User Interface Evaluation - Assignment 4

Group 6

Amber Kiewiet

Dian Arends

Frans Simanjuntak

March 16, 2017

1. Introduction

This report presents the evaluation of **myPrint** [1], an online printing system at University of Groningen, using two evaluation methods, think aloud and questionnaires. Unlike the methods used in the previous assignments, these evaluation methods are involving users to assess the system.

For the think aloud method, the experiments were based on the procedures discussed in [1]. The two methods are coming from other protocol, the flexible Boren & Ramey protocol [2] and the rigid Ericsson & Simon protocol [3]. The experiment was carried out by the team members.

For the questionnaire method, the paper "A Comparison of Questionnaires for Assessing Website Usability"[4] was used as the base for creating questionnaires. Three questionnaires were created in order to evaluate the system: System Usability Scale (SUS), Computer System Usability Questionnaire (CSUQ), and Sixpoint System Usability Scale (SSUS).

2. Think Aloud Method

As mentioned earlier, for the think aloud method, the experiments used two protocols namely the flexible Boren & Ramey protocol and the rigid Ericsson & Simon protocol. The Ericsson & Simon protocol does not allow the experimenter to interact with the subject, except for a "keep talking" reminder when a subject falls silent for more than 15 to 20 seconds [1]. While the first protocol doesn't allow the participant to interact with the examiner, the Boren & Ramey protocol is different in three respects. In this method, the experimenter gives "mm-hmm" tokens as feedback and as reminders to continue verbalization after 15-20 seconds of silence, the experimenter is allowed to repeat a single word to trigger

[1] https://myprint.rug.nl/

clarification, and the experimenter is allowed to encourage a user who is stuck by giving a non-directive suggestion [1].

In order to conduct the experiment, we came up with four tasks for the participants to perform:

1. Try to change the default printing preferences from 'Black and White' to 'Color' and 'Staples off' to 'Staples on'.
2. Try to print a document from your computer.
3. If it is possible, try to purchase €15,- credits for printing.
4. How do you set a notification when your printing credits are lower than €2,50?

The results of both methods are discussed in the following subsections.

2.1 Ericsson & Simon Protocol

The E&S protocol was carried out by two team members, one acted as a participant and the other as experimenter. The first task was to change the default printing preferences. After logging in and landing on the homepage, the participant said: "Where do I go know". The participant was not sure where to go, but after a quick look around he went to 'MyProfile'. The participant changed the preferences and clicked on 'Save'. He got no response from the system and said: "OK, it's done. But.. I'm not sure whether.. It doesn't say anything. I'm confused". After saying that the participant was silent for a few seconds, so it felt natural for the experimenter to interact with the participant, but it was not allowed. After the silence the participant went on with the second task.

The second task was to try to print a document. The participant went back to the homepage and changed the preferences back to how they were before task one. This was not part of the task, but the experimenter was not allowed to ask why the participant did that. The participant selected a file and clicked on 'Print'. After that the participant said: "Oh oh. I cannot see any uploaded documents.. How do I do that?" That's where the experimenter tend to interact with the participant, but this was not allowed. After a short silent and a "keep talking" from the experimenter, the participant said: "I don't know". This is where some interaction may have gained more valuable information. The participant completed this task and continued to the third task.

The third task was to try to purchase 15 euro credits for printing. The participant immediately saw the 'Purchase credits' button and clicked on it. Then the participant was a bit confused and said: "15.. I can't

do that. But I think I can do it twice.. Buy 10 first and then 5..” So the participant first purchased 5 euro and after that 10 euro without any trouble.

The fourth and last task was to change the notification settings to 2,50 euro. This task took long to complete. The participant was not sure where to find it and where to go, and said: “Notifications.. Hmm I don’t know where. Should it be here in the menu or something?” He could not find it there, so he said: “Or maybe I can see it in help”. Then the participant went to ‘Help’ and typed ctrl + ‘F’ and typed in ‘notification’, but did not find anything. The participant tried to type in ‘Alert’ and that worked, so he said: “Oh that’s it. Okay I have to go to ‘MyProfile’ and then click on ‘Change alerts’. So he went to ‘MyProfile’ and ‘Change alerts’ and said: “Done”.

The participant in the first evaluation (E&S protocol) had to get used to thinking aloud, so a few times the participant felt silent. It was also difficult for the participant to speak when focusing on the task, because it did not feel natural. So at the beginning the participant had to get comfortable with thinking aloud, but at the end it got way better. As Krahmer and Ummelen (2004) write, the first think-aloud method (E&S protocol) does not allow the experimenter to interact with the participant, except for a “keep talking” reminder when a participant falls silent for more than 15 to 20 seconds. For the experimenter sometimes it was hard to only say “keep talking”, because it feels more natural to interact with the participant. Overall the participant said for three times to be confused. That is where an interaction between the participant and the experimenter may have gained valuable information, but it was not allowed in this method. All tasks were completed by the participant himself in the E&S protocol, but a bit slower than normally. The vocalization of the thoughts made the participant slowdown in completing the tasks, but in the end it worked.

2.2 Boren & Ramey Protocol

The evaluation was carried out by two team members, one acted as participant and the other as examiner. The first task to perform was to change the default printing preference. At the very beginning, the instruction seemed not very clear to the participant because the user changed the printing preference on the homepage while the goal was to change the default printing preference which placed under ‘MyProfile’. Since the participant made a mistake, the examiner explained the actual task and guided how to do that. After the participant changed her default printing preferences, she was very surprised with the results because the system informed nothing about the changes she made. She was expecting an alert or notification came out saying that the changes are successfully saved or fail. So, she asked the examiner

about it and the examiner explained that the changes she made have been saved successfully even though there is no alert.

For the second task, the participant completed the task very quickly because the print button was placed on the homepage of the website which made it easier for participant to perform the task. However, again, the participant was surprised with the printing alert since it doesn't stand out for long. Beside, the user can't see the list of documents she already uploaded. She got confused and she asked this issue to the examiner. The examiner explained that the documents she uploaded have been recorded in the system and she could check the documents on the printer machine.

The next task is to purchase €15 printing credit. The purchase button is easy to spot because it's placed on the homepage at the bottom of page. Before purchasing the credit, the very first thing the participant did was to check her remaining balance. After that she clicked the purchase button and tried to purchase €15. However, she got stuck again because the available amount provided on the screen is €5, €10, €20, and €50 while the goal of the task is to purchase €15. She tried to click somewhere else on the screen to figure it out how to do it but nothing worked. Since the participant got confused, the examiner asked the participant what's going on. The participant explained that she could not purchase €15 at once, so she proposed to do it twice, €5 and €10 respectively. The examiner agreed because that's the only way to do that. Then, the user purchased the credit twice and the most interesting thing is by the end of each purchase action, the participant never forgot to refresh the page in order to check whether the printing credit has been updated or not.

The last task of the experiment is to set a notification when the credit is lower than €2,50. At the beginning, the participant seemed lost and had no idea how to do that. Since the examiner noticed this, the examiner informed the participant where the alert setting was placed. The participant followed the advice and finally she found the alert setting. Then, she changed the value of credit from €1.50 to €2.50 and submitted the changes afterwards. Since the participant made some changes, the system notified her by popping out alert "Notification threshold saved". The participant seemed very happy with this notification.

From this experiment, we can see that the participant can perform some tasks very easy such as print the document and purchase the credit because those features were placed on the homepage, so it's easy to spot them. However, for the other tasks such as change the default printing preference and set the printing notification, these tasks seem difficult to do because these features are not easy to find on the website

since they are placed under 'MyProfile'. If the participant was not guided by the examiner to perform these tasks, she needed extra time to find out how to do it which could be done by surfing the entire website or reading the documentation.

2.3 Differences between both protocols

The biggest and most predictable difference between the two protocols was the interaction. In the first method (E&S) the experimenter was not allowed to say more than a "keep talking" after a silence of 15 to 20 seconds. In the second method (B&R) the experimenter was allowed to do the following things of interaction according to Krahmer and Ummelen (2004): (1) the experimenter gives "mm-hmm" tokens as feedback (and as a reminder to continue talking after 15-20 seconds of silence), (2) the experimenter is allowed to repeat a single word to trigger clarification, and (3) the experimenter is allowed to encourage a user who is stuck by giving a non-directive suggestion [1]. So for the first method, the participant did not explained many things. The second method gave the experimenter the opportunity to get more information out of the participant, just by triggering for clarification. Another case in which the interaction was different was when the participant got a bit stuck. By the first method (E&S) the participant had to find out himself, and after trying some things it did work, but the participant sometimes got a bit irritated. By the second method (B&R) when the participant got stuck, the experimenter was able to give a non-directive suggestion after a few seconds. So they sort of got out of it together. At the second method the participant was less confused or irritated.

3. Questionnaires

Tullis & Stetson conducted a study to determine the effectiveness of some of the standard questionnaires used for assessing the perceived usability of interactive systems [4]. According to their results, the most effective questionnaires for a small number of participants where the System Usability Scale (SUS) and the Computer System Usability Questionnaire (CSUQ). Therefore, we decided to use these two questionnaires. Furthermore, in a 'real world setting' we would choose to question 12 students, employees and guest from the University of Groningen (RUG). The experiment of Tullis & Stetson showed that chosen questionnaires appear to reach an asymptote at a sample size of 12 participants [4]. Therefore, we would randomly assign 12 participants to the SUS, 12 to CSUQ, and 12 to our own questionnaire. However, for this assignment we acted as participants and filled in the questionnaire ourselves after finishing the think aloud method. By filling in the questionnaire after these protocols, we hope to collect more information on problems in the area of the perceived usability of myPrint.

3.1 System Usability Scale (SUS)

This questionnaire consists of ten questions, which Tullis & Stetson adapted for the evaluation of a website by replacing the word "system" in every question with "website" (see Appendix A). Each question is a statement and a rating on a five-point scale of "Strongly Disagree' to "Strongly Agree" [4]. We placed the SUS on SurveyMonkey[2].

3.2 Computer System Usability Questionnaire (CSUQ)

This questionnaire is composed of 19 questions, which Tullis & Stetson adapted for the evaluation of a website by replacing the terms "system" or "computer system" with "website"(see Appendix A). Each question is a statement and a rating on a seven-point scale of "Strongly Disagree" to "Strongly Agree" [4]. We placed the CSUQ on SurveyMonkey, but we had to divide the questionnaire in a part one[3] and two[4] since we used a free account on the website.

3.3 Sixpoint System Usability Scale (SSUS)

For this assignment, we also decided to create our own questionnaire based on the earlier mentioned SUS. We only changed the size of the scale, since Tullis & Stetson state that the questions yielded the most reliable results across different sample sizes [4]. The original SUS uses a five-point scale, while our version uses a six-point scale. This way, the user cannot stay neutral whilst answering the questions. We placed our SSUS on SurveyMonkey[5].

3.4 Results

When we compare the results of the three questionnaires, a few interesting findings relative to the perceived usability of myPrint arise. First of all, results of both the SUS and SSUS showed that the participants found the website unnecessarily complex. This is in line with the results of the two think aloud protocols. Besides, the participants unanimously disagreed with the fifth statement '*Various functions in the website are well integrated*' (Table 1).

[2] https://nl.surveymonkey.com/r/PYK97DF
[3] https://nl.surveymonkey.com/r/3PLYG8M
[4] https://nl.surveymonkey.com/r/35622KB
[5] https://nl.surveymonkey.com/r/G72VMM6

Table 1: Results SUS statement 5

	Strongly disagree	Disagre	Neutra	Agre	Strongly agree	Totaal	Gewogen gemiddelde
Outcome	0,00% 0	100,00% 3	0,00% 0	0,00% 0	0,00% 0	3	2,00

However, the results of the SSUS show that one participant somewhat disagrees with the comment (Table 2). Therefore, the problem with the lack of integration of various functions might not be as severe as the SUS initially suggests.

Table 2: Results SSUS statement 5

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly agree	Totaal	Gewogen gemiddelde
Outcome	0,00% 0	66,67% 2	33,33% 1	0,00% 0	0,00% 0	0,00% 0	3	2,33

A third problem arose in the area of the consistency of myPrint. According to the SUS results, two participants agreed and one strongly agreed with the sixth statement '*I thought there was too much inconsistency in this website*'.

Table 3: Results SUS statement 6

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Totaal	Gewogen gemiddelde
Outcome	0,00% 0	0,00% 0	0,00% 0	66,67% 2	33,33% 1	3	4,33

But then again, when the participants were given an extra option on the scale ("Somewhat agree"), the problem appears to be not as severe as the results in Table 3 suggest. When we compared the SSUS results with the previous table, the participant who previously chose 'Strongly agree' changed his/her answer to 'Agree'. Another participant also changed his/her previous answer, but he or she chose 'Somewhat agree' instead of 'Agree'.

Table 4: Results SSUS statement 6

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly agree	Totaal	Gewogen gemiddelde
Outcome	0,00% 0	0,00% 0	0,00% 0	33,33% 1	66,67% 2	0,00% 0	3	4,67

Finally, the results of the fifteenth statement of the CSUQ show a perceived problem in the area of the organization of the information on the webpages. The participants unanimously (somewhat) disagree with the statement: '*The organization of information on the website pages is clear'* (Table 5). This is in line with the results of the think aloud protocol. One participant of the this protocol stated that he did not know where to find information about the notification settings or where to change them.

Table 5: Results CSUQ statement 15

	Strongly disagree (1)	(2)	(3)	(4)	(5)	(6)	Strongly agree (7)	N.v.t.	Totaal	Gewogen gemiddelde
Outcome	0,00% 0	0,00% 0	100,00% 3	0,00% 0	0,00% 0	0,00% 0	0,00% 0	0,00% 0	3	3,00

However, since the participants did not choose the options 'Disagree' and 'Strongly Disagree', the problem with the information organization of myPrint might not be perceived as severe as the results of the protocols suggests.

4. Process

For this assignment, we divided the workload for the different methods. First of all, we created four different tasks for the think aloud method together and choose which questionnaires we were going to use after the protocols based on the article of Tullis & Stetson [4]. Next, Dian acted as a participant during the Boren & Ramey protocol and Frans during the Ericsson & Simon protocol. During the performance of the four chosen tasks, two group members were observing. During the Boren & Ramey protocol Frans acted as interactive experimenter and during the Ericsson & Simon protocol Dian acted as experimenter. Amber placed the questionnaires on SurveyMonkey. After the think aloud method, everybody filled in the three questionnaires. Finally, Dian reported on the results of the Ericsson & Simon protocol, Frans on the results of the Boren & Ramey protocol, and Amber on the results of the three questionnaires. Together we combined and adjusted these reports into one paper.

5. References

[1] Emiel Krahmer and Nicole Ummelen. Thinking about thinking aloud: A comparison of two verbal protocols for usability testing. Professional Communication, IEEE Transactions on, 47(2):105{117, 2004.

[2] Ted Boren and Judith Ramey. Thinking aloud: Reconciling theory and practice. Professional Communication, IEEE Transactions on, 43(3):261{278, 2000.

[3] K Anders Ericsson and Herbert A Simon. Verbal reports as data. Psychological review, 87(3):215, 1980.

[4] Thomas S Tullis and Jacqueline N Stetson. A comparison of questionnaires for assessing website usability. In Usability Professional Association Conference, pages 1{12, 2004}.

Appendix A

Figure 1: CSUQ

Overall Reaction to the Website		1	2	3	4	5	6	7		NA
1. Overall, I am satisfied with how easy it is to use this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
2. It was simple to use this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
3. I can effectively complete my work using this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
4. I am able to complete my work quickly using this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
5. I am able to efficiently complete my work using this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
6. I feel comfortable using this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
7. It was easy to learn to use this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
8. I believe I became productive quickly using this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
9. The website gives error messages that clearly tell me how to fix problems	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
10. Whenever I make a mistake using the website, I recover easily and quickly	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
11. The information (such as online help, on-page messages, and other documentation) provided with this website is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
12. It is easy to find the information I need	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
13. The information provided by the website is easy to understand	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
14. The information is effective in helping me complete the tasks and scenarios	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
15. The organization of information on the website pages is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
16. The interface of this website is pleasant	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
17. I like using the interface of this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
18. This website has all the functions and capabilities I expect it to have	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
19. Overall, I am satisfied with this website	strongly disagree	○	○	○	○	○	○	○	strongly agree	○

Figure 2 :SUS

	Strongly Disagree				Strongly Agree
1. I think I would like to use this website frequently.	○	○	○	○	○
2. I found the website unnecessarily complex.	○	○	○	○	○
3. I thought the website was easy to use.	○	○	○	○	○
4. I think I would need Tech Support to be able to use this website.	○	○	○	○	○
5. I found the various functions in this website were well integrated.	○	○	○	○	○
6. I thought there was too much inconsistency in this website.	○	○	○	○	○
7. I would imagine that most people would learn to use this website very quickly.	○	○	○	○	○
8. I found the website very cumbersome to use.	○	○	○	○	○
9. I felt very confident using the website.	○	○	○	○	○
10. I need to learn a lot about this website before I could effectively use it.	○	○	○	○	○