Tribhuvan University Journal
Vol. 38, No. 2: 83-92, December 2023
Research Directorate, Tribhuvan University (TU),
Kathmandu, Nepal
DOI: https://doi.org/10.3126/tuj.v38i2.60770

CONSTRUCTION AND VALIDATION OF ALTERNATIVE USABILITY SCALE

Pralhad Adhikari

Tri-Chandra Multipple Campus, TU, Kathmandu
Corresponding author: Pralhad.adhikari@trc.tu.edu.np

Received date: 1 Feb. 2023- Accepted date: 25 Nov. 2023

ABSTRACT

Usability is the extent of a system being user-friendly. The aim of this research was to make a test to measure usability based on five-factor model of usability by Nielsen (1993). Total of 755 participants (M_{age}=26.1, SD=5.86) representing the users of 21 types of products or services gave responses. Psychometric analyses were done. The test with fifteen items did not fit the model of its parent theory as revealed by confirmatory factor analysis. Principal component analysis was conducted and some items were deleted. Finally, a test of 12 items with two subscales was made. The participants were given a survey in English but they were not its native speakers. The future research can take native speakers and more diverse products' or services' users for psychometric analyses of the 15-item version.

Keywords*:* learnability, memorability, errors, satisfaction, efficiency, system design

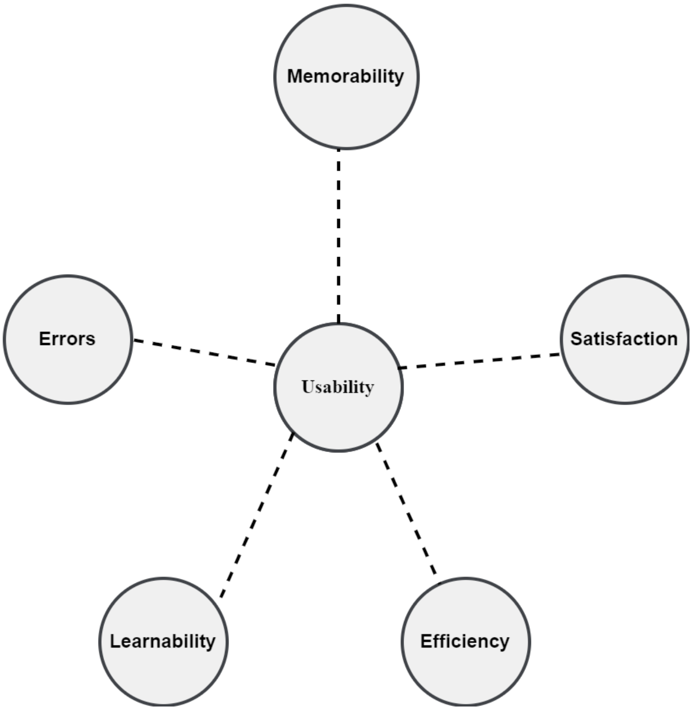
INTRODUCTION

Usability is the extent of a system being user-friendly. It denotes how easy or intuitive a service or product is to use (Wickens *et al.,* 2014). Usability has five components: learnability, efficiency, memorability, errors, and satisfaction (Lee *et al.,* 2017, p. 58). This five-factor model (Nielsen 1993) is used as the theoretical framework for this study. Learnability refers to the extent of system being easy to learn so that user can quickly start to use the system. Efficiency is the degree to which system is productive. Memorability is the extent to which a user can return to the system after some period and use it easily. Errors is the number of mistakes during the use of system. Satisfaction is the degree of liking during use of a system.

Usability is concerned with effectiveness, and achievement of goals also. Other authors such as Jordan (2002) consider the following as components of usability: guessability, learnability, performance, system potential, and re-usability. Guessability is the cost of using a product to perform a new task for the first time. It is similar to learnability. System potential is the maximum amount of performance possible with a product. It is similar to efficiency. Re-usability is the other name of memorability. The other terms associated with usability are effectiveness, cognitive load, simplicity and ease of use (Weichbroth, 2020).

Figure 1

Five-factor Model of Usability by Nielsen (1993): It is the Theoretical Framework for this Study.



Usability testing is a formative evaluation technique and related to iterative development process (Lee *et al.,* 2017, p. 58). So, it can be used to identify problems with design in a system and is helpful for its improvement. Improvement of product or service is a continuous process. A way to do usability testing is to let the users of system self-report their experiences with system. At this point arises a need to develop a test to assess usability. The usability testing can be done with five users in each dimension of usability. During the system design phase, multiple parallel usability testing is better than a single usability testing (Lee *et al.,* 2017, p. 58). It is helpful to find areas for improvement in a system. Design team incorporates the findings and conclusions from each phase of usability testing. Usability testing once is not enough. Ideally, five rounds of usability testing with each system design are needed, but doing more than five times is more beneficial. Users are directly involved with the performance of the system, and should be involved in usability testing. This study aims to develop a reliable and valid test to assess usability in the backdrop of most popularly used scale named the System Usability Scale (SUS, Brooke, 1996) being very old and probably outdated.

In psychometrics, it is a common practice to revise psychological tests occasionally to keep them current, fair, reliable and valid (Cronje *et al.* 2022). However, such task can be formidable (Butcher, 2000). The older tests may get dated with the change in behaviors of people or may need revision to be still useful for the public. Since the SUS is 27 years old already, seeking its alternative is necessary. Hence, this study is justifiable.

METHODS INSTRUMENTS AND PROCEDURE

Google Forms was used to create an online survey. Based on Nielsen (1993), a scale with 15 items was created with the help of Wickens *et al.*, (2014) and included in a questionnaire with the System Usability Scale (Brooke 1996), some demographic questions, and three open-ended questions about points of satisfaction about the system, weaknesses of the system, and ways to better the system. There was an additional question to inquire the purpose to use the system. This approach to test construction is the rational approach in which there is a good link between content of scale and definitions of parent construct (Ruscio, 2015). This approach is deductive (Burisch, 1984). The system usability scale (SUS) is standardized popular measure to assess perceived usability and is quick to use (Lewis, 2018). This scale was used to test the validity of scale being developed in this study. The new scale was named the Usability Scale for System Design (USSD) and users could rate each item in 5-point Likert scale ("strongly disagree" through "strongly agree").

Participants

Total of 755 participants (M_{age}=26.1, SD=5.86) were surveyed. There were 445 females (M_{age}=25.2, SD=4.90), 306 males (M_{age}=27.4, SD=6.82), and four participants preferred not to reveal their gender. The users or customers of 21 types of products or service gave responses as shown in the table 1 below. Users aged 12 through 64 were included. The participants were found using snowballing and convenience sampling techniques. The sample consisted of educated individuals who could respond to the stimuli using Google Forms. The participants were mostly from Kathmandu valley. However, some participants were from various parts of the country and some were the Nepalese living abroad.

Table 1

Products the Participants were Asked Response for

Product	N	Percent
TikTok	65	8.6
Viber	63	8.3
Laptop	60	7.9
Smart phone	60	7.9
YouTube	60	7.9
Instagram	53	7.0
Teams	51	6.8
Zoom	36	4.8
Smart watch	34	4.5
e-Sewa	32	4.2
Facebook	30	4.0
LinkedIn	30	4.0
Pathao	30	4.0
Skype	30	4.0
WhatsApp	30	4.0
Gmail	29	3.8
Daraz	28	3.7
NetFlix	22	2.9
Others (Pumori, Messenger, Google Meet)	7	1.6

DATA ANALYSIS

Excel was used to organize data and Jamovi 2.3.21 was used to process the data. USSD was based on 5-factor model of Nielsen (1993). So, confirmatory factor analysis (CFA) was done. The model did not find a significant fit. Hence, principal component analysis (PCA) was conducted. PCA is a technique for exploratory factor analysis (EFA) when we aim to summarize most of the variance of variables in small number of factors

(Zhang & Luo, 2019). PCA is preferred when the main purpose is data reduction.

RESULTS

Reliability: The internal consistency of USSD as revealed by Cronbach's α was .77. This is acceptable reliability. McDonald's ϖ was .80 and it indicates good internal reliability.

Validity: The total scores of USSD and SUS correlated significantly, r=.71, p<.01. This establishes the concurrent validity. This type of validity is established when a test correlates with other tests that measure the same construct (Best & Kahn, 2014, p. 295). Qualitative data (in response to three open-ended questions) analysis warrants a separate paper but 643 persons gave some reasons for satisfaction with system they responded for. Among all, 645 respondents pointed out some weaknesses, and 635 respondents offered some suggestions to better the system. These facts can implicitly indicate the validity of the scale. Among all, 693 participants reported some purpose of using the system they responded for.

CFA: The CFA model did not show a significant fit, χ^2 (80) =778, p<.001. The fit indices did not meet criteria for good fit as shown in table below:

Table 2

Fit Measures

CFI	TLI	RMSEA
0.771	0.700	0.107

Note. CFI= comparative fit index, TLI= Tucker–Lewis index, RMSEA= root mean square error of approximation

The good fit in CFA needs to meet the following criteria (Adhikari, 2020, p. 55): p (for χ^2) >.05, CFI>.9, TLI>.9, and RMSEA<.08. None of the criteria have been met in the model of this study. The χ^2 and RMSEA are used for absolute fit and the TLI and CFI are used for incremental fit (Ahmad *et al.,* 2016).

PCA: Bartlett's test of sphericity revealed that data in this study are suitable for PCA, χ^2 (105) =3131, p<.001. KMO measure of sampling adequacy was .86 and indicated that sample was adequate. There were three factors as shown in table 2 below, and no correlation was seen between them.

Table 3

Factor Loadings

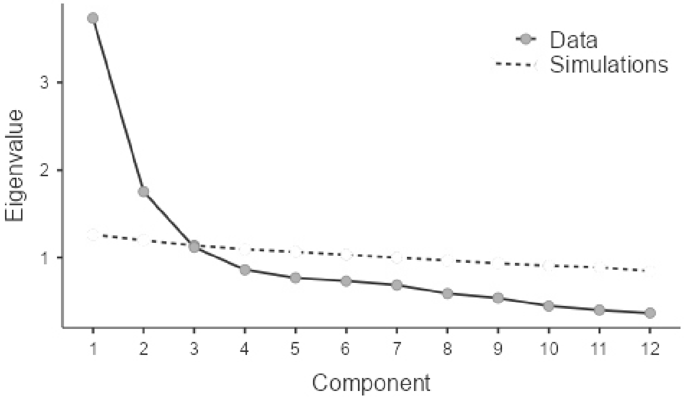
Dimension related to theory		Component			
	Item	**1**	**2**	**3**	**Uniqueness**
Efficiency	USSD4	0.771			0.382
Learnability	USSD3	0.758			0.411
Memorability	USSD7	0.757			0.424
Learnability	USSD1	0.732			0.414
Satisfaction	USSD13	0.704			0.484
Satisfaction	USSD15	0.695			0.452
Memorability	USSD9'	0.39	0.736		0.305
Memorability	USSD8'		0.718		0.415
Errors	USSD11'		0.639		0.568
Errors	USSD10		-0.536		0.659
Learnability	USSD2'		0.483		0.724
Efficiency	USSD5'			0.694	0.478
Efficiency	USSD6	0.386	-0.302	0.614	0.382
Errors	USSD12'		0.43	0.552	0.508
Satisfaction	USSD14'	0.303	0.407	0.456	0.535

Note. Varimax rotation was used. Loadings below .3 have been repressed. ' means reverse-scored item. USSD=Usability Scale for System Design

Removing the three items (i.e., 9, 12, 14) with conflicting components (or cross-loading), two factors only remained as shown by the scree plot given below in figure 1.

Figure 2

Scree Plot Made After Removing Three Items



The summary of the scores is given in the table 4 below. These statistics can be used as norms for future studies.

Table 4

Usability Score Summary

Test Version	M	SD	Minimum	Maximum	Q_1	M_d	Q_3
15-item	52.8	6.86	24	75	48	53	57
12-item	42.4	5.43	21	60	39	42	46

DISCUSSION

A reliable and valid test was made. It can be used as an alternative to the SUS or other tests available (such as Baumgartner *et al.* 2019, Borsci *et al.* 2022, Brooke 1996). The final test had the following 12 items as shown in table 4.

Table 5

Final Items Remaining in the Scale

Sub scale	Items	Loading
Factor 1 (labelled as short/direct)	4 This system is efficient to use.	0.785
	7 The steps needed to operate this system are easy to remember.	0.744
	3 This system has understandable display and instructions.	0.738
	15 I like this system.	0.725
	13 This system is pleasant to use.	0.716
	1 This system was easy to learn for me.	0.702
Factor 2 (labelled as long/indirect)	6 I can finish my task on time because of this system.	0.702
	*11 When I make errors in this system, I cannot easily recover from them.	0.678
	*8 I have to learn all steps over when I use this system after a gap of some time.	-0.629
	10 This system induces few errors.	0.555
	*2 I could not rapidly start getting some work done in this system.	0.384
	*5 This system has not increased my productivity.	0.702
Deleted items	*9 I have confused steps in this system because they are complex.	
	*12 I am afraid this system will cause accident.	
	*14 I am not satisfied with this system.	

*Note. * reverse-scored. Jamovi suggested to reverse item 10.*

The confirmatory factor analysis did not support the original theory of the scale. So, exploratory factor analysis was carried out and the PCA gave two dimensions of the scale. Summary in Table 4 can be used as norms for future studies.

There are several challenges to develop measures of usability like their validity (do they actually measure usability (Hornbæk, 2006)? Usability can be measured in three ways: user-oriented way, product-oriented way and user-performance way (Bevana *et al.,* 1991). The three kinds can be supported by contextual orientation too. The difficulty of measuring usability in systems is the complexity of the system. For example, Facebook is more than just a single product now. It has other features like Watch, Marketplace, and Games in addition to its Home with posts of friends and followed Pages. Many people do not know all its features. Since all users cannot know all the features of a product or service, they report the usability of the features they use. So, this scale also measures the perceived usability.

CONCLUSION

There were some limitations in this study. The sample used in it had English as the second language. So, the future studies can verify the structure of scale (with all 15 items) among native speakers and on the same system. Participants may not understand some items when they are non-native speakers (Finstad, 2006). In this study, a system (e.g., Viber) was inquired generically. In other words, the study was not focused on specific segment like mobile users, tablet users, computer users, android users or iOS users. Future studies can keep the mode of the use consistent. Moreover, most of the systems used by participants in this study were software. So, the scale with 15 items can be tested for systems other than computer or mobile applications in the following studies. The other types of reliability and validity need to be tested in future.

REFERENCES

Adhikari, Y. R. (2020). *Prevalence of Professional Quality of Life (ProQOL) and Its Influence on the Personal Distress of Doctors in Nepal*. University of Nicosia.

Ahmad, S., Zulkurnain, N. N. A. & Khairushalimi, F. I. (2016). Assessing the validity and reliability of a measurement model in Structural Equation Modeling (SEM). *British Journal of Mathematics & Computer Science*, *15*(3), 1-8.

Baumgartner, J., Frei, N., Kleinke, M., Sauer, J. & Sonderegger, A. (2019). Pictorial system usability scale (P-SUS): Developing an instrument for measuring perceived usability. *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, 1-11.

Best, J. W. & Kahn, J. V. (2014). *Research in Education* (10th ed.). Pearson Education.

Bevana, N., Kirakowskib, J., & Maissela, J. (1991). What is usability. *Proceedings of the 4th International Conference on HCI*, 1-6.

Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The Chatbot usability scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing*, *26*(1), 95–119.

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*. Taylor & Francis, pp. 189-194.

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*. https://doi.org/10.1037/0003-066X.39.3.214

Butcher, J. N. (2000). Revising psychological tests: Lessons learned from the revision of the MMPI. *Psychological Assessment*, *12*(3), 263-271. https://doi.org/10.1037/1040-3590.12.3.263

Cronje, J. H., Watson, M. B., & Stroud, L.-A. (2022). Guidelines for the revision and use of revised psychological tests: A systematic review study. *Europe's Journal of Psychology*, *18*(3), 293-301. https://doi.org/10.5964/ejop.2901

Finstad, K. (2006). The system usability scale and non-native English speakers. *Journal of Usability Studies*, *1*(4), 185–188.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*(2), 79–102.

Jordan, P. W. (2002). *An introduction to usability*. Taylor & Francis.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). *Designing for people: An introduction to human factors engineering* (3rd ed.). CreateSpace.

Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction*, *34*(7), 577-590. https://doi.org/10.1080/10447318.2018.1455307

Nielsen, J. (1993). *Usability Engineering*. Academic Press.

Ruscio, J. (2015). Rational/theoretical approach to test construction. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology* (pp. 1–5). Wiley Online Library. https://doi.org/10.1002/9781118625392.wbecp454

Weichbroth, P. (2020). Usability of mobile applications: A systematic literature study. *Ieee Access*, *8*, 55563–55577.

Wickens, C. D., Lee, J. D., Liu, Y. & Gordon-Becker, S. (2014). *An Introduction to Human Factors Engineering* (2nd ed.). Pearson.

Zhang, L. & Luo, W. (2019). Application of exploratory factor analysis in language assessment. In V. Aryadoust & M. Raquel (Eds.), *Quantitative Data Analysis for Language Assessment :Volume I*. Routledge.