Assignments — Week 14 | Design | **Usability Testing**

Image source

In this assignment, you will design and carry out a *mini* usability test of your Module 3 deliverable, *the shopping assistant*, in three steps. In the first step, you will make some decisions on the *why*, *what*, *how*, and *whos* of the study and write a two-page test plan that reflects your decisions. Next, you will recruit two volunteers from among classmates, family, and friends who can help you with your testing, and you will execute your test plan to collect quantitative and qualitative data on the use and experience of the shopping assistant. Finally, you will analyze your data and translate your findings into design insight. Your deliverables for the assignment will be your test plan from Step 1, the data you collected in Step 2, and a report of your findings and a discussion of their design implications in Step 3.

Step 1. Design a "mini" usability test. In this step, you will make some decisions about the format and design of a brief *formative* usability test and develop a *test plan*. First, you will determine two desired outcomes for your study. You can choose from five Es we have discussed in class (*effective*, *efficient*, *engaging*, *error tolerant*, and *easy to learn)*, the three dimensions of the ISO definition of usability (*effective, efficient, satisfactory*), or related concepts or outcomes (e.g., desirability, learnability, discoverability) that best fit to what you would like to evaluate. These will serve as your desired outcomes. Next, for each outcome, you will develop *questions, tasks*, and *scenarios* that will guide your testing. Then, you will choose two metrics: one performance, one self-report. Your deliverable will be a test plan that communicates these decisions and serves as a guide for the moderator (you) to run the test. The steps in the checklist below will help you in your decision-making and writing of your test plan and the form below that will help you draft your test plan. Your test plan should not exceed two pages.

Usability Test Design Checklist

- ❏ Choose <u>two</u> intended **outcomes**, e.g., effective, efficient, engaging, error tolerant, easy to learn, usable, satisfactory, etc.
- ❏ For each outcome, formulate a **question**, e.g., "To what extent are users satisfied with the shopping assistant" or "What is the overall usability of the shopping assistant?"
- ❏ For each question, devise a **task** using your shopping assistant that can help you assess how well your design meets the outcome. The task description should capture what you expect the users to do to successfully perform the task.
- ❏ For each task, develop a **scenario** that will provide context and guidance to the user. The scenario should prompt the user to perform the task you developed.
- ❏ Choose <u>two</u> **metrics** for measurement: one performance, one self-report. Examples of performance measures include task success (e.g., number of task substeps completed), time (e.g., seconds), or errors (e.g., number of deviations from expected use). For self-report measures, you can use the SUS questionnaire or all or part of the USE questionnaire.
  - ❏ Templates for [SUS](#) and [USE](#).
- ❏ Write out your **test plan** using the form on the next page. Your plan should have three sections: (1) overview, (2) study design, and (3) test procedure. The overview section will briefly describe the context (including the "what" of the usability test, i.e., the scope of your interim or final design), the general goals for the testing, and the intended outcomes of the test. The study design section will outline your questions, tasks, and scenarios and your metrics. In test procedure, you will provide a step-by-step plan for the test in the form of a checklist.
  - ❏ You can see an example usability test plan from Barnum (2011) [here](#). Your plan will not be as detailed as this example and should be *at most* two pages.

Usability Test Plan

Overview

We have built a shopping assistant which could help the people with disabilities to search and products and make the purchases. In addition, it also has the functionality on navigation, query, and filters. The shopping assistant is building on top of the WiscShop which is a platform for student, faculty, and UW alumni to buy the school merchants and swags. The scope of the testing is still in the first phase of the final design, since we still need to improve and training our natural language part. Thus, in current situation, we may have all the features implemented, but it may face some functionality problem due to the training sets. The goal for our study is to test the effective and engaging of the program which are the important features of the products. I hope after the study, we could learn how to improve our training set in order to reduce the error rate of the program. In details we want to test the engaging on to what extent are users found that the shopping assistant could help them perform the task through voice. And for the effective, we hope to learn how accurately shopping assistant completed the tasks based on the order or the instruction from the users.

Study Design

Engaging: To what extent are users found that the shopping assistant could help them perform the task through voice.

The task for this outcome will be that users added multiple of an item, and then remove them from the cart. After that, he will also need to add multiple of another item to the cart. Since we are testing the engaging of the program, this task has many involvements including adding to cart, removing from cart, and searching or query which are the main intents of our shopping assistant. The scenarios will be when users are comparing with different products and they have a hard time to make decision. At that moment, they may need to switch back and forth by adding and removing item to or from cart.

Effective: How accurately shopping assistant completed the tasks based on the order or the instruction from the users.

The task will be that users need to find the item that under category Sweatshirts named Bucky Crew Neck Sweatshirt with the tag logo and check out that item. Because this task has many information that needs to process by the shopping assistant and there are many ways to searching or narrow down the products, it's important to show the effective side of the program. The scenario will be when user want to locate a product that they hope to buy or they heard from their friends, it may provide the system with different information since they cannot remember the detail of the products. All the information piece will help the program to locate the products which are tested by this task.

For the performance metrics, I used the task success. Because we hope to learn the engaging and effective of the program. And for the self-report metrics, I used the System Usability Scale (SUS) questionnaire which are provided by the professor in templates.

Test Procedure

The following test plan combine two tasks together to have a better measurement and have a connected scenario.

You are going to buy some products from the WiscShop website. This is your first time using the system. It has the products and merchants related to UW and you may need to find and make the purchase by performing certain task.

Scemario 1: You are a person who have a hard time to make decision. You find a Sweatshirts with tag logo first and add them to the cart. However, you then decide to buy another item called Wisconsin Football Hat. The moderator should watch the detail procedure as follows:

1. Log in to the system with your username 12345 and password 54321

2. Find the category that WiscShop provided

3. Go to the Sweatshirts category page

4. Try to find the products that are under tag logo

5. Go to the product page that are has the tag logo

6. Add 5 of that products to the cart

7. Go to the cart page

8. Remove the item that you just added

9. Find another item that is under category hat name as Wisconsin Football Hat

10. Add 10 of Wisconsin Football Hat to the cart.

Scenario 2: Right now, there are 10 Wisconsin Football Hat in your cart, you decide to check out those items. The moderator should watch the detail procedure as follows:

11. Go to the cart page

12. Review the item located inside the cart

13. Confirm the item inside the cart

Time for answering post-test questionnaire.

Step 2. Execute your test plan. In this step, you will identify <u>two</u> volunteers to help you test your shopping assistant. They can be your classmates, friends, or family members. It is acceptable to pair up with a classmate and trade taking each other's test. You can use any version of your shopping assistant as long as you have a working prototype and choose to focus on any aspect of it. You can capture performance measures during the test, e.g., by timing them, counting errors, taking notes, or by recording them and watching later. You can present self-report measures on paper or on a computer screen after they perform all scenarios. Finally, be sure to make qualitative observations and ask questions, e.g., "you seemed surprised by that response, what were you expecting," to your participant where appropriate during and/or after the study. The deliverable for this step will be your data in table and/or text format pasted below. For performance, questionnaire, and qualitative data, provide the raw numbers or text that you will later organize and analyze in Step 3.

I performed the usability test with my friend Houqi, and Xianjie. Both are the senior computer science students from UW-Madison who have certain technical skills on coding and programming. The following are the data for each from each of them.

Houqi:
Performing Metrics:
    Task of success: 9
    He failed on the task 4, 5, 9, 13
Self-Report Metrics:

SUS Questionnaire[12]

Houqi:

#	Questions			
1	I think that I would like to use this system frequently.	Strongly disagree 1 2 3 ④ 5 Strongly agree	3
2	I found the system unnecessarily complex.	Strongly disagree 1 ② 3 4 5 Strongly agree	3
3	I thought the system was easy to use.	Strongly disagree 1 2 3 4 ⑤ Strongly agree	4
4	I think that I would need the support of a technical person to be able to use this system.	Strongly disagree 1 ② 3 4 5 Strongly agree	3
5	I found the various functions in this system were well integrated.	Strongly disagree 1 2 ③ 4 5 Strongly agree	2
6	I thought there was too much inconsistency in this system.	Strongly disagree 1 ② 3 4 5 Strongly agree	3
7	I would imagine that most people would learn to use this system very quickly.	Strongly disagree 1 2 ③ 4 5 Strongly agree	2
8	I found the system very cumbersome to use.	Strongly disagree 1 ② 3 4 5 Strongly agree	3
9	I felt very confident using the system.	Strongly disagree 1 2 3 ④ 5 Strongly agree	3
10	I needed to learn a lot of things before I could get going with this system.	Strongly disagree 1 2 ③ 4 5 Strongly agree	2

28

Total   70

[1] Based on Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.
[2] Scoring Guide for SUS
1. For odd items, subtract one from the user response. For even-numbered items, subtract the user response from 5.
2. The subtractions scales all values from 0 to 4 where four is the most positive response.
3. Add up the converted responses for each user and multiply that total by 2.5, which will convert the range of possible values from 0 to 100 instead of from 0 to 40.

Qualitative Metrics:
Even Houqi has the background in Computer Sciences, he had a hard time using the system for certain stage. At the stage 4, when he performed the procedure, he did not know what to say to the shopping assistant because he did not know what kind of the tag available to users. Thus, he was stuck on that stage. On the stage 5, he said go to the third to the left to the shopping assistant. Since the assistant can only understand the order or the name of the products, it did not give him a response on it. For the task 9, even we provide him with the product name and category information, the user still could not find the product. In addition, in the cart he said check out, the page did not go to the review page. At the end, I asked two questions to him, one is about whether he need any time for learning the products. He said it make take time to familiar with the products. In addition, I asked whether there's any result surprised him. He said sometimes the natural

language part needs to be improved since the system cannot understand the graph location on the screen.

Xianjie
Performing Metrics:
    Task of success: 10
    He failed on the task: 3, 9, 13
Self-Report Metrics:

SUS Questionnaire[12] Xianjie

#	Questions			
1	I think that I would like to use this system frequently.	Strongly disagree 1 2 3 4 (5) Strongly agree	4
2	I found the system unnecessarily complex.	Strongly disagree 1 (2) 3 4 5 Strongly agree	3
3	I thought the system was easy to use.	Strongly disagree 1 2 (3) 4 5 Strongly agree	2
4	I think that I would need the support of a technical person to be able to use this system.	Strongly disagree (1) 2 3 4 5 Strongly agree	4
5	I found the various functions in this system were well integrated.	Strongly disagree 1 2 (3) 4 5 Strongly agree	2
6	I thought there was too much inconsistency in this system.	Strongly disagree (1) 2 3 4 5 Strongly agree	4
7	I would imagine that most people would learn to use this system very quickly.	Strongly disagree 1 2 (3) 4 5 Strongly agree	2
8	I found the system very cumbersome to use.	Strongly disagree 1 (2) 3 4 5 Strongly agree	3
9	I felt very confident using the system.	Strongly disagree 1 2 3 (4) 5 Strongly agree	3
10	I needed to learn a lot of things before I could get going with this system.	Strongly disagree 1 (2) 3 4 5 Strongly agree	3

30

Total 75

[1] Based on Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.
[2] Scoring Guide for SUS
  1. For odd items, subtract one from the user response.For even-numbered items, subtract the user response from 5.
  2. The subtractions scales all values from 0 to 4 where four is the most positive response.
  3. Add up the converted responses for each user and multiply that total by 2.5, which will convert the range of possible values from 0 to 100 instead of from 0 to 40.

Qualitative Metrics:
Xianjie as Houqi has the background in computer science. However, he still failed two tasks. When he performed the task 3, the system did not go to the category page. Instead, it tells the user of all the category available to them. I believe this could be an error from the training which the system did not recognize the difference between showing category and go to the category page. Moreover, Xianjie also failed the test on task 9, since he did not find the way to get into the product page. And he still believes we need to find the product by visiting the category first, and then choose the product. But actually, you can visit the product page by just calling the name. In addition, he has the same problem as Houqi on figuring out the review page. I asked the same question that I asked Houqi about whether he need any time for learning the products. Xianjie said he need sometime

on figuring out the structure of the products, such as category, tags, and name. However, after that it may require less effort on using the system. In addition, regarding the things surprised him. He said he thought the searching process will be when a user says a general name or type of a product, the system will show him the best match. But for our assignment, it may be a little bit harder to implement.

Step 3. Analyze and report your findings. In this step, you will clean, consolidate, and analyze your results and translate them into design insight. For your quantitative data, calculate the average values from your metrics and report the averages. For self-report data, if you used SUS, follow the scoring method included in the template and give your shopping assistant a grade (e.g., "D") and level of acceptability (e.g., "high marginal") using the guide below.[1] If you used a subscale of USE, such as "ease of use," average out the scores for all items to arrive at a single value and average out the values for both of your test participants. For qualitative data, categorize your notes and observations into a minimum of <u>two</u> high-level findings. If the quantitative data or the qualitative comments from your two participants vary significantly, you can also comment on these differing views. Report your findings in narrative form and end your report with high-level design insight and recommendations for how your shopping assistant might be improved. Your report should not exceed a page.

---
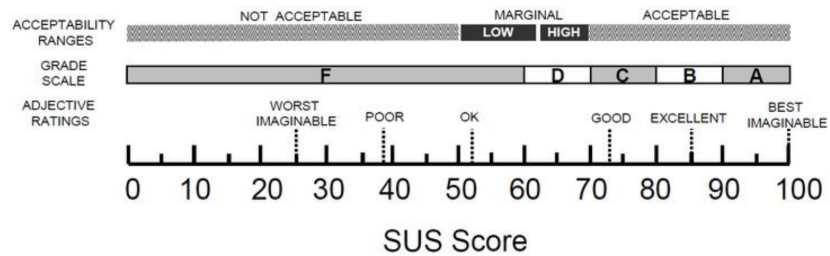
[1] Based on Brooke, J. (2013). SUS: a retrospective. *Journal of usability studies*, 8(2), 29-40.

Usability Findings

Quantitative Summary

Performing Metrics:
Two volunteers earned a score of 9 and 10 respectively. On average the score for performing metrics is 9.5.
Self-Report Metrics:
Two volunteers earned a score of 75 and 70 respectively. On average the score is 72.5 for the self-reported metrics. This is a score range for the high D and low C which is on the middle of the acceptability range.
The quantitative summary shows that the program is acceptable; however, there's a large room that can be improved.

Qualitative Summary

1. Based on the findings from the quantitative metrics, I think it also shows that the acceptability is only in the middle range. This could mean a lot to us which can also reflect from the answer that user provided to us. For our program, it has the context, which sometimes the user needs to speak certain phase in order to complete the tasks. This is also showing in the response from users that they may need more time on learning using the shopping assistant.

2. The product finding seems have a problem. If user speaks about the visit certain category, the program report all the category available showing on the page. This misinterpretation could also happen in another phase which means we may need to train our phase set.

3. The searching has a problem since it may not provide the best match results. We cannot return a list of best matches based on what user said. However, due to the limitation of the program, it may hard for us to manipulate or improve it.

Conclusions

1. (Related to Finding 1) Adding detail response for each of the phase or context. Since some of the tasks we need to perform them in order, we need to inform user about what they need to do next. Thus, providing them a detail response could be a great way to solve this kind of the problem. We could end with a question sentence such as do you want to do something? This could help to mitigate the confusion and require less time on learning how to use the system.

2. (Related to Finding 2) Adding more training phase into the Dialogflow. Since sometimes the system could misinterpret the meaning of the user's instruction, it shows that we do not have a great amount of training phase. Thus, adding more to it may help to relieve the problem that user faced before.