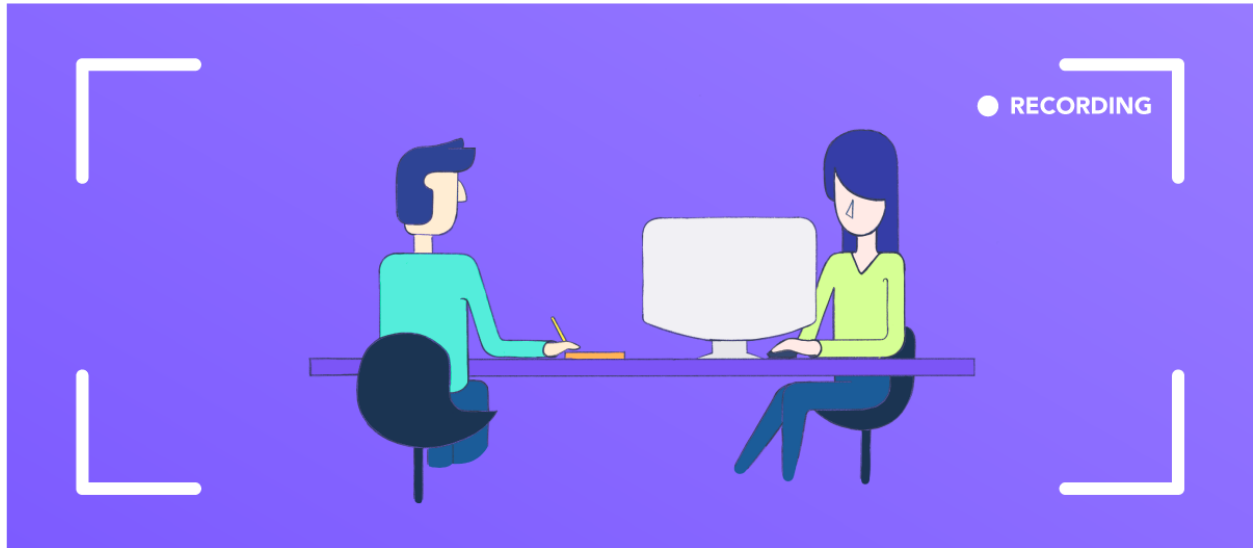


# Dialogflow $\gamma$ (3 Points)

## Usability Testing



[Image source](#)

In this assignment, you will design and carry out a *mini* usability test of your Dialogflow 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<sup>1</sup>

## *Overview*

<Describe the general context and scope of your study>

The general context for this test is to collect user feedback and data on how they use the WiscShop agent to perform any actions in the shop. Then, the collected information will be analyzed to improve the usability of the shopping assistant. The scope of the study focuses on the efficiency and error tolerance of the shop agent.

<Outline study goals, including intended outcomes, in the form of questions>

The first intended outcome of the study is “efficient.” The main question is “how efficiently and quickly the user can navigate, browse, obtain information, and purchase using the shopping assistant?” The second intended outcome of the study is “error tolerant.” The main question for this outcome is “does the shopping assistant prevent errors and help the users recover and learn from their mistakes?”

## *Study Design*

<For each question, describe the task and the scenario>

For the first question of efficiency, one task can be that the user is trying to find a few products in the shop, add them to the cart, and check out. Expectedly, the user should be able to find their desired products using the filtering options including the categories and tags. Then, they should be able to add these products to the cart and visualize them in the cart as well. Lastly, they should be able to confirm their cart and place an order using the shopping assistant. The scenario to provide context and guidance for the users can be providing appropriate feedback for each action in the series. For example, if the user is trying to add products (e.g. jump around shirts) to the cart, the shopping assistant should inform the user of if the addition is successful, how many of the products are added, and perhaps what they should do next.

For the second question of error tolerance, one task can be that the user is trying to find, navigate to, or buy products that do not exist in the shop. Expectedly, the user should be able to navigate the shop and try to find their desired item. After using the filtering options including the categories and tags, they cannot find their item. This is an error because the desired item is not included in any of the categories or tags in the shop, so the shopping assistant may struggle to respond in some cases. The scenario to provide context and guidance for the users is to first inform the user of the item not found in the shop. Then, the shopping agent should provide the user with items that are already in the shop and perhaps suggest related items to the user’s search. For example, if the user is trying to find a Michigan hat in the

---

<sup>1</sup> Or use the [Usability Test Plan template](#)

shop, which is not offered, then the agent should be able to inform the user of the unavailability and perhaps recommend the user buying a badger hat.

<Describe your measures>

For the first intended outcome, the performance measure is the time that it takes for the user to complete the task. Time is a significant factor when evaluating efficiency. Then, the self-report measure is the SUS questionnaire.

For the second intended outcome, the performance measure is the ratio of the user's frustration, which is the number of times when the user feels frustrated from the error divided by the total number of errors the user experiences using the app. If the ratio is high, it means that the user did not receive enough help or guidance from the agent when an error occurred. If the ratio is low, it means that the user was able to recover from most of the errors when they occurred. Then, the self-report measure is also the SUS questionnaire.

### *Test Procedure*

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

- **Overview/Briefing:** The participants will be briefly introduced to the WiscShop interface and the shopping assistant. They should not be allowed to use the mouse to navigate the web shop. since the usability test is for the shopping assistant. The moderator should ensure that the participants do not have any questions before starting the test.
- **Scenario/Task 1:** The participant wants to find two products that exist in the shop: Wisconsin Football Hat and Jump Around Shirt. The quantities for these two products are 1 and 3 accordingly. Then, the participant wants to add these products to the cart and check out (confirm and place an order). The log in information should be provided by the moderator since it is not the focus of this task: the username is oliver and the password is 12345.
- **Data collection:** During the previous process, the moderator should record the time for the user to complete the task. After the task, the moderator will ask the participant to complete the SUS questionnaire for scenario/task 1 and ask them their experience of using the app and the agent so far.
- **Scenario/Task 2:** The participant wants to find a few products that do not exist in the shop: Michigan Football Hat and Biff Wolverine Plush. They may use any filtering tools provided in the shop. They should try as hard as they can to find these products based on the information provided by the agent.
- **Data collection:** During the previous process, the moderator should record the total number of times when the user encounters any errors and also the number of times when they become frustrated and helpless as they try to recover from the errors. Then, the moderator will ask the participant to complete the SUS questionnaire for scenario/task 2 and ask them their experience of using the app and the agent when errors happen.



## 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.

---

<your data>

### Scenario/Task 1

	Time to complete task 1 (seconds)	SUS Questionnaire 1
Participant A	3 minutes 51 seconds	 SUS Questionnaire A1
Participant B	4 minutes 18 seconds	 SUS Questionnaire B1

- Participant A had some trouble navigating around the shop. Once he was informed that he could use the categories to navigate, the process became a lot easier.
- Participant A searched for a category that does not exist in the shop. The agent responded with no such category.
- Participant A had some trouble navigating to the tees page. There seemed to be a misunderstanding of providing more information on tees and navigating to the tees page for the agent.
- When adding the items, Participant A added the Wisconsin Football Hat, which has a default count of 1. Then, when adding 3 Jump Around Shirts, he added 1 first and realized he needed 3. So, he added another 2 using the agent without any problem.
- Participant B had some trouble working with the agent because there seemed to be an unstable connection between the agent and the API.
- Participant B directly asked "search for ..." and the agent responded with the product details.
- Participant B only asked for the items in the cart before placing an order. He did not actually go to the cart page to visualize the cart.

### Scenario/Task 2

	Number of error times during task 2 (count)	Number of frustrated error times during task 2 (count)	Frustrated error ratio (unitless)	SUS Questionnaire 2
Participant A	5	1	0.2	SUS Que...
Participant B	8	5	0.625	SUS Que...

- After the first task, participant A became more familiar with the system. He first searched by category and navigated to the hats page. He did not see any item named Michigan Football Hat. Then, he tried to add the Michigan Football Hat to the cart. The shopping assistant did a good job by saying “Sorry! I can’t find the item Michigan Football Hat.”
- Participant A asked “what is a plush?”, which the agent responded well by providing different tags of plushies in the shop.
- Participant B also did not have too much trouble to figure out the two items in the task did not exist in the shop. However, the API connection was in and out sometimes and the agent would respond not available sometimes, so he had to try multiple times for a request from time to time.

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

# Usability Findings

## *Quantitative Summary*

<Calculate averages and summarize>

The average time for completing scenario/task 1 is 244.5 second, or 4 minutes and 5 second approximately. The average ratio of the number of frustrated error encounters over the total number of error times is 41.25%. The average score using the calculation in the SUS method is 64, which is above OK and is a letter grade of D.

## *Qualitative Summary*

<Categorize your notes and report a minimum of two findings>

The first important finding from the notes is that different people say and describe things very differently. For example, when navigating the page, participant A would say “go to ... page” or “navigate to ... page.” However, participant B would just say the single word of the page. Before this test, I would think that I had enough training examples for the intents. But now, I think that training examples are never enough. The agent struggles a lot when different people describe the same thing or intention very differently and may respond with different results.

The second important finding from the notes is that context and simple training/help during the test process are very important when interacting with apps and agents. If the moderator does not provide any training or help for the users, the users can be clueless when interacting with the agent. It would take a long time for the user to figure out some interaction issues by themselves. For example, if the user is informed of the context that there are categories and tags in the shop without having to figure out by themselves, then the process would be less time consuming and easier for the user to navigate.

## *Conclusions*

<Report high-level insight and recommendations for how to improve the agent>

Regarding the first finding from above, the high-level insight is to provide more training examples for all the intents. One recommendation is to collect different descriptions and texts from different people. For the same intent or task, people describe it very differently. Therefore, collecting information and descriptions from a larger population instead of the developers themselves would help in this situation.

For the second finding from above, the high-level insight is to provide help and support for the users through their interaction with the web shop and the shopping assistant. For example, when the user first logs in, the agent can have more informational guidance and introduction to the app. In this way, the user is able to quickly understand the app and interact with the agent more comfortably and confidently.