❒   398

Web usability test in 60 seconds: a theoretical foundation and empirical test

Imran Mahmud[1], Farzana Sadia[2], Mostafijur Rahman[3], Sharmin Ahmed[4], Didarul Islam[5]
[1]Graduate School of Business,UniversitiSains Malaysia,Penang , Malaysia
[1,2,3,5]Department of Software Engineering,Daffodil International Univrsity, Bangladesh
[4]Department of Electrical and Electronic Engineering, Daffodil International University, Dhaka, Bangladesh

Article Info	ABSTRACT
	The aim of this research is to explore the possibility of web usability test in 60 seconds rather than unlimited time by customers. Usability was tested by two major testing methods: system usability scale (SUS) and NetQu@l.60 customers as two groups were involved in the experimental design procedure. The assessment included an online shopping website where one group tested in 60 seconds and other group had 3 days to test. Result shows that there are significance differences in SUS based testing and no significant differences in NetQu@l based testing. Altogether, these results provide further support that SUS based usability testing can be implemented in 60 seconds time frame without imposing additional cognitive load on customers.

Corresponding Author:

Imran Mahmud,
Graduate School of Business,
UniversitiSains Malaysia,
Penang , Malaysia.
Email: imranmahmud@usm.my

1.   INTRODUCTION

There are several researches have been conducted on web based services within last decades. Researchers induced several usability techniques like System usability scale (SUS) [1], Software usability measurement inventory (SUMI) [2], Questionnaire for user interface satisfaction(QUIS) [3] to evaluate websites. Customer acceptance testing is very crucial for voluntary technologies like web based market place, online shopping, web based games and services. The budget and time associated with web based system testing are often more costly than overall system development. As the economy based on digital technologies are growing rapidly, IT companies and professionals are paying lots of attention on system testing and user acceptance testing to create more positive influence on their business. User acceptance testing is often called the last phase of testing and it is very important because it is not the tester but the users will use it in daily basis. However, users might take longer time for testing because of non-IT experience.

For this purpose, [4] developed a five second testing (5ST) method for users based on their first impression of the website to save time and cost. Author also suggested that 5ST captures very first moment of website users but his investigation result suggested that it might not be very good idea to test web usability.

To solve this problem, based on the foundation of [4] research in this paper we conducted an experiment. Our experimental design made two groups of online customer to conduct web usability test in 60 seconds and unlimited time (3 days). The reasons to choose 60 second are given below

a)   According to [5], to evaluate intuitive speed and visio-spatial cognitive processing, 60 seconds is suitable.

b) Reference [6] used 60 seconds to test cognitive ability of people in UK.
c) In USA, for cognitive loads testing, psychologists give 60 seconds for the patients to list some information [7].

If the usability test results vary between two groups, IT companies might choose one method over another to test the web. If the two examinations produce same result, then stakeholders can chose 60 second test over long period testing which will save more money and time. We induced system usability scale and NetQu@l to test the usability of an online shopping website. Previous researches on usability testing were different. Researchers used chatting window, layout, navigation, shortcut keys, hand control and battle screen buttons as key factors for usability [8]. Ref. [9] conducted observational study to measure usability for augmented reality for tourism object. Perceived usefulness, ease of use, compatiability, self-efficacy and intention to use from [10] were used as the key constructs for measure usability for linus students in the research of [11]. SUS was also used in both industrial [12] and health settings [13].

The rest of this paper is organized as follows: at first, we review some literature related to usability testing method. After that previous studies on experimental design were discussed. Next we discussed our research methodology which includes participants profile, study design, scale, and survey questionnaire and data analysis techniques. Data analysis result, discussion and conclusion were followed after that.

2. RESEARCH METHOD

2.1. Participants

The present study was conducted among 60 online customers of 3 different online shopping page of Facebook. There were 11 female (18.33%) and 49 male (81.66%) software testers. Among the participants, 78% had 1- years of experience of online shopping. We message them via Facebook and they voluntarily participated. Participants were randomly assigned for test 1 (Unlimited test) and test 2 (60 second test). Participants profile is shown in Table 1.

Table 1. Participants' Profile

Category	Frequency	Percentage
Gender		
Male	49	81.66%
Female	11	18.33%
Online shopping experience		
1-3 Years	47	78.33%
4- 6 Years	11	21.33%
More than 6 years	2	1%

2.2. Study Design

Software testers participated in this present study. Two groups were formed by a random allocation. The 60SEC group comprises 30 participants and the unlimited test group comprises 30 participants. An e-commerce site was provided to them for testing. The time duration was three working days. After three days, they evaluated the website based on system usability scale and NetQu@l.

Two experimental conditions were provided for this research. 60SEC condition: for this situation, the index page of the website will be displayed.
a) Testers can move around the page, click, or download anything within 60 seconds. After 60 seconds, the website will redirect the tester to the survey questionnaire page.
b) UT condition: for this condition, tester can browse and test the website for 3 working days. Testers from UT condition can click the questionnaire button any time when they finished testing.

Two participant groups were offered both experimental conditions. Moreover, to each condition, participants were assigned randomly. Table 2 shows the instructions provided for each group

Table 2. Instructions

Group	Instructions
60SEC group	• Please visit the website, http://imstriker.com/rioMarket/index.html
	• Click on 60 Seconds button.
	• Explore the website within 60 seconds.
	• After 60 seconds, you will be redirected to the questionnaire page available in google forms.
	• Please evaluate the website based on the questionnaire.

Group	Instructions
UT group	• Please visit the website, http://imstriker.com/rioMarket/index.html • Click on unlimited time button. • Please check this website within 3 working days. • If you finished early, you can click on the respond to the survey button available on the menu bar. • Please assess the website based on the questionnaire available on the google form.

2.3. Scales and Questionnaire

We used both system usability scale (SUS) and netqu@l to assess the website for both experimental group. Both instrumental and non-instrumental features of a website can be evaluated by SUS and NetQu@l. See the Table 3 for classification.

Table 3. Classification of SUS and NetQu@l

	Scale	Non-Instrumental	Instrumental
	SUS		x
Netqu@l	Information quality		x
	Design		
	Ease of use	x	x

SUS inquires users to evaluate their level of agreement or disagreement to the 10 statements. For reporting results, we used a scoring template. This template turns the individual survey ratings into a single SUS score based rating according to Brooke's standard scoring method (statement ratings are manipulated to get a common 0-4 rating and then multiplying the sum by 2.5 to get a score. The score can range from 0-100). The questionnaire of SUS was adapted from the research [11, 14, 15]. Questionnaire contained 10 items. The items are given in Table 4

Table 4. Items of SUS

Group	Instructions
Q1	I think that I would like to use this system frequently
Q2	I found the system unnecessarily complex
Q3	I thought the system was easy to use.
Q4	I think that I would need the support of a technical person to be able to use this system.
Q5	I found the various functions in this system were well integrated.
Q6	I thought there was too much inconsistency in this system.
Q7	I would imagine that most people would learn to use this system very quickly.
Q8	I found the system very cumbersome to use.
Q9	I felt very confident using the system
Q10	I needed to learn a lot of things before I could get going with this system.

NetQu@l scale consists of 15 items. The items are categorized by design (3 items), information quality (6 items) and ease of use (6 items) of the system. We calculated the values of netQu@l by taking an average of all the responses given in Table 5.

Table 5. Items of NetQu@l

Category	Dimension
Ease of use	This site is easy to use
	It is easy to move about and to find out what are looking for on this site
	The layout of this site is clear and simple
	It is easy to find information
	The organization and layout of this site make it easier to find information
	This site has interactive features
Information quality	This site provides a detailed information about the products or services sold
	On this site, I can find unique or different products or services
	This site offers good illustrations of products or service sold out
	This site offers a wide variety of products and service
Design	This site is pretty
	This site is visually attractive
	This site is creative

The response scale for all items is a positive 5 point Likert scale [16] where 1 = strongly disagree, 2= Disagree, 3= slightly agree, 4= Agree, 5= Strongly Agree.

2.3.1 Data Analysis Strategy

We conducted independent sample t-test between data of the UT and 60SEC group. This comparison actually made between SUS score and Netqu@l score for both group. SPSS version 21 was used to test the data following the research of [11, 14].

3. RESULTS AND ANALYSIS
3.1. Reliability Analysis

At first we tested the reliability for both SUS and NetQu@l. Based on the responses, we found the reliability Cronbach's Alpha of our items are 0.81 and 0.89 respectively for SUS and NetQu@l. According to [17] reliability of survey questionnaire must be more than 0.7. As a result, we can conclude that our both set of questionnaire was understandable to our testers and satisfied the reliability benchmark.

3.2. Comparison of Scores

We conducted independent sample t-test between the SUS scores of UT and 60SEC group. Results are given in the Table 6 below.

Table 6. Result of SUS

Group	Mean	SD	Std. Error	Value
UT	62.39	16.339	3.483	t value= -0.468
60SEC	64.66	15.873	3.384	p value= 0.642

Result indicates the mean SUS score are 62.39 and 64.66 for UT and 60SEC group respectively. After independent sample t test result (t = -0.468, p = 0.642), we can conclude that there are no significant differences on system usability of this web between UT and 60SEC groups. After comparing SUS scores between two experimental groups, we compared the result of NetQu@l. Results are shown in the Table 7 below.

Table 7. Result of NetQu@l

Group	Mean	SD	Std. Error	Value
UT	3.266	0.699	0.149	t value= -0.239
60SEC	3.78	0.734	0.156	p value= 0.021

Comparison result of NetQu@l indicates there are significant difference (t= -0.239, p = 0.021) between the score of UT and 60SEC group as p value is less than 0.05. To get more accuracy, we also compare the result by category of NetQu@l. Results are shown in the Table 8.

Table 8. Categorized result of NetQu@l

NetQu@l	Category	UT	60SEC	Mean difference	p value	Remark
	Ease of use	3.43	4.00	0.57	0.014	Significant
	Information Quality	3.15	3.56	0.41	0.123	Not Significant
	Design	3.15	3.78	0.63	0.022	Significant

Result shows that, despite there are no significant difference on the score of information quality, however, in case of ease of use and design, there are different opinions of the testers. Therefore, we can imply that NetQu@l result will be different if manager provide the testers different time frame specifically for ease of use and design. Assessment result might be similar in case of testing information quality.

4.   CONCLUSION

A user's initial impression of a website is formed quicker than the ability to discern how that website is useful to them. Lindgaard et al. (2006) [18] showed that users could form an initial impression of a website in only 50 milliseconds, in the same way as if the site were presented to them for 500 ms. Also, Dahal (2011) [19] discovered that if a webpage required 2.66 seconds to attract the attention of users, their first impression had already been developed after just 180 ms.

This research explored the potential differences between testing modes of 60SEC test and unlimited test based on the research of [4]. The differences between the scores of both experiment suggests that in case of SUS, both experimental groups got similar result. In case of NetQu@l, both experimental groups achieved different scores. This study suggests us not to test design and ease of use in 60 seconds.

Regarding the forming first impression, this research showed that 60SEC test is not enough to measure all the qualities of a website. In general, the result of the study showed that, for non-instrumental quality (design) and instrumental quality (ease of use), our results is different between two groups. Regarding 60SEC test, testers focused only the links and download options. However, this experiment implies that to test system quality and information quality, software companies do not need to spend much time on testing as our two experiments results are not too different.

Additionally, this study investigated that in what way the 60SEC method can be applied and through what users construct their first impression using this time-specific method. Some scientific validations of 60SEC method were established by us, which needs further investigation. Like some other studies, this research is also got limitation because of small sample size. In future, experiment should be designed with bigger sample size. Additionally, as English is not a native language in Bangladesh, some participants faced difficulties to test this English language based website in 60 seconds [20].

This research we only conducted experiment based on system usability scale and NetQu@l. In future research, researchers can use other usability scale following the research of [21, 22] repoted like User experience questionnaire (UEQ) [23], Computer System Usability Questionnaire (CSUQ) [24], Usability Metric for User Experience (UMUX) [25]. The current research is another brick in what we believe, is a promising wall of study in the area of usability testing mode.

REFERENCES

[1]   J. Brooke, SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189-194). London: Taylor and Francis, 1996.

[2]   J. Kirakowski, & M. Corbett, "SUMI: The software usability measurement inventory". *British journal of educational technology*, vol. 24 no. 3, pp. 210-212, 1993.

[3]   B.D. Harper, & K.L.Norman, "Improving user satisfaction: The Questionnaire for User Interaction Satisfaction". Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference (pp. 225-233), Santa Monica, CA. Human Factors and Ergonomics Society, 1993.

[4]   G. Gronier, ", Measuring the first impression: testing the validity of the 5 second test, *" Journal of Usability Studies*, vol. 12 no.1, pp. 8-25, 2016.

[5]   E. Ferreira, D. Ferreira, S. Kim, P. Siirtola, J. Röning, J.F. Forlizzi &A. K. Dey, " Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults," In Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), IEEE Symposium on (pp. 39-48). IEEE, 2014.

[6]   S. McFall, "Understanding Society-UK Household Longitudinal Study: Cognitive Ability Measures. Understanding Society User Manual," 2013.

[7]   G. Bressolles,," La qualité de service électronique: NetQu@ 1 Proposition d'uneéchelle de mesureappliquée aux sites marchands et effetsmodérateurs., " *Recherche et Applications en Marketing (French Edition)*, vol. 21 no.3, pp. 19-45, 2006.

[8]   I. Tahyudin, D. Intan, S. Saputra, Haviluddin.," An Interactive Mobile Augmented Reality for TourismObjects at Purbalingga District, " *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 1 no.2, pp. 375-380 ,2016.

[9]   F. D Davis.,R.P. Bagozzi, & P.R. Warshaw, " User acceptance of computer technology: a comparison of two theoretical models, *"Management science*, vol. 35 no. 8, pp. 982-1003, 1989.

[10]   K. Awang.K, .S.N.W. Shamsuddin, I. Ismail, N.A. Rawi, M.M. Amin, "The usability analysis of using augmented reality for linus students, *"Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 13 no.1, pp. .58-64, 2019.

[11]   W. P. Wong, V. Veneziano & I. Mahmud, "Usability of Enterprise Resource Planning software systems: an evaluative analysis of the use of SAP in the textile industry in Bangladesh". *Information,"Development,* vol. 32 no.4, pp. 1027-1041, 2016.

[12]   B.Scholtz, I. Mahmud, & T.Ramayah (2016). "Does usability matter? An analysis of the Impact of Usability on Technology Acceptance in ERP Settings". *Interdisciplinary Journal of Information, Knowledge, and Management*,vol 11, pp. 309-330, 2016

[13] T. Bhuiyan, I. Mahmud & B. Alam . "Usability analysis of SMS alert system for immunization in the context of Bangladesh". *Int J Res Eng Tech*, vol. 2 no.11, pp. 300-5, 2013

[14] V. Veneziano, I. Mahmud, A. Khatun & W.W. Peng,. *"Usability analysis of ERP software: Education and experience of users' as moderators"*. In The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014) (pp. 1-7). IEEE, 2014.

[15] S. McLellan, A. Muddimer & S.C. Peres (2012). "The effect of experience on System Usability Scale ratings". *Journal of usability studies*, vol. 7 no. 2, pp. 56-67. 2012.

[16] T. Lam & A.J. Klockars,"Anchor point effects on the equivalence of questionnaire items", *Journal of Educational Measurement*, vol. 19 no. 4,pp. 317-322, 1982.

[17] J. Nunnally, , "Psychometric methods, Newyork: McGrawHill.,1978.

[18] G. Lindgaard, G. Fernandes, C. Dudek, & J. Brown. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, vol. 25 no. 2, pp. 115–126, 2006.

[19] S. Dahal. "Eyes don't lie: Understanding users' first impressions on website design using eye tracking". Master's Thesis, *Information Science and Technology*, Missouri University of Science and Technology. 2011.

[20] I. Mahmud., T.Ramayah & S. Kurnia, " To use or not to use: Modelling end user grumbling as user resistance in pre-implementation stage of enterprise resource planning system", *Information Systems*, vol. 69, pp. 164-179, 2017.

[21] A. Hinderks, M. Schrepp, F. J. D. Mayo, M.J. Escalona & J. Thomaschewski. "Developing a UX KPI based on the user experience questionnaire". *Computer Standards & Interfaces*, vol. 65, pp. 38-44, 2019.

[22] J.R. Lewis. "Measuring perceived usability: The CSUQ, SUS, and UMUX". *International Journal of Human–Computer Interaction*, vol. 34 no. 12, pp. 1148-1156.2018.

[23] J.R. Lewis, "Psychometric evaluation of the PSSUQ using data from five years of usability studies". *International Journal of HumanComputer Interaction*, vol. 14, pp. 463–488. 2002

[24] K. Finstad, "The usability metric for user experience". *Interacting with Computers*, vol. 22, pp. 323–327, 2010

[25] K. Finstad, "Response to commentaries on "The usability metric for user experience". *Interacting with Computers*, 25, 327–330, 2013.