Empirical study on the use of two remote asynchronous usability testing methods

S B Yudhoatmojo and S F Sutendi

Faculty of Computer Science, Universitas Indonesia, Kampus UI, Depok, 16411, Jawa Barat, Indonesia

E-mail: satrio.baskoro@cs.ui.ac.id

Abstract. Usability testing is used to ensure that the application can be used by the users and according to the user needs. Basically, there are two usability testing methods: the conventional method and the remote method. In this study, we focus on the remote method (more specifically on the remote asynchronous method). The remote asynchronous method means that the tester and the evaluator of the test result are separated by time and location. Within this method, we compared the use of forum and diary. We analyzed the use of these two methods in remote asynchronous method. The goals of this study are to find out which of these two are more efficient in identifying the usability problems, and the advantages and disadvantages of these two methods. The result of our study shows that the diary enables us to identify more usability problems. Our further analysis shows the reason for this is due to the constant daily exposure to the application during the whole testing process. This also allows the tester to become more familiar with application which leads to a faster time in finishing the test than the forum method. Nevertheless, the number of respondents who were able to finish completely the test scenario using the diary method is less than the forum method. We probed this matter further and found that the reasons are they can be persistence in working on the test every day and/or they just completely forgot to do the test on certain days.

1. Introduction

Usability testing is performed to test the usability level of an application. This testing allows the possibility to develop a better application[1]. The result of this test is also beneficial for the users because it can be used as a consideration for further improvement of the tested application. In every test, there is someone who run test (we called this tester) and the one who evaluate the result (we called this evaluator). Basically, there two usability testing methods, that is the conventional method and the remote method. The conventional method restricts the tester and the evaluator to be in the same room at the same time, while the remote method does not.

In our study, we focus on the remote method especially on remote asynchronous method where the tester and the evaluator are separated by time and location. Within this method, we focus more on the use of forum and diary. We compared the two remote asynchronous methods. The following is our research questions:

RQ-1: Which of the two remote asynchronous methods (the forum and diary) is more effective in identifying usability problems?

RQ-2: What are the advantages and disadvantages of the two remote asynchronous methods?

This paper is organized as follows. Section 2 describes our literature review to support our study.
Section 3 shows the research methodology we used. Section 4 discusses the result of our study. Section 5 concludes this paper.

2. Literature Review

This section describes the literature review that we have conducted to support our study. The reviewed literatures consist of literatures about user experience, usability, usability testing, remote asynchronous usability testing method and related works.

User experience is one of important aspects in a development of a product. User experience or UX is defined as a process for identifying the expected customer needs of the developed product[2]. There are seven factors which influence the user experience[3]: useful, usable, desirable, findable, accessible, credible, and valuable. Usability is a quality attribute for measuring the easy-to-use level of an interface design[4]. While on ISO 9241-11, it is stated that usability is a measurement against the usefulness level of a product based on its effectiveness, efficiency and satisfaction on the context of the product’s objectives[5]. Usability has five components[4]: learnability, efficiency, memorability, errors, and satisfaction. Problems related to usability can be categorized into three categories[6]: (1) Critical – this problem leads to possibility for a user to make a mistake in doing his/her tasks; (2) Moderate – this problem occurs on some users to make a mistake and slows down his/her tasks; and (3) Minor – this problem causes the user to doubt the application.

Usability Testing is an important sub-activity in developing an application. This type of testing is important to the designer because without it the designer can be assured if the developed application is usable and according to the user needs or not[7]. An effective usability testing takes (1) a mature test plan; (2) user sample recruitment; (3) test result analysis and report as important issues[8].

There are two usability testing methods, the conventional method and the remote method. The conventional usability testing method restricts the evaluator of the testing and the user samples who are assigned as the testers to be in the same place at the same time. The remote usability testing method is the opposite of the conventional one. Furthermore, the remote usability testing method is divided into two types which is based on the testing time. The two types are moderated usability testing (or remote synchronous usability testing) and unmoderated usability testing (or remote asynchronous usability testing). The former is conducted in different locations by the tester but the evaluator is accompanying the tester virtually[9]. This method is suitable for complex tasks where the execution steps are unstructured[10]. The latter is conducted without any assistance from the evaluator[9]. This method is suitable for relatively simple tasks with specific execution steps[10].

Remote asynchronous usability testing is a usability testing where the user and the evaluator are separated between time and location. This method takes a longer time to finish and the number of usability problem identified is relatively smaller than the remote synchronous usability testing[11]. Nevertheless, this method is useful for evaluator who has busy schedule in performing usability testing and, for testing which requires large amount of participants[12]. There are two ways in performing remote asynchronous usability testing that we used in our study: Forum and Diary.

In forum or forum-based online reporting and discussion method, the user is given several tasks. At the end of the tasks, the user must create a note about the problems encountered during performing the tasks. The tasks are done only once and the duration of each task is recorded. The user reports the problems in the forum and discusses those problems with other users. The rule of game is that the user must first check whether there is another user who already post the same problem prior to the user. If it is already existed then the user does not have to post it again but the user must positively or negatively response to the post.

In diary-based longitudinal user reporting, the testing is performed on several days. On the first day, the user is given several tasks and if the user finished them then the user must report the problems encountered and the duration for completing each task. The following days, the user is given additional tasks[11]. This method is suitable for performing a test in which the respondents are distributed and are not able to meet with the evaluator in person[7].
In previous study, the researchers compared four methods (three remote asynchronous usability testing and one remote synchronous usability testing)[11]. The four methods are: (1) the conventional user-based laboratory test (Lab); (2) the user-reported critical incident (UCI); (3) the forum-based online reporting and discussion (Forum); and (4) the diary-based longitudinal user reporting (Diary). There were 40 respondents which are divided evenly for each method. The result of this study shows the remote asynchronous usability testing is able to identify usability problem. Nevertheless, the number of usability problem identified is half of the number identified by remote synchronous usability testing.

3. Research Methodology

This study uses two approaches: quantitative and qualitative. The quantitative approach is used for analysing the numerical data obtained in the study, such as the duration time for performing the test. On the other hand the qualitative approach is for analysing the forum and diary contents. Our research methodology consists of ten steps. The first one is identifying the problem we are trying to experiment on. The second step is doing a literature review for our laying the foundation of our study. Third, we design the experiment. Fourth, we determine the respondents. Five, we design the testing scenario. Six, we do a readability test against our testing scenario in which we may return to step five if we need to redesign the scenario. Seven, we perform a pilot testing. If the pilot testing result is poor then we may need to go back to step six. Eight, we perform the actual usability testing. Nine, we analyse the data. Ten, we conclude our study.

The application we tested is an in-house application developed by our faculty. It is a student peer-reviewed application which is designed for assessing group work contributions. We design the same tasks for the forum and diary methods. The main difference is the duration. The forum method lasted for one day while the diary lasted for five days where on the following days we add new tasks. For the forum, the user writes the notes on the given online discussion forum while, for the diary, the user writes the notes on Google Docs.

The analysis of qualitative data is conducted using a focus group discussion with several experts to identify the usability problems. For the quantitative data, we used the paired-sample T test on SPSS 22 software. This approach is used to compare the significance of the average total duration of each task on both methods.

There are two groups of respondents which represent the two groups of user; the lecturer and the student. We chose 30 students using convenience sampling. According to [13], 30 samples is a minimum amount of sample for analysing statistic data, but, along the way of this study, the actual respondents fall into 26 students. This is still suffice because according to [14], usability testing has five as the minimal number of respondents and 20 respondents for quantitative data. As for the lecturers, we took four lecturers as our respondents. These four lecturers are all lecturers of a course which uses the application in assessing the group work contribution.

The testing scenarios design went through four times readability tests and the testing scenario design used considerations from [15]. The readability tests are conducted to four people in which, two of them are the respondents. Once passed the readability tests, we are ready to do the usability testing.

4. Result

This section describes the result of the study. We analysed both the qualitative and quantitative data where the analysis are presented in their own sections.

4.1. The Qualitative Data Analysis

The data used in the analysis are forum posts for the forum usability testing method and diary entries for the diary usability testing method. Both of these data are analysed using a focus group discussion (FGD) to identify the number of usability problems encountered by the respondents. This FGD is used for obtaining a consensus about the usability problems identified by the expert analysts in the FGD. The usability problems are categorized into one of these classes: critical, moderate, minor, as
mentioned in Section 2 about usability problems. After each expert, independently, identified the usability problems, they discussed them to obtain the final decision about the usability problems they identified and their categorization. Table 1 shows the amount of usability problems grouped by the type of respondents and the usability testing method.

|               | Critical | Moderate | Minor | Total |
|---------------|----------|----------|-------|-------|
| Lecturer Forum| 3        | 0        | 2     | 5     |
| Student Forum | 2        | 1        | 1     | 4     |
| Lecturer Diary| 2        | 0        | 0     | 2     |
| Student Diary | 3        | 2        | 0     | 5     |

Table 1. The number of usability problems from the forum and diary usability testing method grouped by respondent type

For the lecturer group respondent, the number of usability problems decreases in the diary method. Further analysis shows that lecturers are not novice users and plenty of experiences with the application. They also able to overcome some obstacles in using the application, thus they did not write them on the diary. Opposite to the lecturer group respondent, the student group respondent shows an increase in the usability problems using the diary method.

4.2. The Quantitative Data Analysis

The quantitative data used in our analysis is the time duration of the testing. The average time of each method is compared using paired sample T-test. This method is chosen because the respondents are working on the same tasks using two different methods. Table 2 shows the total time duration for lecturer group respondent and Table 3 is for the student group respondent.

Table 2. Total time duration of the testing for the lecturer group respondent

| Respondent | 1  | 2  | 3  | 4  | 5  |
|------------|----|----|----|----|----|
| Forum      | 273.8 | 306 | 181 | 370.6 | 205 |
| Diary      | 283  | 252 | 103 | 496 | N/A |

Table 3. Total time duration of the testing for the student group respondent

| Respondent | Forum | Diary | Respondent | Forum | Diary |
|------------|-------|-------|------------|-------|-------|
| 1          | 360   | 102.5 | 14         | 374   | 136   |
| 2          | 150   | 108.5 | 15         | 360   | 177.17|
| 3          | 540   | 243.5 | 16         | 480   | 190   |
| 4          | 420   | 221   | 17         | 122.6 | 191.5 |
| 5          | 900   | 216.67| 18         | 270   | 293.33|
| 6          | 420   | 150   | 19         | 360   | 388.33|
| 7          | 292   | 257.25| 20         | 540   | 221.5 |
| 8          | 330   | 195.33| 21         | 360   | 236   |
| 9          | 1080  | 112.83| 22         | 410   | 412.67|
| 10         | 304   | N/A   | 23         | 360   | 230.17|
| 11         | 660   | 318.167| 24       | 420   | 194.83|
| 12         | 372   | 185   | 25         | 143   | 152.67|
| 13         | 480   | 195   | 26         | 134   | N/A   |

The comparison of the duration is conducted on each group. The Saphiro-Wilk test against the lecturer group respondents shows a value of 0.602 where the significant value of this test is over 0.05 (p > 0.05). Thus, this group is normally distributed. Next, we test the data using paired-sample T-test using SPPS 22 software. The significant value using two-tailed test is 0.08. This means there is no significant difference between the duration of the test performed using the forum and the diary. The test shows no significant difference between the two although the average duration for the diary is
144.694 seconds while in forum is 285.867 seconds. These differences seem to not cause any significant difference in the pair-sample T-test.

For the student group respondent, the result of Saphiro-Wilk test shows that the data is not normally distributed (the value is 0.000), thus we needed to transform the data. The transformation of the data is conducted by removing two outliers, in which is the ones with duration over 500 seconds. Once we removed the outliers, we retested them again using Saphiro-Wilk test and the result is normally distributed (the value is 0.08). The result of the paired-sample T-test shows the significant value of 0.000 (p < 0.05). This shows there is a significant difference between the test duration time between the forum and the diary. In using the forum method, the average duration time of the test is 373.8 seconds while in the diary is 218.2 seconds.

4.3. The Comparison between Forum and Diary Usability Testing Method
The test respondents executed the test scenarios using two remote asynchronous usability testing methods, that is the forum and diary. The experiments lasted for two weeks for each group of respondents. The aim is to determine which method is more effective to be used and to identify the usability problems. Table 4 shows the advantages and disadvantages of both methods based on our empirical study.

|                  | Forum                                      | Diary                                      |
|------------------|--------------------------------------------|--------------------------------------------|
| **Advantages**   | The number of respondents who are doing the test are larger than the diary method | The number of usability problems identified are larger than the forum method |
|                  | The execution of the test is easier because it is only executed once per-week | The duration time in executing the test is lesser than the forum method |
|                  | The ability to see problems from someone else’s point of view via shared problems in the forum | The comments about the application are more objective |
| **Disadvantages** | The respondent has tendency to be influenced by other respondents during the discussions held in the forum | This method allows the user to be more familiar with the application |
|                  | The respondents are less participated in the discussion forum | This method is harder to do because it occupies the time to do the tasks one at a time, day after day. |
|                  | The number of usability problems identified are lesser that the diary method | It takes more effort to remind the respondents to work on the testing scenarios |

In our findings, we found that the declining number of respondents, who work on the test scenarios using the diary method, is due to one of the constraints of the method. The constraint is that the respondent must work on the test scenarios every day for one week. The persistence in doing this is difficult because some respondents have other matters to do on that day and sometimes they just completely forgot to do the test. The lesson that we can learn from this is that for future reference, we as the evaluator must actively remind the respondent to do the test.

Other thing that we found is that the diary method seems able to identify larger number of usability problems than the forum method. We can exhibit this in the student group respondent but not in the lecturer group respondent. We suspect that since the lecturers have been using the application, in practice, longer than the students through numerous semesters teaching the same courses, which use the application, then the lecturers are already familiar with the functionalities and the problems existed within the application. Thus, they can alleviate the encountered usability problems.

For the forum method, respondents can find usability problems from other respondents by looking at the discussion forum. The problem can be similar to the one encounter by the respondent or completely a new one. This is one of the benefits of forum method. The respondents gain more insight about usability problems of the application by sharing and discussing about the usability problems.
Nevertheless, this method can cause a tendency for the respondents to share the same usability problems and get influence by other respondents in sharing the problems.

5. Conclusion

This paper described the result of our study about comparing the use of two remote asynchronous usability testing applied known as the forum and the diary method. We used a peer group work review application, which is an in-house developed application, which we use in our faculty. The respondents who are performing the test comprise of two groups: the lecturer group and the student group. The respondents work on the same test scenarios using those two different methods. We found that the number of respondents using diary method is lower than the forum method. This is because the respondents need to work on the test scenario one day at a time. But, the familiarity level about the application increases better using the diary method. The result shows that the diary method is faster than the forum method in terms of the duration of working on the test scenario. This caused by constant exposure to the application daily throughout the whole process of the usability testing. This method can improve three usability components: memorability, learnability and efficiency.

References

[1] Thompson K E, Rozanski E P and Haake A R 2004CITC5 '04Here, there, anywhere,” in Proceedings of the 5th conference on Information technology education. p 132

[2] Norman D and Nielsen J 2018 The Definition of User Experience (UX) Nielsen Norman Group. [Online]. Available: http://www.nngroup.com/articles/definition-user-experience/. [Accessed: 12-Jun-2018].

[3] Morville P 2004, “User Experience Design,” Semantic Studios, 2004. [Online]. Available: http://semanticstudios.com/user_experience_design/. [Accessed: 01-Jun-2018].

[4] Nielsen J 2018 , “Usability 101: Introduction to Usability,” Nielsen Norman Group, 2012. [Online]. Available: http://www.nngroup.com/articles/usability-101-introduction-to-usability/. [Accessed: 01-Jun-2018].

[5] “ISO 9241-11:1998(en) - Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability,” ISO, 1998. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en. [Accessed: 15-Jul-2018].

[6] Sauro J 2013 , “RATING THE SEVERITY OF USABILITY PROBLEMS,” Measuring U, 2013. [Online]. Available: http://www.measuringu.com/blog/rating-severity.php. [Accessed: 01-Jun-2018].

[7] Y. Rogers Y, Sharp H, and Preece J 2011 WileyInteraction design : beyond human-computer interaction

[8] “Usability Testing,” usability.gov. [Online]. Available: https://www.usability.gov/how-to-and-tools/methods/usability-testing.html. [Accessed: 01-Jun-2018].

[9] Nelson E T and Stavrou A 2011, “Advantages and Disadvantages of Remote Asynchronous Usability Testing Using Amazon Mechanical Turk,” Proc. Hum. Factors Ergon. Soc. Annu. Meet.55(1) pp 1080–1084

[10] “Remote Testing,” usability.gov. [Online]. Available: https://www.usability.gov/how-to-and-tools/methods/remote-testing.html. [Accessed: 01-Jun-2018].

[11] Bruun A, Gull P, Hofmeister L and Stage J 2009 Let your users do the testing in Proceedings of the 27th international conference on Human factors in computing systems. p 1619

[12] Andreasen M S, Nielsen H V, Schroder S O and Stage J 2007What happened to remote usability testing? in Proceedings of the SIGCHI conference on Human factors in computing systems. p 1405

[13] Cohen L, Manion L and Morrison K 2018Research Methods in Education. (Routledge:Routledge)

[14] Nielsen J 2012How Many Test Users in a Usability Study? Nielsen Norman Group

[15] Dumas J S and Redish J 1999Intellect BooksA practical guide to usability testing.