CS-639 Building User Interfaces, Fall 2020, Professor Mutlu

# Dialogflow (3 Points)

## Usability Testing

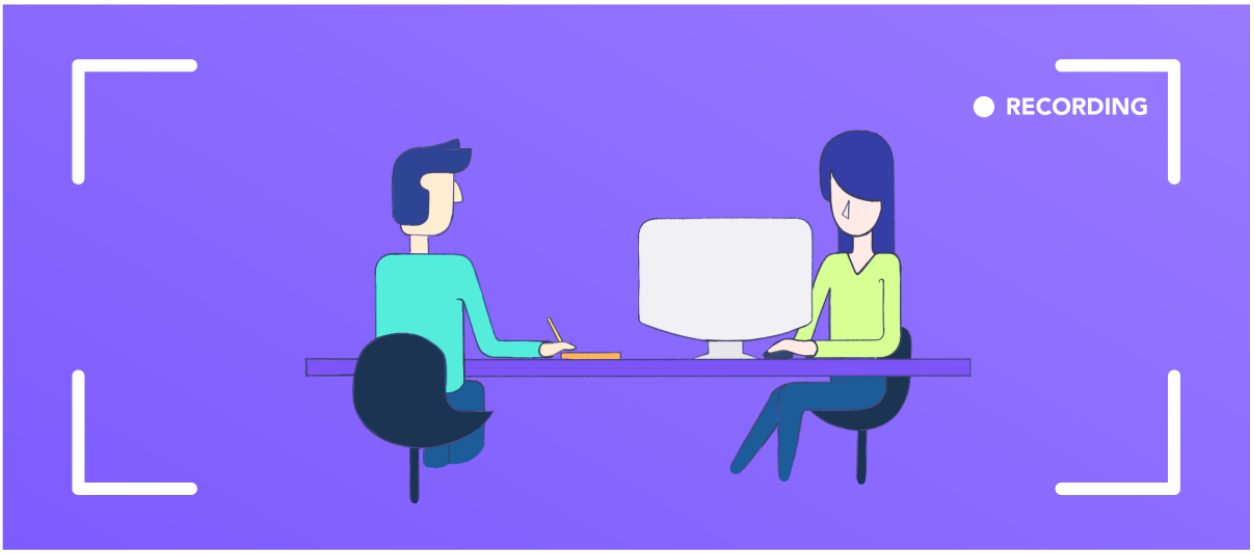



Image source

In this assignment, you will design and carry out a *mini* usability test of your Module 3 deliverable, *the shopping assistant*, in three parts:

**Part 1—Designing A "Mini" Usability Test (0.8 Point):** In the first part, 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.

**Part 2—Executing Test Plan (1.4 Points):** 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, over videoconferencing, to collect quantitative and qualitative data on the use and experience of the shopping assistant.

**Part 3—Analyzing & Reporting Findings (0.8 Point):** Finally, you will analyze your data and translate your findings into design insight.

## Submission Details

Your deliverables for the assignment will be your test plan from Part 1, the data you collected in Part 2, and a report of your findings and a discussion of their design implications in Part 3, all as a single PDF document submitted to Canvas.

*Note:* Your assignment will be graded on the contents of this report and not the usability of your system. If you find that your agent is hard-to-use or unintuitive, you can be honest with your outcomes.

## **Part 1:** Designing A "Mini" Usability Test (0.8 Point)

In this part, 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. Your study should be in the form of a remote *moderated* usability test conducted over videoconferencing, e.g., Zoom. 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[1]

## *Overview*

The intended outcomes of this usability test is to evaluate the efficiency and satisfaction dimensions of the shopping assistant created with Dialog Flow for WiscShop web application to users. The questions I will focus on addressing with this test are: 1. Overall, how satisfied are the users with their experience interacting with the shopping agent to select and purchase items? And 2. How efficient is the shopping assistant in aiding users finding information compared to just using the website?

The test will be conducted remotely with two volunteers. The scope of the test will include navigation from page to page feature(i.e. From home page to cart, or a category page back to home page), query about the cart, review and confirm the purchase with all the items in the cart, adding and removing products to cart, and filtering items by task. Also, the scope may include queries about categories, tags, or products(if the user found the need to use these features, but it is not the main tested features). This basically included the entirety of the shopping assistant, since it only had limited functionalities.

## *Study Design*

1.Overall, how satisfied are the users with their experience interacting with the shopping agent to select and purchase items?

I am using the following task to measure the satisfaction of users with the shopping assistant: Find and add 2 Wisconsin Running Shorts, 1 Wisconsin Legging, and 3 W Cloud Pillow to Cart, and purchase the items. The scenario is as follows: User is buying Christmas presents for their 3 siblings, the user decided to buy a W Cloud Pillowfor each of them, in addition to a pair of running shorts to two of their siblings that loves running and a pair of leggings for the other sibling who loves to collect leggings.

2.How efficient is the shopping assistant in aiding users finding information compared to just using the website?

I am using two tasks to compare the efficiency of the assistant vs. the website on its own.

First, the user will navigate to the “bottoms category” and eliminate items displayed to the ones with “grey” and “logo” tags. A scenario would be that the user is a freshman at Madison and wants to get some school gears. The user would like a pair of pants or shorts. Their favorite color is grey and they would like a logo on the product so that they could show off their school spirit.

Second, the user will record the number of products in each category in a cart full of different products. The scenario is that the user is buying products for their company, and the company can only reimburse the cost of the purchase when they are given a list of categories of products purchased. For this task, the content of the cart will be different for each trial per user, this is to eliminate the possibility that the user was simply going off of memory for the second trial.

The metrics of measurement that will be used are the following:

- Performance measurement: recording time needed for each task on each medium(WiscShop website or agent) and recording number of clicks or commands typed by the user. These will be

[1] Or use the Usability Test Plan template

quantitative data. These quantitative data will mostly be used to evaluate the efficiency of the shopping assistant.

- Self-Report: I will use the "Ease of Use" and "Satisfaction" sections in the USE questionnaire along with a few interview questions to ask the user about how engaged they were when using the assistant. The interview will start with: Tell me about the things you liked or disliked about using the agent. Then, I will follow up on their specific answers. The reason I am also using the "Ease of Use" section from the USE questionnaire is that people tend to find the items that are easy to use to be more enjoyable and pleasant to use. The questionnaire from USE will be quantitative data with an average of the final scores for each section, and the interview responses will be qualitative data.
- There will also be other qualitative data collected throughout the testing process, where observations of the users are recorded.

## *Test Procedure*

- ❑ The test will begin with an introduction of the website and shopping assistant. The moderator will also explain the general usage of the shopping assistant and encourage users to think out loud.
- ❑ The moderator will time 1 minute for the user to explore the shopping assistant and the website to familiarize with the set up.(This is because the study is not testing for learnability or ease to learn. Also, 1 minute is not too long of a time period for the user to understand every aspect of the assistant, but just to feel comfortable using it, which is necessary because the following tasks will be timed.)
- ❑ **First Scenario**: You are buying Christmas presents for your 3 siblings. You decided to buy a W Cloud Pillow for each of them, in addition to a pair of running shorts to two of your siblings that loves running and a pair of leggings for the other sibling who loves to collect leggings. Use the shopping assistant to achieve this purchase.
- ❑ **Second Scenario**: You are a freshman at Madison and would like to get some school gears. Your favorite color is grey and you want a logo on the product so that you can show off your school spirit, also you'd like a pair of pants or shorts. Find the item(s) that fulfills your needs and show them on the screen. There is no need to choose one.
  (An order of using the website or the shopping assistant will be chosen randomly)
  You will perform this task twice, once with each medium.
- ❑ **Third Scenario**: You are buying products for your company, and your company can only reimburse the cost of the purchase when they are given a list of categories of products purchased. As a result, you will need to make a simple table that records the number of products in your cart from each category of WiscShop.
  (An order of using the website or the shopping assistant will be chosen randomly)
  You will perform this task twice, once with each medium.( the cart will be refreshed to contain new items for each trial)
- ❑ The moderator will provide the user with a set of questionnaires to fill out.
- ❑ The moderator will interview the user on the things they liked or disliked about using the shopping assistant.

## **Part 2:** Executing Test Plan (1.4 Points)

In this part, you will identify two volunteers to help you test your shopping assistant over videoconferencing, e.g., Zoom, Microsoft Teams, Webex, etc., choosing a system that allows remote control of your computer (see documentation on conducting remote sessions where you give control of your computer to your partner for Zoom, Teams, Webex). 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 part 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 Part 3.

---

Volunteer 1: user familiar with the agent
(Task in order that it is performed)

| Task | Time (min:sec) | Number of clicks/ commands | Observations |
|---|---|---|---|
| Task 1 (Agent) | 2:51 | N/A | - User added 2 running shorts to cart twice(4 in total) because the agent response wasn't clear(no number was mentioned.<br>- User attempted to delete running shorts from cart using the command "shorts", but it did not work because shorts is not an entity of products, but a tag. |
| Task 2 (Agent) | 1:51 | 9 | - user typed in a command and was interpreted by agent as a log in command.<br>- user explored a little before being able to add two tags(did them separately and realized that it clears at every command) |
| Task 2 (Website) | 0:15 | 3 | - The user went through the task very quickly. |
| Task 3 (Agent) | 0:26 | 1 | - User was familiar with the general idea of the shopping assistant, and was able to fetch the information needed directly from the agent. |
| Task 3 | 0:55 | 3 | - user scroll across the cart page several times |

| | | | |
|---|---|---|---|
| (Website) | | | - The cart did not display all items that had been added to cart.(In the GUI)<br>- The category of leggings wasn't clear to the user, and the user misinterpreted it to be of category bottoms. |

| **Ease of Use** | | | | | | | |
|---|---|---|---|---|---|---|---|
| It is easy to use. | *Strongly disagree* | 1 | 2 | 3 | **4** | 5 | *Strongly agree* |
| It is simple to use. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| It is user friendly. | *Strongly disagree* | 1 | 2 | 3 | **4** | 5 | *Strongly agree* |
| It requires the fewest steps possible to accomplish what I want to do with it. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| It is flexible. | *Strongly disagree* | 1 | 2 | **3** | 4 | 5 | *Strongly agree* |
| Using it is effortless. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| I can use it without written instructions. | *Strongly disagree* | 1 | 2 | **3** | 4 | 5 | *Strongly agree* |
| I don't notice any inconsistencies as I use it. | *Strongly disagree* | 1 | 2 | 3 | **4** | 5 | *Strongly agree* |
| Both occasional and regular users would like it. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| I can recover from mistakes quickly and easily. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| I can use it successfully every time. | *Strongly disagree* | 1 | 2 | 3 | **4** | 5 | *Strongly agree* |
| **Satisfaction** | | | | | | | |
| I am satisfied with it. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| I would recommend it to a friend. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| It is fun to use. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| It works the way I want it to work. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |
| It is wonderful. | *Strongly disagree* | 1 | 2 | 3 | **4** | 5 | *Strongly agree* |
| I feel I need to have it. | *Strongly disagree* | 1 | 2 | **3** | 4 | 5 | *Strongly agree* |
| It is pleasant to use. | *Strongly disagree* | 1 | 2 | 3 | 4 | **5** | *Strongly agree* |

What he liked:

- High tolerance - smart agent compare to other shopping agents
- Easy to navigate, also can see all categories in the homepage
- add/remove from cart, have access to the cart
- Simple graphic.
- Polite as expected(from shopping assistants)

What he did not liked:

- Difficult to learn, did not know some commands.
- Cart GUI did not display everything

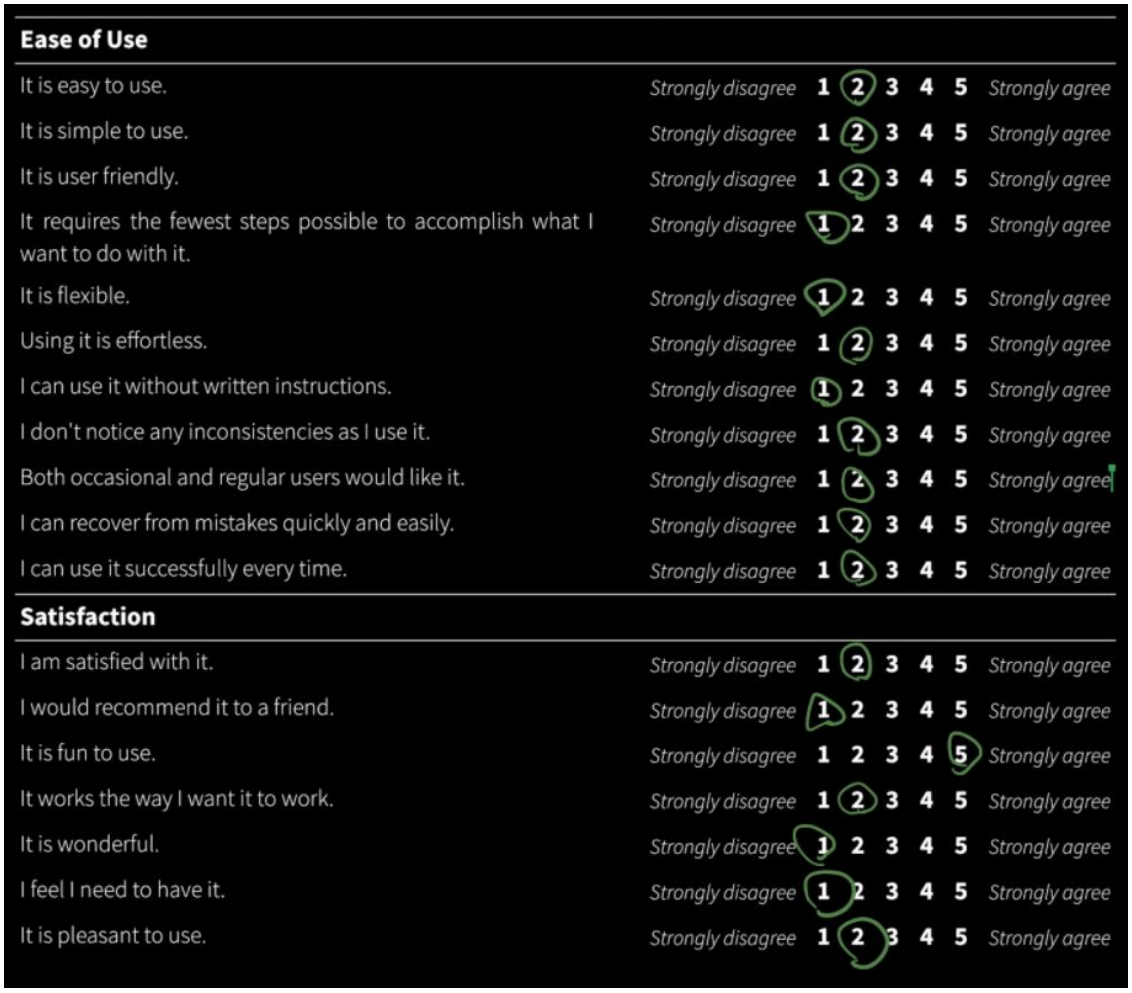
(my note: Maybe some suggestions could be given by the agent, similar to where many websites linked to messenger and have suggested questions or options to choose from)

---

Volunteer 2: first time user

(Task in order that it is performed)

| Task | Time (min:sec) | Number of clicks/ commands | Observations |
|---|---|---|---|
| Task 1 (Agent) | 1:57 | N/A | - When the user attempted to add an item to cart, the agent led her to the products page.<br>- the user did not realize that she could add 2 items to the cart at once at first, but realized later. |
| Task 2 (Agent) | 1:41 | 11 | - User mistakened leggings for bottoms<br>- User had to test adding tag command to find out that it clears after each command.<br>- User initially searched for the tags in the homepage, which did not work. |
| Task 2 (Website) | 0:10 | 3 | - user went through the task very quickly. |
| Task 3 (Website) | 1:33 | 6 | - user went into different product's page trying to figure out their category by reading the descriptions |
| Task 3 (Agent) | 0:12 | 1 | - user learned from previously using similar commands and was able to figure out the simple command to achieve this task |



| Ease of Use | | | | | | | |
|---|---|---|---|---|---|---|---|
| It is easy to use. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It is simple to use. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It is user friendly. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It requires the fewest steps possible to accomplish what I want to do with it. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It is flexible. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| Using it is effortless. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| I can use it without written instructions. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| I don't notice any inconsistencies as I use it. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| Both occasional and regular users would like it. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| I can recover from mistakes quickly and easily. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| I can use it successfully every time. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| **Satisfaction** | | | | | | | |
| I am satisfied with it. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| I would recommend it to a friend. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It is fun to use. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It works the way I want it to work. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It is wonderful. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| I feel I need to have it. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |
| It is pleasant to use. | Strongly disagree | 1 | 2 | 3 | 4 | 5 | Strongly agree |

What she liked:

- Agent is friendly
- The navigation to different categories is easy
- The agent understood context

What she did not like:

- The agent is not smart enough, meaning that it did not always understand her commands even when it was clear

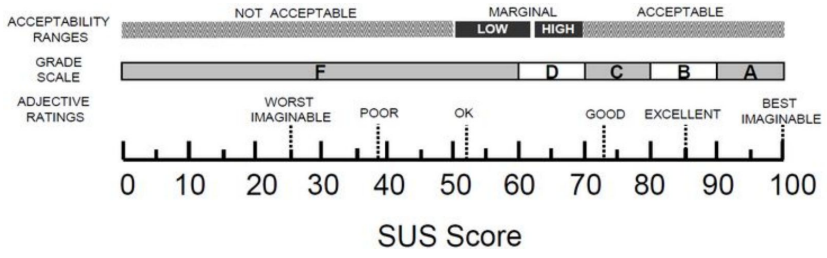
(My note: Maybe she is looking for something like siri, that can adapt to anything user said. However, this could be hard to achieve as the agent mainly relied on training phrases to understand, and I wasn't able to get very conclusive with the training phrases.)

## **Part 3:** Analyzing & Reporting Findings (0.8 Point)

In this part, 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.[2] If you used a subscale of USE, such as "ease of use," average out the scores for all items to arrive

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

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.



# Usability Findings

## *Quantitative Summary*

| | Volunteer 1 | Volunteer 2 |
|---|---|---|
| Task 1 Time** | **2:51** | **1:57** |
| USE Average Score | **4.39** | **1.72** |
| Task 2 Time | Agent time is 1:36 **slower** than the website | Agent time is 1 :31 **slower** than the website |
| Task 2 Clicks/Commands* | **6** | **8** |
| Task 3 Time | Agent time is 0:29 **faster** than the website. | Agent time is 1:21 **faster** than the website. |
| Task 3 Clicks/Commands* | **-2** | **-5** |

*Calculated difference using (agentData - websiteData) **Times are all in min:sec format

Comparing the time it took for volunteers to complete task 2, we could see a pattern of users performing the task much faster with the website setting. Also, there are significantly more commands used in the agents than clicks used in the website. In this case, the website has been shown to be more efficient in allowing users to filter items by tags. As for task 3, we received a different outcome compared to task 2, with the time it took with the agent being faster than the website for both volunteers. There are less commands in agents than clicks in the website as well. Though the differences are not as significant as the ones from task 2, it is fair to conclude that the agent is more efficient than the website in this case.

The USE scores shows that the first user is rather satisfied with the shopping assistant while the second one is not satisfied.

## *Qualitative Summary*

The first user reported that some suggestions would help make the interaction with the agent easier and more engaging, and the second user reported that the agent is not smart enough to understand commands that only seemed natural to humans. Thus, a general problem is that the commands that the agent understands are too limited, causing there to be dissatisfaction for users.

Both volunteers mentioned that the agent had a good attitude, with volunteer 1 commending the agent for being polite and volunteer 2 liking the friendliness of the agent. With a polite and friendly agent, the users are more likely to be satisfied with this shopping assistant.

## *Conclusions*

From conducting the test with 2 volunteers, my conclusion is that the users would become more satisfied with the agent if it is more well-trained to understand the different commands and to show suggestions to support the learning curves of first time users. Also, the agent is more efficient in supporting the users with higher level tasks, such as summarizing items or providing reports; however, when it comes to lower level tasks the website itself may be more efficient.

Overall, many upgrades could be done to increase its efficiency and user satisfaction. This includes adding more intents that allow for the capabilities to perform tasks with one step, more training phrases to yield better understanding of user commands, and agents providing suggestions allowing the user to have a smaller learning curve.