CS-639 Building User Interfaces, Fall 2019, Professor Mutlu

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 two 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 two **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

The goal of the usability test is to collect feedback from the users visit and shop on WiscShop website via a voice assistant. This test is going to find out which designs of the voice assistant will make it difficult for the users to visit and shop on the WiscShop website.

Outcome 1: Efficient – Do users find the product he wants more quickly.

Question: What is the time the users find a desired product from home page?

Outcome 2: Error Tolerant – Can the voice assistant tolerate more user mis-inputs.

Question: How many times the voice assistant cannot understand the user with a slice of time or a flow of operations?

Study Design

Outcome 1: Efficient – Do users find the product he wants more quickly.

Task: Given a particular product, let a new user interact with the voice shop assistant and record the time he reaches the goal page from home page.

Scenario: The user is a new user to the WiscShop website. He wants to browse the website, learning the reviews about the ratings of the products and then buy some products for exercising. To be specific, go to the details page of 'Bucky Badger Keychain' and add it to cart after reading its average ratings.

Outcome 2: Error Tolerant – Can the voice assistant tolerate more user mis-inputs.

Task: Give a list of descriptions of flow of tasks, (e.g. check out from cart), let the new user perform the flow of tasks and record how many times the voice assistant cannot understand the user.

Scenario: The user has already added the products he wants to the shopping cart and wants to check the shopping cart. To be specific, ask to user reduce the number existing product in cart to one and confirm the cart then check out.

Metrics:

Time-on-task: Record how long users in scenarios will take complete the tasks.

Unit: seconds.

SUS Questionnaire: calculate the total points, 60 for pass.

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

Test Procedure

Participants: The users must never use the WiscShop shop assisstant before.

The moderator will welcome the participant and have him/her sign the video consent form.

Next, the moderator will explain the facilities and ask the participant to think out loud.

Scenario 1: 1) Start from home page,

2) Go to the page of ‘plushes’,

3) Go to details of “Bucky Badger Keychain”,

4) Ask for ratings of the keychain,

5) Add one keychain to cart

Scenario 2: 1) Start from a cart with one “Wisconsin Sweatpants” and two “Wisconsin Running Shorts”,

2) Change the number if “Wisconsin Running Shorts” to one,

3) Confirm the cart, cart should be one “Wisconsin Sweatpants” and one “Wisconsin Running Shorts”,

4) Place the order

Step 2. Execute your test plan. In this step, you will identify two 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.

Participant 1: Zhanhe Lyu,

friend,

have experience in using voice assistants, but never used WiscShop voice assistant.

Results:

	Time(sec)
Scenario 1	25
Scenario 2	26

Question	1	2	3	4	5	6	7	8	9	10
Points	2	3	4	2	3	3	3	3	4	1

total: 56, not pass

Other feedback: The voice assistant cannot tolerate sentence without space.

Participant 2: Bojun Xu,

non-English native speaker,

have experience in using voice assistants, but never used WiscShop voice assistant.

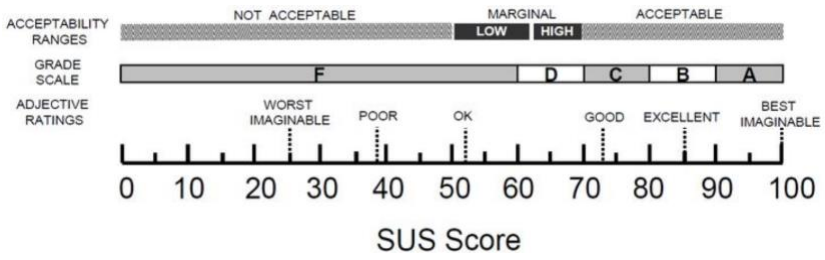
Results:

	Time(sec)
Scenario 1	27
Scenario 2	20

Question	1	2	3	4	5	6	7	8	9	10
Points	2	4	3	1	4	2	4	1	4	1

Other feedback: The voice assistant cannot recognize a as one.

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 two 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

Scenario 1:

Average completion time: 26s

Scenario 2:

Average completion time: 23s

Average SUS:

54 points, in low-marginal

Conclusions

In my usability testing, the first scenario is focused on how fast a user can find the product he wants. In this case, the feedback from my volunteers is especially crucial because new users may not be good at using the voice assistant. Both my volunteers report they don't know where to start (i.e. there is lacking in hints). The second scenario is basically focused on if the user's input can be effectively recognized. In other words, although there may be more hints than the first scenario, the second scenario requires more tolerance on the user's ambiguous and wrong inputs. From my volunteer, he reports the assistant cannot recognize the case when space does not exist.

To make improvements, I have these ideas:

1) Provide more hints from the assistant:
   a. If the user is new to the assistant, give the user some sample inputs for the user to learn how to input.
   b. Give user options. (e.g. if the user wants to buy a tee-shirt, the assistant provides options when he goes to a category page: "Which color would you like? red, black or white".

2) Increase error tolerance:
   a. Provide more training instances for the machine learning model, the more the better.
   b. Check existing training instances to optimize the training instances.