PAON usability testing using system usability scale

To cite this article: K Setemen et al 2019 J. Phys.: Conf. Ser. 1165 012009

View the article online for updates and enhancements.
PAON usability testing using system usability scale

K Setemen\textsuperscript{1,a}, L J Erawati Dewi\textsuperscript{1,b}, and I K Purnamawan\textsuperscript{1,c}

\textsuperscript{1}Department of Informatics Management, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia

E-mail: \textsuperscript{a}k.setemen@undiksha.ac.id, \textsuperscript{b}joni.erawati@undiksha.ac.id, \textsuperscript{c}purna@undiksha.ac.id

Abstract. One of the important variables in developing software is users’ satisfaction of the software. Hence, there is a need to measure users’ satisfaction of the software’s that have been developed. The measurement of the users’ reactions covers some aspects such as users’ satisfaction, ease of use, efficiency, and whether the systems that have been developed can meet the users’ needs. This measurement is often called usability testing. It can be done using various instruments that have been developed by experts and communities in the field of computer. One of them is System Usability Scale (SUS). This article discusses an online peer assessment using usability testing. The instrument used was the usability scale system that was developed by John Brooke in 1986. The measurement using SUS yielded a mean score of 80.00. This mean score shows that the online peer assessment can be accepted and it can be categorized as grade B with Good rating. Hence, the online peer assessment is a feasible system.

1. Introduction

PAON (Online Peer Assessment) is a web-based system that was developed for assessing of a course. The process of using this system involves peers in a study group or a class. In addition to assessment made by the peer, there is also the assessment made by the teacher. Both are combined with loadings in accordance with setting or are based on the teacher’s discretion. The competitive advantages in this computer-based peer assessment are: 1) the teacher can design the course to be taught; 2) the teacher can assign students to a class to be taught; 3) the teacher can make a group of assessments; 4) the teacher can write or select rubrics to be used in the assessments; and 5) the teacher can make or start an assessment.

PAON is a system that has just been developed. Hence, it needs to be tested in terms of its acceptability and the users’ satisfaction with the system. This test can cover some aspects such as the users’ satisfaction, ease of use, efficiency, and the more important is whether the system can meet the users’ needs. This testing is often called usability testing. According to Erik, the usability of a product can be tested from two different perspectives: “ease-of-use” and “quality-in-use”\cite{1}. Furthermore, he stated that the first perspective has a limited scope. Ease and convenience are determined by the characteristics of the software themselves. Hence, usability is part of the quality of products.

Usability is the capacity of the software to be understood, learned, used and liked by users, when they are used in certain conditions\cite{2}. Nielsen \cite{3}, defines usability using 5 (five) quality components: 1) Learnability, the ability to which the users can learn the existing design for the first time; 2) Efficiency, how quickly the users can do their tasks; 3) Memorability, the ease with which the users can reuse the design after a long time they have not used it; Errors, the number of errors made by the users, the
seriousness of the consequence of committing them and how easily they can recover from the errors; and 5) Satisfaction, the pleasure that the users can have from the use of the existing designs. The principal goal of usability testing is for obtaining answers to the question: Is the product that has been developed useful to meet the need of the users. Research on usability has been done by many researchers such as Erik [1], Aprilia, et al. [4], Mansor, et al.[5], and Arh, et al.[6]. On the whole, the results described that usability testing is very useful in doing a research on products that have been developed. They described whether the products that have been developed can satisfy the users’ needs or not. Hence, the result of usability testing can be used as input for implementing a system or for improving it further in order the system can 100% meet the users’ expectation. Usability testing can use instruments that have been developed by experts and human computer interaction (HCI) communities. Some of the instruments are free of charge as long as the source is mentioned, but many have licenses, which means that if they are used the users have to pay to the owners. However, the most important thing to remember in using instruments is that the users should use standardized ones. According to Garcia [7], the advantages of using standardized instruments are: 1) Quantity, the standardized measurement enables the practitioners to report details better than when doing the measurement by using their own instruments; 2) Communication, it is easier for the researchers to communicate their findings when referring to a standard of measurement; and 3) Quick Comparisons, by using standardized questionnaire it is easy to compare different design iterations during the development process. Furthermore, Garcia [7], mentioned some standardized questionnaires, i.e., Software Usability Measurement Inventory (SUMI), Post-Study System Usability Questionnaire (PSSUQ), and the System Usability Scale (SUS).

This paper discusses a usability testing of the PAON system using the SUS. SUS was developed by John Brooke in 1986. It is very suitable for measuring usability of a computer system from the perspective of the users [8]. The SUS instrument was also selected since it is a standardized usability instrument with 43% level of popularity from various unpublished usability studies [7]. On the basis of research results, according to Bangor [9], SUS is very reliable (alpha = 0.91) and it is useful for various types of interface. This study also concluded that there was a small significant correlation between age and SUS score (SUS score decreased with the increase in age), but there was no effect on gender.

2. Method

This was a descriptive study. The intended descriptive study was the one that describes a phenomenon about an activity that can be done by a peer assessment called PAON. More specifically, this study surveyed users’ satisfaction through a usability test using the SUS questionnaire. The SUS questionnaire consists of 10 items containing statements that represent feelings or responses of the users to a product interface. The statements in the SUS questionnaire look like what is shown in Table 1. This questionnaire uses Likert scale (a five points scale) using the criteria of “strongly disagree” “disagree”, “neutral”, “agree”, and “strongly agree”. The ten statements are divided into positive and negative categories [8]. The positive statements have odd numbers (numbers 1,3,5,7, and 9) and the negative ones have even numbers (numbers 2,4,6,8, and 10). The calculation of the score for each positive statement is done by subtracting 1 score from every score given by the respondent. While the calculation of the score for each negative statement is done by subtracting a score given by a respondent from 5. Hence, the minimal score of each item is 0 and the maximal score is 2.5, so that the final score of each respondent ranges from 0 to 100. When it is written in a mathematical formula, the calculation of the SUS score for each respondent is as the following formula:

\[
SUS \text{ Score} = \{(S_1-1)+(5-S_2)+(S_3-1)+(5-S_4)+(S_5-1)+(5-S_6)+(S_7-1)+(5-S_8)+(S_9-1)+(5-S_{10})\}*2.5
\]

Note: \( S_i = \text{ the-}i \text{ item statement} \)

| Code | Statement |
|------|-----------|
| S1   | I think that I would like to use this website frequently |
Then, this SUS questionnaires were distributed to the respondents selected as the sample in this study. The sample used in this study consisted of 60 students of Informatics Management of the Universitas Pendidikan Ganesha. The sample was selected using a purposive sampling technique. This technique was used since at the same time the members of the sample were used as the sample of a limited trial of the system that has been developed.

3. Results and Discussion

The general description of the PAON system using uses case diagram is shown in Figure 1. Use case diagram is a series of activities that are interrelated and form a system of activities that are regularly done by an actor. In Figure 1 it is shown that two actors are involved in this system: a teacher and a student. The activities that can be done by the student are registration into the system, joining a class based on the material wanted, uploading tasks/tests, and assessing others. In this system, the teacher will write his or her course and open classes for the course. Then, he or she will open the classes if he or she is ready to accept students who will join the classes. Then, based on the classes that have been opened, the students are then combined into a class that is opened by him or her. The students will receive tasks or tests that they have to upload into the system according to the deadline that has been specified. When the task and test uploading time is over, the teacher will make groups of assessment based on the existing classes. These groups of assessment are used to evaluate the tasks/tests that have been uploaded by the students. Thus, the focus is to evaluate the tasks or tests done by their classmates.

![Figure 1. Use case diagram PAON system](image-url)
The SUS questionnaires have been distributed to 60 respondents. The results of the questionnaire completion were then calculated using the formula in formula (1). The final score result of the SUS for each respondent looks as shown in Table 2. The table shows that the mean score of SUS questionnaire is 80.00. It was this mean score that was used as the criterion score of the users’ acceptance of the system that was developed.

| R-i | Score-i | Final Score-i | R-i | Score-i | Final Score-i | R-i | Score-i | Final Score-i |
|-----|---------|---------------|-----|---------|---------------|-----|---------|---------------|
| R1  | 35      | 87.50         | R21 | 33      | 82.50         | R4  | 34      | 85.00         |
| R2  | 33      | 82.50         | R22 | 29      | 72.50         | R42 | 28      | 70.00         |
| R3  | 34      | 85.00         | R23 | 27      | 67.50         | R43 | 36      | 90.00         |
| R4  | 33      | 82.50         | R24 | 31      | 77.50         | R44 | 31      | 77.50         |
| R5  | 31      | 77.50         | R25 | 35      | 87.50         | R45 | 34      | 85.00         |
| R6  | 35      | 87.50         | R26 | 31      | 77.50         | R46 | 32      | 80.00         |
| R7  | 34      | 85.00         | R27 | 30      | 75.00         | R47 | 33      | 82.50         |
| R8  | 31      | 77.50         | R28 | 31      | 77.50         | R48 | 33      | 82.50         |
| R9  | 32      | 80.00         | R29 | 34      | 85.00         | R49 | 30      | 75.00         |
| R10 | 32      | 80.00         | R30 | 33      | 82.50         | R50 | 32      | 80.00         |
| R11 | 34      | 85.00         | R31 | 30      | 75.00         | R51 | 31      | 77.50         |
| R12 | 29      | 72.50         | R32 | 31      | 77.50         | R52 | 35      | 87.50         |
| R13 | 32      | 80.00         | R33 | 35      | 87.50         | R53 | 30      | 75.00         |
| R14 | 33      | 82.50         | R34 | 31      | 77.50         | R54 | 30      | 75.00         |
| R15 | 32      | 80.00         | R35 | 31      | 77.50         | R55 | 32      | 80.00         |
| R16 | 36      | 90.00         | R36 | 33      | 82.50         | R56 | 35      | 87.50         |
| R17 | 28      | 70.00         | R37 | 32      | 80.00         | R57 | 30      | 75.00         |
| R18 | 30      | 75.00         | R38 | 34      | 85.00         | R58 | 34      | 85.00         |
| R19 | 32      | 80.00         | R39 | 33      | 82.50         | R59 | 32      | 80.00         |
| R20 | 30      | 75.00         | R40 | 28      | 70.00         | R60 | 30      | 75.00         |

Mean score 80.00

Note:
R-i the i\textsuperscript{th} respondent
Score-i the total score of the i\textsuperscript{th} respondent
Final Score -i the final score of the i\textsuperscript{th}

The next step is to convert the mean score of the SUS into an acceptance level category, grading scale, and adjective rating. The categories for the SUS scores were developed by Bangor et al [9]. The acceptance level categories are classified into “Not Acceptable”, “Marginal”, and “Acceptable”. The grading scales are classified into grades “A”, “B”, “C”, “D”, and “F”. The classification for the adjective ratings is divided into “Worst Imaginable”, “Poor”, “OK”, “Good”, “Excellent”, and “Best Imaginable”. Each of them is based on the SUS mean scores as shown in Figure 2.

Figure 2. SUS score (Source: Bangor [9])
On the basis of Figure 2, a system can be accepted if it has a minimal SUS score of 70. Therefore, based on the result of the usability testing result of the PAON system, it turned out that it has an SUS score of 80.00, which means that the system can be accepted. In terms of the grading scale, the PAON system has grade “B”. In addition, based on the adjective ratings classification, the PAON system has the “Good” classification. According to Sauro, an SUS score can show promoters and detractors [10]. Furthermore, Sauro stated that the SUS mean score for promoters is 82 and for detractors is 67. Thus, if one wants a system to have a good power of promoters, then the SUS mean score has to be 80 at the minimum. In the light of the result of this usability testing, the PAON system is feasible to be used and implemented as an online assessment system.

4. Conclusion
Usability testing of the PAON system was done using the SUS questionnaire. The SUS was developed by John Brook in 1986. It is very suitable for use in measuring the usability of computer system from the users’ perspective. The SUS instrument was selected since it is a standardized usability instrument with 43% popularity level from various unpublished usability studies. Based on the results according to Bangor, SUS is very reliable (alpha = 0.91) and is useful for various types of interface.

The result of the testing of the PAON system using the SUS questionnaire showed that the mean score of the SUS questionnaire was 80.00. Based on this score the PAON system can be accepted as a system that can be implemented further. Furthermore, based on the SUS score conversion according to Bangor, the PAON system has grade B and falls into the Good adjective rating. Hence, according to this result, the PAON system is a system that is feasible to use further.

References
[1] D. Erik and P. W. M. V. Cisa, “Questionnaire based usability testing,” Conf. Proc. Eur. Softw. Qual. Week, Brussels, no. November, pp. 1–9, 1998.
[2] ISO 9126, “Information technology — Software product quality,” Iso/iec Fdis 9126-1, vol. 2000, pp. 1–26, 2000.
[3] J. Nielsen, “Usability 101: Introduction to Usability,” 2012. [Online]. Available: https://www.nngroup.com/articles/usability-101-introduction-to-usability/.
[4] I. H. N. Aprilia, P. I. Santoso, and R. Ferdiana, “Pengujian Usability Website Menggunakan System Usability Scale Website Usability Testing using System Usability Scale,” J. IPTEK-KOM, vol. 17, no. 1, pp. 31–38, 2015.
[5] Z. Mansor, Z. M. Kasirun, S. Yahya, and N. H. Arshad, “The Evaluation of WebCost Using Software Usability Measurement Inventory ( SUMI ),” Int. J. Digit. Inf. Wirel. Commun., vol. 2, no. 2, pp. 197–201, 2012.
[6] T. Arh and B. J. Blažič, “A Case Study of Usability Testing - the SUMI Evaluation Approach of the EducaNext Portal,” WSEAS Trans. Inf. Sci. Appl., vol. 5, no. 2, pp. 175–181, 2008.
[7] A. Garcia, “Choosing the Right Usability Questionnaire | ChaiOne,” pp. 1–7, 2013.
[8] J. Brooke, “SUS: A Retrospective,” J. Usability Stud., vol. 8, no. 2, pp. 29–40, 2013.
[9] A. Bangor, P. Kortum, and J. Miller, “Determining what individual SUS scores mean: Adding an adjective rating scale,” J. usability Stud., vol. 4, no. 3, pp. 114–123, 2009.
[10] J. Sauro, “DOES BETTER USABILITY INCREASE CUSTOMER LOYALTY?,” https://measuringu.com/usability-loyalty/, 2010. [Online]. Available: https://measuringu.com/usability-loyalty/. [Accessed: 31-Aug-2018].