market share. For consumers, a

poorly designed product may lead to stress, frustration and time wasting.[3–4] The results from usability evaluation studies can be used to evaluate the possible success of a product with its intended market, to compare two or more similar products/services and to provide feedback about the design.[5–6]

There are an increasing number of methods available for evaluating product designs for usability, which can be classified as either empirical or non-empirical methods. The empirical (also called user-based) methods involve observing users interacting with products or asking them to express their perceptions of a product's usability, while non-empirical (also called expert-based) methods involve an expert making a judgement about a product's usability or a structured check of a product's design qualities.[7–8] An empirical approach may be preferable as there is no substitute for involving the users when it comes to improving user-product interaction. However, in the case of constraints (e.g. user involvement difficulties, confidentiality issues, financial or time restrictions) a non-empirical approach may be preferable.[7]

The System Usability Scale (SUS) was developed by Brooke[9] as a survey scale for evaluating the usability of a given product or service. The SUS has several desirable features that make it an appropriate tool for general usability practitioners. The SUS, as a reliable, quick and easy method, provides a single score on a scale that is easily understood by the wide range of people.[10] This tool is also flexible enough to assess the usability of a wide range of products and services.[10] The SUS is composed of 10 statements related to various aspects of usability that are scored on a 5-point scale of strength of agreement. Final scores for the SUS range from 0 to 100, with higher scores indicating higher perceived usability.

As noted earlier, many techniques and methods have been employed in usability, most of which have been developed for use in English-speaking populations. The cross-cultural adaptation of existing English language questionnaires is therefore necessary for use in other populations. Thus, it was decided to conduct a study to validate the Persian version of SUS for people with Persian (Farsi) language in the world, although this has not been done in any other languages. The aim of the present study was to develop and evaluate the psychometric properties of a Persian version of SUS.

Materials and Methods

Study participants

In this descriptive- methodological study, a random sample of 202 students from the Tabriz University of Medical Sciences, Iran were recruited to express their experiences with the university food reservation system using the SUS questionnaire. Demographic details including age, gender, and educational level of each participating student were also recorded. Those students who had used this system at least once a week for the duration of one year were included in the study. The study period was between September and November 2013.

Instrument (SUS)

The SUS was developed as part of the usability engineering program in integrated office systems developed at the Digital Equipment Company.[9] The SUS comprises 10 items that relate to the usability of the system or product. When it comes to selecting a usability questionnaire, usability practitioners may have several choices (for example, Software Usability Measurement – SUMI[11], Usefulness, Satisfaction and Ease of Use –USE[12], Poststudy System Usability – PSSUQ[13]). The distinct advantage of SUS is the ease with which the measure may be applied. The SUS has proved its value over the years as pointed out by Brooke[14] ". . . it is now possible to choose SUS and be confident it is a valid and reliable measuring tool, be able to make a comparison between the scores you achieve and some normative standards (and thus with other systems or products), and to have some idea . . . of . . . system . . . You can do this quickly, and you can be fairly sure that you are getting reliable results by asking a small number of users

and even, maybe, basing it on their first impressions of the system." It takes about one minute to complete the questionnaire, and no training is required.

The respondents to the SUS are asked to rate the usability of a given product/system using items on a 5-point self-report scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). There are five positive statements (items with odd numbers) and five negative statements (items with even numbers), which alternate. The overall SUS score is calculated by taking 1 from all the scores on items with odd numbers and subtracting the scores from 5 for all the items with even numbers. The sum of the scores for ten items is then multiplied by 2.5 to give an overall SUS score between 0 (extremely poor usability) and 100 (excellent usability).

The Persian conversion of the SUS was conducted in a forward–backward translation process. The forward translation was carried out by a specialist in ergonomics. Afterwards, the back translation to the original English of the Persian version was performed by two professional translators with experience of living in English-speaking countries. Some modifications were then made on the Persian version after comparing to English back–translation with original edition. For possible linguistic problems, the questionnaire was completed by 30 participants and appraised and some amendments were made following feedback from participants.

In qualitative evaluation, the SUS was reviewed for content validity by a panel of 10 experts and few items were revised based on the feedback from expert panel. In quantitative evaluation, a survey involving questions in two general parts was administered to the expert panel members. The experts provided their answers to these questions based on a 4-point scale. The first part included questions regarding relevancy, clarity and simplicity, which were used to compute Content Validity Index (CVI), and the next part contained a question regarding the necessity of each item, that was used for calculation of Content Validity Ratio (CVR). Based on the number of expert panel members, CVI and CVR values higher than 0.75 and 0.62, respectively, were considered appropriate.[15] It should be noted that 4 out of 10 items were revised based on the quantitative results and qualitative recommendations in the first step evaluation phase. For the quantitative evaluation of the content validity, CVI and CVR showed satisfactory results for each item (both CVI and CVR ranged between 0.80 and 1.00), and consequently for the SUS (Table 1).

Some authors have suggested that 5 or more participants per item or a total sample size of 200 is sufficient for factor analysis.[16–17] Thus, a sample of 202 participants was recruited in this study, which indicates a good sample size.

Table 1: The scores of Relevancy, Clarity, Simplicity, CVI and CVR for SUS

Item	Item content	Relativity	Clarity	Simplicity	CVI	CVR
1	I think that I would like to use this system frequently.	1.00	0.70	0.80	0.83	1.00
2	I found the system unnecessarily complex.	0.90	0.80	0.90	0.87	1.00
3	I thought the system was easy to use.	1.00	1.00	1.00	1.00	1.00
4	I think that I would need support of a technical person to be able to use this system.	1.00	0.90	0.90	0.93	1.00
5	I found that the various functions in this system were well integrated.	0.90	0.70	0.80	0.80	1.00
6	I thought that there was too much inconsistency in this system.	1.00	0.80	0.80	0.87	1.00
7	I would imaging the most people learn to use this system very quickly	0.80	0.90	1.00	0.90	0.80
8	I found the system very cumbersome to use.	0.90	0.90	0.80	0.87	0.80
9	I felt very confident using the system.	0.80	0.90	1.00	0.90	0.80
10	I needed to learn a lot of things before I could get going with this system.	1.00	0.90	0.90	0.93	1.00
Total		0.85	0.77	0.81	0.81	0.85

Data Analysis

The data were presented as Mean (SD) for the SUS scale and study quantitative variables and N (%) for qualitative variables. Ceiling and floor effects were appraised using percentage of scores at the boundaries of the scaling range (e.g. 0 and 100). Floor or ceiling effects are matter of concern if more than 15% of respondents achieve the lowest or highest possible score, respectively.[18] Cronbach's alpha and Intraclass Correlation Coefficient (ICC) were computed for evaluating scale internal consistency and stability reliability, respectively. Values higher than 0.7, was considered as satisfactory.[19]

Exploratory Factor Analysis (EFA) was used for structure detection. The purpose of structure detection is to examine the underlying (or latent) relationships between the variables. EFA was carried out by Principal Axis Factoring (PAF) extraction method and utilizing Varimax Rotation with Kaiser Normalization. The Scree plot procedure was used for deciding on the number of factors to be extracted.[19] KMO (Kaiser-Meyer-Olkin) measure of sampling adequacy, Bartlett's test of sphericity and total variance explained were used for the evaluation of model sufficiency. High values of KMO (more than 0.7) generally indicate that a factor analysis may be useful with the data. Bartlett's test of sphericity tests the hypothesis that a correlation matrix is an identity matrix, which would indicates that variables are unrelated and therefore unsuitable for structure detection. Small values (less than 0.05) of the significance probability indicate a satisfactory factor analysis. Factor loading values of 0.3 or higher were considered as an important relationship between items and factors.[20] Confirmatory Factor Analysis (CFA) was conducted to assess how well the EFA extracted model fits to observed data. The weighted least squares estimation method was used with weighted matrix of asymptomatic covariances. Fit indices and reasonable values of these indices were considered as χ^2 / df < 5, Root Mean Square Error of Approximation (RMSEA) < 0.08 and also, Comparative Fit Index (CFI) > 0.9. Statistical analysis was performed using SPSS 13.0 (SPSS Inc., Chicago, IL) and LISREL 8.80 (Scientific Software International Inc., 2007). *P*-values less than 0.05 were considered as significant.

Procedure and Ethical Considerations

Permission to use the original SUS was gained by the original author. The Ethical Committee of the Tabriz University of Medical Sciences reviewed and approved the study. The participants were informed about the general nature of the study and were assured of the privacy of their records. The written informed consent for the study was obtained from all the participants.

Results

Sample characteristics

There was a total of 210 university students, from which 202 students (124 males and 78 females) returned the completed questionnaires (response rate = 96.19%). The age of participants ranged from 19 years to 29 years (mean = 22.1 years; SD = 2.3 years). Among them, 71.8% (n = 145) were undergraduate and 28.2% (n = 57) were postgraduate students.

Feasibility

The results showed no Ceiling effect or Floor effect for the Farsi version of SUS.

Content Validity

In qualitative evaluation of the measure, experts gave written feedback on the clarity and relevancy of the content of the SUS items to the Iranian culture. In general, the content validity of the measure was supported in this stage. It is also notable that some of the items were improved based on the qualitative suggestions of the expert panel.

Reliability

The Cronbach's alpha was 0.79 for the scale, which indicates adequate internal consistency (more than 0.7). ICC for the scale (value = 0.96; 95% CI = 0.91 to 0.99) also

showed a satisfactory (more than 0.7) test retest reliability.

Construct Validity

For evaluating construct validity, the results of both EFA and CFA were presented for the items of the SUS.

EFA

In this analysis, KMO (Kaiser–Meyer–Olkin) measures of sampling accuracy were 0.81, which established the model sufficiency for these subscales. Bartlett's test of sphericity gave $P < 0.001$, which was in line with KMOs.[21] Total variance explained of the scale was 36.85%. The Scree plot supported the uni-dimensionality of the SUS (Fig. 1).

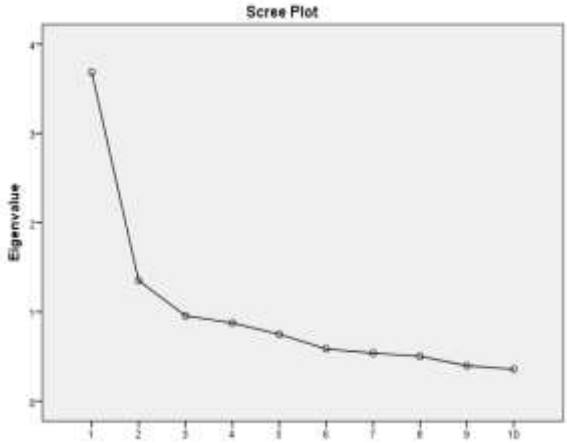

Fig. 1: Scree Plot of the SUS

A cut off value more than 0.3 for factor loadings proposed that all items associated to the SUS have been convincingly loaded on the scale (Table 2).

CFA

The CFA model showed sensibly good fit indices (Chi2 / df = 1.04 < 5; SRMR = 0.051 < 0.1, RMSEA = 0.014 < 0.08, CFI = 0.998 > 0.90, NFI = 0.945 > 0.90, NNFI = 0.996 > 0.90, GFI = 0.973 > 0.90, AGFI = 0.948 > 0.90). Therefore, the model fit was established logically by the indices and there was a good support for the one-factor structure of the SUS.[20, 22–23]

Table 2: Exploratory Factor loadings for SUS

Items	SUS
Q8	0.690
Q3	0.663
Q7	0.640
Q2	0.632
Q4	0.615
Q9	0.554
Q10	0.466
Q1	0.404
Q5	0.388
Q6	0.314

In the next step the relationships were evaluated based on this model. The values of factor loadings specified that the items had significant loadings on the one-factor solution (Fig. 2). Standardized factor loadings ranged from 0.30 to 0.71 on the scale, with all items representing moderate to strong factor loadings[24] (Fig. 2).

The statistical significance of the two-factor correlations ($r = 0.70$, $P < 0.01$) supports the hypothesis that the two factors (OE and EE subscales) are highly related dimensions of self-efficacy for childbirth.

Discussion

The present study has resulted in the development and validation of the Persian version of the SUS for usability studies in Iran and other Persian-speaking communities. Usability studies, which are very important for both the designers of the products and consumers, have been very limited in Iran. This may be attributed mainly to the lack of valid and reliable tools in this regard. Therefore, Persian-speaking communities have a great need for valid and reliable tools and instruments to measure users' perception of the usability of a wide range of products and services. The results of this study showed that the Persian SUS is a valid and reliable tool for measuring usability, with psychometric properties consistent with the original English version.[9]

Feasibility

No floor or ceiling effects were observed for the total score of Persian SUS.

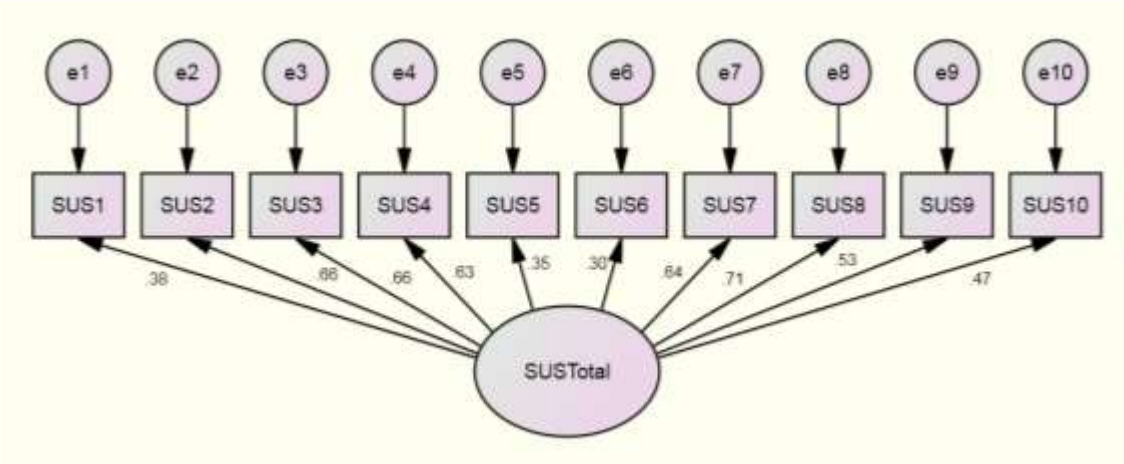

Fig. 2: Path diagram revealing the standardized parameters relating items to SUS. All parameters were statistically significant (All $P < 0.05$)

When more than 15% of the study participants produce the maximum or minimum possible score on a scale, it represents a ceiling or floor effect, respectively, which can be considered as a measuring limitation of the instrument.[18] The lack of floor and ceiling effects reassures the usability practitioners of the validity of this version of the SUS. No other study reported the measure of ceiling and floor effects of this instrument, and therefore comparison may not be established.

Content validity

Content validity of the SUS was approved based on both qualitative (e.g. comments from the panel members) and quantitative results (e.g. the level of agreement among expert panel members, with CVR and CVI values higher than 0.85 and 0.81, respectively). The original instrument did not use CVR and CVI measures for the essentiality of the items and for simplicity, relativity, and clarity of the subscales, respectively.[9] No other study also reported these measures in their evaluation of the SUS in which to compare our results.

Reliability

In reliability analyses, the absolute ratings (e.g. transformed responses for Statements 2, 4, 6, 8, and 10 so that all scales had 1 as the negative and 5 as the positive) for the 10 statements were used to calculate Cronbach's alpha. The SUS showed good internal consistency (alpha = 0.79). The Cronbach's alpha found in our study is in agreement with those reported previously in English language studies, which ranged from 0.85 and 0.92.[10,25–27] In addition, the value for ICC (0.96) to assess test-retest reliability indicated excellent reproducibility. The high ICC for the Persian SUS demonstrates a strong stability of the SUS over time. Neither the original study,[9] nor other studies have evaluated this measure for the SUS.

Factorial (construct) validity

A factor analysis was conducted to determine whether or not the 10 items on the scale address just one dimension (e.g. usability), as expected.[9] The results indicated that there was only one significant factor for the ten SUS statements and supported the one-factor structure of the SUS.[20,22-23] This result is in line with a previous report by Bangor et al.,[10] who demonstrated that the SUS questionnaire, as a whole, reflects participants' perception of the overall usability of a product or system and, therefore this confirms that practitioners should analyze and report only the overall SUS scores.[9]

Limitations and future studies

The measure was used among a sample of students from the Tabriz University of Medical Sciences to express their experiences with the university food reservation sys-

tem, and therefore further studies testing other systems and populations may be required to approve the applicability of the SUS as a totally accepted practical measure in Iranian population. The factor analyses were conducted on the same sample, and thus it is also recommended to run the CFA on a separate sample.

Conclusions

The present study aimed to validate and culturally adapt the System Usability Scale (SUS) for the Iranian population. The results of the current study demonstrated high degrees of feasibility, reliability and validity for the Persian version of the SUS as a measure of usability assessment. The psychometric properties of the Persian version of the SUS are comparable with those of the original English version, and therefore Iranian usability practitioners and researchers may use it for usability evaluations and for research purposes.

Acknowledgements

The authors would like to acknowledge all participants who collaborated in this study.

Competing interests

The authors declare that there is no conflict of interest.

Funding

There was no funding for this study.

References

1. ISO WD. 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). USA: International Organization for Standardization;1998.
2. Jordan PW. An introduction to usability. London: Taylor & Francis;1998.
3. Mack Z, Sharples S. The importance of usability in product choice: A mobile phone case study. *Ergonomics* 2009;52:1514–1528.
4. Zhang Y, Luximon A. Usability of mobile phone. In: Karwowski W, editor. International encyclopaedia of ergonomics and human factors. Boca Raton: Taylor & Francis;2006.
5. Baber C. Evaluation in human–computer interaction. In: Wilson JR, Corlett N, editors. Evaluation of Human Work. Boca Raton: Taylor & Francis;2005.
6. Harvey C, Stanton NA, Pickering CA, McDonald M, Zheng P. A usability evaluation toolkit for In-Vehicle Information Systems (IVISs). *Appl Ergon* 2011;42:563–574.
7. Jordan PW. Usability and product design. In: Karwowski W, editor. International encyclopaedia of ergonomics and human factors. Boca Raton: Taylor & Francis;2001.
8. Dillon A. Evaluation of software usability. In: Karwowski W, editor. International encyclopaedia of ergonomics and human factors. Boca Raton: Taylor & Francis;2006.
9. Brooke J. SUS: A "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability evaluation in industry. London: Taylor and Francis;1996.
10. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum Comput Stud* 2008;24:574–594.
11. Kirakowski J, Corbett M. SUMI: The Software Measurement Inventory. *Br J Educ Technol* 1993;24:210–212.
12. Lund AM. Measuring usability with the USE Questionnaire. Usability Interface: The usability SIG newsletter of the Society for Technical Communications 2001;8(2). Available at: http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html
13. Lewis J. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int J Hum Comput Stud* 2002;14:463–488.
14. Brooke J. SUS: A Retrospective. *Journal of Usability Studies* 2013;8:29–40.
15. Waltz CF, Strickland OL, Lenz ER. Measurement in nursing research. 2nd ed. Philadelphia: FA Davis Company;1991.
16. Bryant FB, Yarnold PR. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold RR, editors. Reading and understanding multivariate statistics. Washington, DC: American Psychological Association;1995.

17. DeVellis RF. Scale development: theory and application. 2nd ed. Thousand Oaks, CA: Sage Publications;2003.
18. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
19. Tinsley HEA, Brown SD. Handbook of applied multivariate statistics and mathematical modeling. New York: Academic Press;2000.
20. Kline RB. Principles and practice of structural equation modeling. 2nd ed. New York: Guilford;2005.
21. George D, Mallery P. SPSS for Windows step by step: a simple guide and reference. 6th ed. Boston: Allyn & Bacon;2003.
22. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods* 1996;1:130–149.
23. Marsh HW, Balla JR, Hau K. An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In Marcoulides GA, Schumacker RE, editors. Advanced structural equation modelling: Issues and techniques. Mahwah, NJ: Erlbaum;1996.
24. Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale, NJ: Erlbaum;1988.
25. Kirakowski J. The use of questionnaire methods for usability assessment. Unpublished manuscript[internet]. Available at: http://sumi.ucc.ie/sumipapp.html
26. Lewis JR, Sauro J. The factor structure of the system usability scale. In: Masaaki K, editor. Human Centered Design. Berlin: Springer Berlin Heidelberg;2009.
27. Kobak KA, Craske MG, Rose RD, Wolitsky-Taylor K. Web-based therapist training on cognitive behavior therapy for anxiety disorders: A pilot study. *Psychotherapy* 2013;50:235–247.