

Field Usability Study (Team)

Due: Wednesday, February 19th, 2020 (by the start of class)

Goals

The goal of this assignment is to learn how to perform a field usability study and to incorporate the results of the study into design changes in your hi-fi prototype. In practice, this “pilot” study would be used to redesign your experiment before running the study with a larger pool of participants.

Prototypes

You will be performing this field test using the prototype you produced on the last assignment after you’ve fixed any issues that you know you wanted to change, implemented missing features, and made changes the teaching staff highlighted.

Participants

You will find **at least five** participants (i.e., volunteers who are not in this class) to work through your benchmark tasks (***the more you can run the better***). Remember, it must be voluntary. You should get the participants to **sign an [informed consent form](#)** and obtain other demographic information (e.g., age, sex, education level, major/occupation, experience with your type of tasks & application, etc.). Make sure they are **representative** of your target customers and that they are not participants that you have used previously during this project (unless you have pre-approval from the teaching staff). **A maximum of 2 Stanford students** are allowed as participants in your study.

Remember that you have a team testing budget of \$100 this quarter to use. For reimbursements, please read the rules and keep track of them on [this spreadsheet](#).

Benchmark Tasks

Your test will use the tasks that you have been using this quarter (unless we asked you to update or you decided to make changes). They should include at least 1 simple task, 1 moderate task, and 1 complex task. These tasks should give good coverage of your UI at this point; if they don’t you should redesign them. Note that applications that have multiple user roles (e.g., Lyft driver vs. rider), may require more than three tasks to test the application properly.

Measures and Observations

Although it will be hard to get statistically significant bottom-line data with only five participants and a prototype, you should measure and report on some important **dependent variables** to get a feel for how it is done (i.e., task time, task completion %, # of errors, etc.).

You will **concentrate on process data** in your presentation. For example, you should instruct your participant to **think aloud (if co-located)**. You should make a **log of critical incidents** (both positive and negative events). For example, the user might make a mistake and you notice it or they might see something they like and say “cool.” When a critical incident occurs, write down in the log what happened and the time.

You should also **record** the test **using a digital video camera or a phone** – note the time that you start recording so that you can find your critical incidents later– **as well as logging software** (see below).

Procedure

Give a brief high level overview of the context of your app without showing them how it works. Do **not** show them how to perform your tasks. You should write-up a script of your overview and follow the same script with each participant.

The participant will then be given task directions (e.g., on a single card) for the first task that **tells them what** they are trying to achieve, **not how** to do it. Make sure that you have given them enough scenario information (e.g., who they are or who their friends are) so they have the context for the tasks. When they are finished, you will give them the directions for the next task and so on. Each participant will perform **all 3** tasks (may differ for teams with two roles). Keep the data separate for each task & participant.

Location

This study will be carried out in a **field setting**. That is, if your application is meant to be used outdoors, then you will test it outdoors. If your application is supposed to be used in a user's home, then you will test it in their home. If its use would happen in a public place, you should test it in a public place. If it is a combination of settings, try it in different types of places. We want to see the tradeoffs of testing in a naturalistic setting.

Video and Clickstream Logging

You must **integrate video and clickstream logging** into your app. Use a tool that accomplishes this for you. For video logging, the built-in screen recorder on iPhone works quite well or the [Record It!](#) app, and there are many good apps on the Android platform, such as [AZ Screen Recorder](#).

For clickstream logging, there are many tools that are free or offer a free trial. For example, Google Analytics / Firebase Analytics (<https://analytics.google.com/analytics/web/>, <https://firebase.google.com/docs/analytics/>), Amplitude (<https://amplitude.com/>), Mixpanel (<https://mixpanel.com/>), and Lookback (<https://lookback.io/>), which has iOS and Android versions, as well as a React Native module for iPhone (but requires “detaching”). **Report on this data and say how it was helpful or not** helpful in your talk.

Results

You must **report your results** (values of dependent variables, summaries of those values, summaries of the process data, and summaries of clickstream data). You should **draw some conclusions** with respect to your interface prototype. **You should also say how your system should change if those results hold with a larger user population.** We want to understand how you would fix your system as a result of what you observed.

Examples:

- [Flutter](#)
- [Educonnect](#)

Deliverables

Presentation

One member of your team will present your project during a **12-minute** slide-based presentation in front of the other students and teaching staff. **If everyone has already presented on your team, someone can go again for the chance to receive a higher presenter grade (we will take the highest of your two presenter scores).** See the grading guidelines for information on how to structure your talk. You must **make the slides available for download on your web site.**

Presentation Guidelines

You will have **12 minutes** for this presentation plus up to 5 minutes for questions. Please practice as we will grade you on how close you are to the time limit (under and over). All team members are expected to work collaboratively on the presentation, though only one team member will deliver the presentation.

Suggested Talk Outline:

1. Project title & team (introduce yourselves)
2. Introduction
 - Introduce the system being evaluated (1 slide) & the **purpose** of the experiment (1 slide)
3. Outline of Talk (tell a story, do not read bullets)
4. Method
 - Participants (who -- demographics -- and how were they **selected** & **compensated**) (1 slide)
 - Apparatus (the **equipment/software** you used and the environment **where** tested) (1 slide)
 - Tasks (1 slide each) - describe each task & **what you looked for** during each task
 - Procedure (1 slide) - describe **what** you did and **how**
5. Test Measures (1 slide) - describe **what** you measured and **why**
6. Results (multiple slides) - Results of the tests (include **lots of images** & **data including clickstream**)
7. Discussion (multiple slides)
 - What you **learned** from the pilot run what you might change for the “real” experiment
 - What **you might change in your interface from these results alone**
 - **Difficulties** of running the experiment *in situ*? Any learnings particular to testing it in the wild? Did clickstream & video logging help? why/why not?
8. Summary of talk
 - What did your **team do**? What were the **key things you learned**?

Grading Criteria

Your grade will be based on the thoroughness of your experimental design, the analysis of your results, and the quality of your presentation. The presentation grading will be broken into two components: the individual grade of the presenter and a group grade for the quality of the content itself. Each bullet/grading category below will be out of 4 points (4=+, 3=check+, 2=check, 1=check-, 0=missing).

Presenter's grades (NAME: _____)

- Organization
 - ___ Project & team introduction
 - ___ Introduction to the system being evaluated & the purpose of the experiment
 - ___ Overview/Outline of talk (1 slide) – don't read this, tell it like a story
 - ___ Participants (who, how recruited, & how compensated)
 - ___ Apparatus & Environment
 - ___ Task & what you looked for
 - ___ Procedure
 - ___ Test Measures
 - ___ Results
 - ___ Discussion (what really matters)
 - ___ Summary of talk
- Presentation
 - ___ Use effective slides (easy to read, understand, good use of visuals/images)
 - ___ Cover required scope in 12 mins (+ 5 minutes Q&A). Practice in advance.
 - ___ Ensure the presenter makes eye contact and projects well. (___ : ___)

Group grade (GROUP NAME: _____)

- Introduction:
 - ___ clear what the system is for?
 - ___ valid reasons for running an experiment?
- Participants
 - ___ good range? demographics?
 - ___ recruiting/compensation clear?
- Apparatus & Environment
 - ___ clear on where the experiment took place and what equipment/software used?
 - ___ appropriate location?
- Tasks & Prototype
 - ___ good range?
 - ___ clear on what you were looking for?
 - ___ prototype implemented sufficiently to test?
- Procedure
 - ___ clear on what participants needed to do?
- Test Measures?
 - ___ appropriate variables recorded/measured?
- Results?
 - ___ clear if participants succeeded?
 - ___ clear on what worked and what did not, and why?
 - ___ clear on what was learned from logged clickstream data?

CS 194H Winter 2020: Assignment 9

Instructor: James Landay

- Discussion?
 - ___ recommendations for a larger study appropriate
 - ___ recommendations for design changes based on the results?
 - ___ recommendations for design changes good ideas?
 - ___ issues / advantages of running the experiment *in situ*?