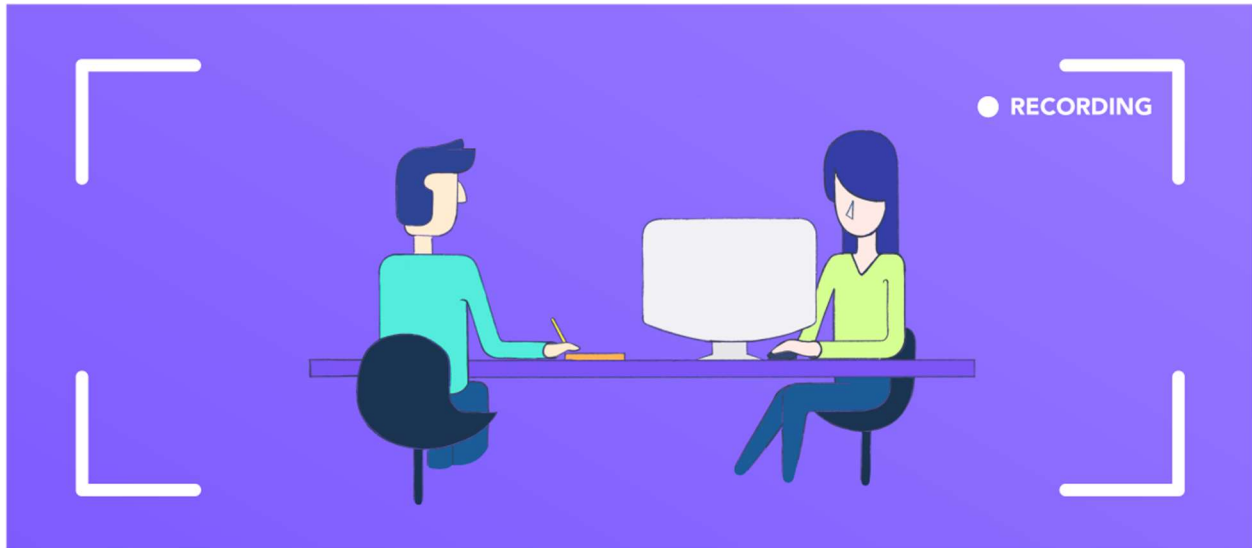


## Assignments — Week 14 | Design | **Usability Testing**



[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 effectiveness of just using speech to browse a website

Can the user effectively navigate the page?

Does the agent recognize vague commands?

Can the agent recover from vague commands?

## Study Design

User effective navigation:

Use speech to login, go to specific sections, and apply filters for the products

Vague commands:

User doesn't specify things like username, password, and whatever page they are navigating to.

Recover:

Would the agent ask for clarification and re-prompt the user or just end the conversation?

Measure: Number of errors (agent not behaving as expected).

SUS questionnaire

## Test Procedure

<Describe the procedure you will follow in the form of a checklist for the study moderator>

<b>1. Just say Login</b>	<b>9. Rigorous testing, repeat 5-8</b>
<b>2. Provide username</b>	<b>10. Navigate to home</b>
<b>3. Provide password</b>	<b>11. End</b>
<b>4. Restart and specify both username and password in one command</b>	
<b>5. Ask the agent to show all items</b>	
<b>6. Agent should ask which category and navigate into the page</b>	
<b>7. Do the same task again but specifying the (another)category.</b>	
<b>8. Ask the agent to narrow the search</b>	

## *Overview*

The efficiency when the user already knows what to buy.

Can the user add item to cart with minimal steps and calls?

Can the user add item to cart with the product name?

Can user remove from cart with product name?

## *Study Design*

Add using general (but specific) term:

Say something like “add the blue hat”

Add using specific product name: Jump Around Shirt etc.

Remove using specific product name: Same as above

Measure: Time taken

SUS questionnaire

## *Test Procedure*

1. **Login with username and password**
2. **Add a white cap into cart**
3. **Add the Jump Around Shirt into cart**
4. **Remove the Jump Around Shirt from cart**
5. **End test**

**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.

---

Effectiveness test

Login : Error. The word Login is treated as a password

Username Password entry : Pass

Show items w/o specify category: Error. Items was treated as the required word/category

Show items w/ category : Pass

Tags : Pass

Rigorous : Same results

Home Navigation : Pass

**4/7 Pass**

I was confused when DialogFlow did not work as expected and took words that weren't intended. I was expecting the agent to ask for username and password, but it only asks for username and took the word as password. The same behavior happened for category page.

---

#	Questions									
1	I think that I would like to use this system frequently.	Strongly disagree	1	2	3	4	5	Strongly agree		
2	I found the system unnecessarily complex.	Strongly disagree	1	2	3	4	5	Strongly agree		
3	I thought the system was easy to use.	Strongly disagree	1	2	3	4	5	Strongly agree		
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		
5	I found the various functions in this system were well integrated.	Strongly disagree	1	2	3	4	5	Strongly agree		
6	I thought there was too much inconsistency in this system.	Strongly disagree	1	2	3	4	5	Strongly agree		
7	I would imagine that most people would learn to use this system very quickly.	Strongly disagree	1	2	3	4	5	Strongly agree		
8	I found the system very cumbersome to use.	Strongly disagree	1	2	3	4	5	Strongly agree		
9	I felt very confident using the system.	Strongly disagree	1	2	3	4	5	Strongly agree		

10	I needed to learn a lot of things before I could get going with this system.	<i>Strongly disagree</i>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<i>Strongly agree</i>
----	--	--------------------------	----------	----------	----------	----------	----------	-----------------------

---

## Efficiency Test

Adding blue hat to cart : 16 seconds

Adding “Jump Around Shirt” to cart : 14 seconds

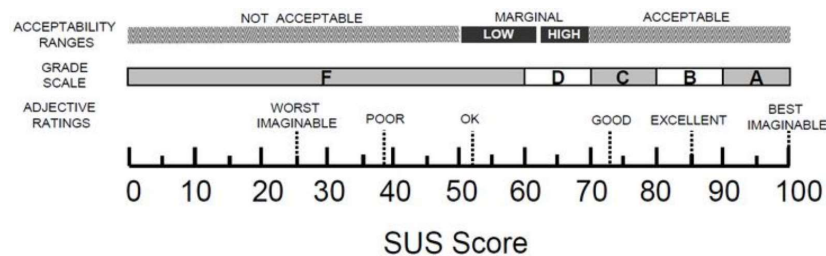
Remove “Jump Around Shirt” from cart : 25 seconds

The test results were efficient, where the agent could understand, process commands and the filtering and functions in the webhook is done in a good amount of time. Remove takes longer because of delete confirmation. The only problem is that for the blue hat test, it should have taken longer since I did not handle the case where the agent did not recommend the closest item when the search failed, hence the short time taken. Time recorded does not include time taken to log in.

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

**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.<sup>1</sup> 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.



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

# Usability Findings

## *Quantitative Summary*

Number of error test passed is 4/7, 57.14% score with scale of 100

SUS score: 67.5. Marginal high

## *Qualitative Summary*

Agent does not recognize vague commands properly, i.e. Login

Search function working properly, (applying tags, navigating to pages)

## *Conclusions*

Since the problem is the entities is incorrectly selected, I think it should be a problem with Dialog Flow. I would have to investigate the training data of the intents or the entity settings of the Dialog Flow, worst case scenario might have to re-write the structure of the intents.

---

## *Quantitative Summary*

Total time taken not including login is: 55 seconds

SUS score: 70 Acceptable

## *Qualitative Summary*

Did not handle case where can't find the specified item.

Adding by item name requires user to have prior knowledge

## *Conclusions*

Right now, I think I should make more error messages instead of one default one when the fetch results in an empty array. This is so that it would provide feedback to the user so they may know how to better phrase their query.