Human Computer Interaction

Usability Testing

Prof. Andrew D. Bagdanov

Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Firenze
andrew.bagdanov AT unifi.it

December 19, 2017

Outline

Overview

Overview

Today

- In this lesson we will see a broad overview of the extremely diverse and dynamic field of usability testing.
- Rather than trying to be comprehensive, in this lecture I will merely introduce the main issues and types of tests.
- At the end of the slides there are some pointers to tools, and more in-depth guides to many of the topics covered here.

The rest of the course

19/12/2016	Material design
22/12/2016	Course wrapup, exam comments

Usability Testing: Introduction

Usability testing: basics

- What is usability testing in software and what are the benefits to the end user?
- Think of it as carefully designed tests to determine whether the application built is user-friendly or not.
- Usability Testing is a black box testing technique.
- It also reveals how comfortable users are in terms of the flow, navigation and layout, speed and content.
- Usability Testing tests the following features of the software:
  - How easy it is to use the software.
  - How easy it is to learn the software.
  - How convenient is the software to end user.

Usability testing: benefits

- Usability testing includes the following five components:
  - Learnability: Can users accomplish basic tasks the first time they encounter the design?
  - Efficiency: How fast can experienced users accomplish tasks?
  - Memorability: When a user returns to the design does she remember enough to use it?
  - Errors: How many errors do users make, how severe are these errors and how easily can they recover from them?
  - Satisfaction: Does the user like using the system?
- Benefits of usability testing to the end user or the customer:
  - Better quality software.
  - Software is easier to use.
  - Software is more readily accepted by users.
  - Shortens the learning curve for new users.

Usability testing: other advantages

- Advantages of usability testing:
  - Usability tests can cover other types of testing such as functional testing, system integration testing, etc.
  - Usability testing can be very economical if planned properly.
  - If proper resources (experienced and creative testers) are used, usability tests can help even before the system is finally released to the user.
  - Usability testing can discover potential bugs in the system not visible to developers.
- Usability testing is broad area and requires a high level of understanding of the field along with a creative mind.
- In this lecture I will introduce the basic concepts and try to give some concrete advice for designing your own usability tests.

Usability testing: bottom line

- The biggest challenge designers face isn't how technology works – it's how humans work.
- What users say versus what they do are two completely different things, and the only way to verify is to test.
- Usability testing is more than a just a checkbox on a list of product requirements – it is the most convincing support for your design decisions.
- Just like in software development, for usability analysis is it essential to test early and test often.
- Every product is different, so there is no magical usability test that will tell you everything you need to know.
- Usability testing helps you see the bottom line of whether your design works or doesn't.

Usability Testing: Defining Goals

Defining goals for usability testing

- The first step in usability research should always be knowing what you want to get out of it.
- This isn't always as easy as it sounds: you must categorize your testing goals and know what type of data is most appropriate.
- Michael Margolis (Google Ventures Design Studio) believes the first step to determining objectives is knowing the right questions to ask.
- It is important to focus the team on research questions:
  - Research question: Why do people enter the website and not watch the demo video?
  - Dictating methods: We need to do focus groups now!

Defining goals: asking the right questions

- Relevant Product Information: Do you know the history of your idea? Do you know what's coming in the future?
- Users: Who uses your product? Who do you want to use your product? Be specific: demographics, location, usage patterns.
- Success: What is your idea of success for this product? Make sure there is consensus.
- Competitors: Who is your competition? How do you compare? What do users expect based on existing tools?
- Research: What do you want to know? What data would help your team best? Is that research already available?
- Timing and Scope: What time frame are you working with for collecting your data? When is it due?

Defining goals: asking the right questions

- Have everyone write down (on sticky notes) questions they have about their users and the interface.
- Collect all the questions and stick them to a board, then try to organize all the questions based on similarity.
- You'll see that certain categories will have more questions than others – these will likely become your testing objectives.
- Your objectives should be simple:
  - Simple: Can visitors find the information they need?
  - Complex: Can visitors easily find our products and make an informed purchase decision?
- These questions form the basis of your goals for UX testing.

Knowing what to measure for UX testing

- Now that you roughly know your goals, you must decide how to apply usability testing to accomplish them.
- What type of feedback would be most helpful for your results?
- Does your team need a graph or a rating scale? Personal user accounts or numbers? Written responses or sound bites?

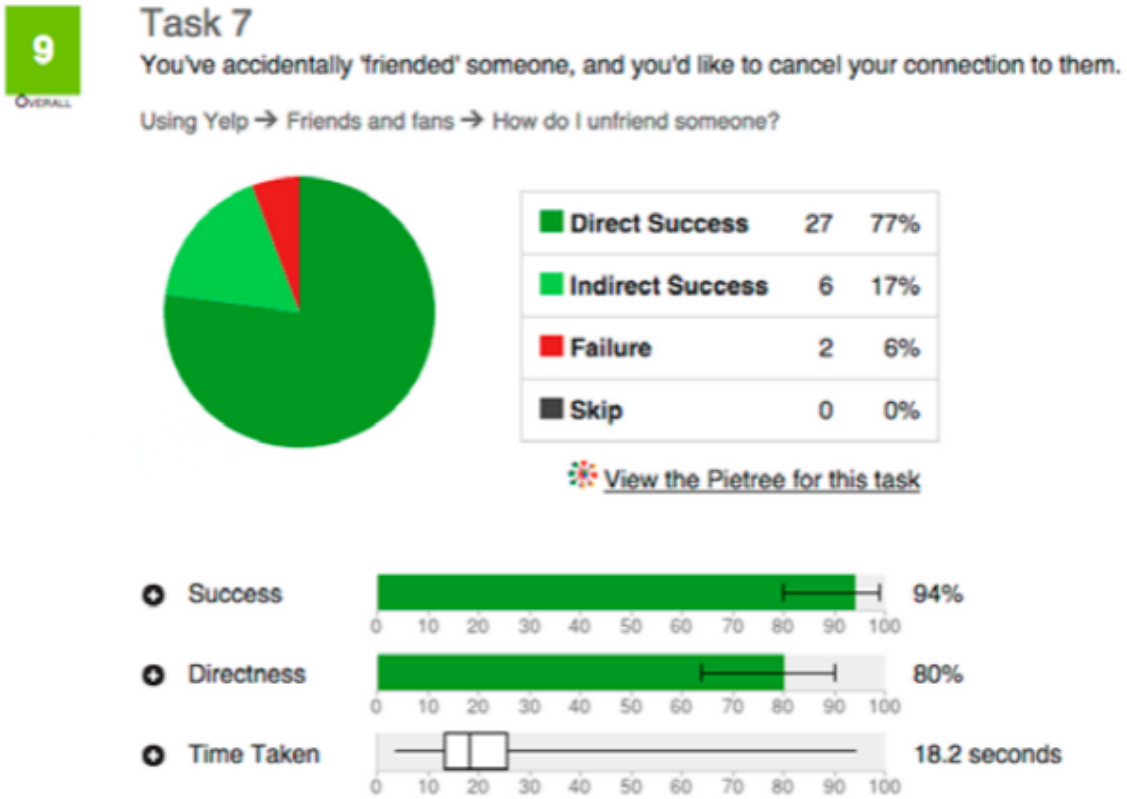
Type	Example	Results
Verbal Response	Describe and demonstrate what, if anything, was most frustrating about this site.	Spoken answers correlate with where a participant is at in the study. Make great clips for a highlight reel.
Multiple Choice	Do you trust this company? • Yes • No	Great for collecting responses that are **categorical**. These can be nominal (cats or dogs?) dichotomous (yes or no) and even ordinal (**Likert** scale agree/disagree).
Rating Scale	How likely are you to return to this site again? *1 2 3 4 5* *Not at all likely* *Very likely*	Good for collecting ordinal variables (low, medium, high) and are very recognizable especially within the United States.
Written Response	What do you think is missing from this page, if anything?	Good for running post-study analysis. How many people used the same answers? Quick quotes for building user stories.

Usability metrics

- These are quantitative data about usability.
- Usability metrics are statistics measuring a user's performance on a set of tasks. The website usability.gov lists some of the most helpful:
  - Success Rate: In a given scenario, was the user able to complete the assigned task?
  - Error Rate: Which errors tripped up users most? These can be divided into two types: critical and noncritical. Critical errors prevent a user from completing a task, while noncritical errors simply lower the efficiency with which they complete it.
  - Time to Completion: How much time did it take to complete the task?
  - Subjective Measures: Numerically rank a user's satisfaction, ease-of-use, availability of information, etc.

Usability metrics: task summarization

- Task summaries look like this:



Quantifying qualitative data: SEQs

- The Single Ease Question (SEQ) is a 7-point rating scale to assess how difficult users find a task.
- It's administered immediately after a user attempts a task in a usability test.
- After users attempt a task, ask them this simple question: Overall, how difficult or easy was the task to complete?

- Labels and values: label the end points only and provide numbers from 1 to 7. Slight changes are far outweighed by the very salient event of the just attempted task. Users are have little problem expressing their frustration or delight.

Quantifying qualitative data: SEQs

- SEQs work well: Despite their simplicity, the SEQ performs about as well or better than more complicated measures of task-difficulty (see this CHI paper for a comparison).
- Ratings of difficulty correlate with other metrics: the correlation between user responses on the SEQ and task-time and task-completion is around r =.5.
- Users respond differently: Some users will make everything a 6 or 7 while others will use the full range of the scale. It's very common for people to use rating scales differently, but these differences tend to average out across tasks.
- SEQs are technology agnostic: A feature of task-difficulty ratings is that users tend to respond to what they expect given the device, fidelity of the interface, and nature of the task.

Quantifying qualitative data: SEQs

- Ask Why?: When users rate a task difficult, it's good to know why they did. When a user provides a rating of less than 5, it's a good idea ask them to briefly describe why they found the task difficult.
- If you're going to ask something: There is quite a bit of evidence that if you're going to ask the users one question about their experience with a task, it should be the SEQ.
- Use it for everything: You can use SEQs to ask about many aspects of usability: the scale can also be "1. Strongly Disagree – 7. Strongly Agree"
- Mix it up: don't always ask questions that demand the same response (1 or 7).

Usability Testing: Type of Test and Participants

UX Testing: the Pilot Test

- No matter what type of test you choose in the end, you should always start with a pilot test.
- Pilot testing is like a test run of your greater user test.
- You will conduct the test and collect the data in the same way you would a real test, but the difference is that you don't analyze or include the data.
- In most cases, something will go wrong with your first test.
- If you want the most reliable data, run a pilot test or two until you feel you understand the process and have removed all the kinks.

UX Testing: types of tests

- There are four general types of usability tests:
  - Scripted use of the product: These tests focus on specific usage aspects. The degree of scripting varies, with more scripting generating more controlled data.
  - Decontextualized use of the product: Tests that don't use the product – at least in the actual testing phase – are designed for broader topics and generating ideas.
  - Natural (and near-natural) use of the product: Seek to analyze common usage behaviors and trends with the product, doing well with data authenticity at the cost of control.
  - Hybrid: Hybrid tests are creative and non-traditional tests geared towards understanding user mentality.

UX Testing: types of tasks

- Each type of test is divided into tasks.
- The first distinction to make is whether tasks will be phrased directly or in terms of scenarios:
  - Direct Tasks: are instructions such as "Find a turkey recipe on the Food Network," or "Learn about wiener dogs on the blog." Direct tasks are more technical in nature, and their unnatural nature can detract from user experience.
  - Scenario Tasks phrase the instructions in a real-life example: "You're going to a high school reunion this weekend. You want to find a nice outfit on the Macy's website". Scenario tasks are more common than direct tasks because the user can forget they are being tested.

UX Testing: types of tasks

- The next distinction to make is whether tasks will be closed or open-ended:
  - Closed tasks: have clearly defined success or failure. These are used for testing specific factors like success rate or time. For example (from a Yelp redesign case study): "Your friend is having a birthday this weekend. Find a venue that can seat up to 15 people."
  - Open-ended tasks are ones the user can complete in several ways. These are more subjective and most useful when trying to determine how your user behaves spontaneously, or how they prefer to interact with your product. For example: "You heard your coworkers talking about Yelp. You're interested in learning what it is and how it works."

UX Testing: finding your audience

- The core component of these usability tests are actual people.
- To think of your participants as merely test subjects is a mistake – they are all individuals with their own personalities and their own way of doing things.
- When focusing in on your test group, it's important not to obsess over demographics.
- The biggest differentiator will likely be whether users have prior experience or are knowledgeable about their domain or industry – not gender, age, or geography.
- If you find you have more than one target group, remember to test each group independently of each other.

UX Testing: recruiting participants

- Knowing who you want for the test is only half – you still need to get them to come.
- These are some effective methods to find test participants:
  - Existing Users: By definition, these are your target users. Even if you're researching a new product, if your company has produced similar products in the past there's a chance they both target the same type of person.
  - UserTesting.com: A website designed specifically for this. It lets you select users by age, gender, location, and even more customizable options. The site delivers audio and video of users actually testing your site or app.
  - Mechanical Turk: Amazon's crowdsourcing network is the cheaper version of UserTesting (but you get what you pay for). If your testing is simple, it is easy to recruit many people for relatively little cost.
  - Hallway Testing: "Hallway" testing is a term that means random, as in whoever if walking by the hallway when you conduct the test. Remember that the farther you get from your target audience, the less helpful the data.

UX Testing: hallway testing

- Hallway testing is a usability test set-up in a high foot traffic area, utilizing bystanders to test your product.
- Participants will be people who happen to be walking down the hall and are able to afford 5-10 minutes of their day.
- Some concrete Hallway testing tips:
  - Location is extremely important. Choosing the optimal location for your hallway test is invaluable. Ensure your location will have heavy foot traffic. Be aware of the timing of your hallway test. Do not schedule hallway tests during inconvenient hours or major events.
  - Plan ahead. As impromptu as hallway testing sounds, it takes a lot of planning. Preparation for a hallway test can start as early as a month before the actual test.
  - Set up early. Give yourself ample time to set up and get the testing team situated. You should arrive at least 30 minutes before you are scheduled to begin testing.

UX Testing: hallway testing

- Concrete Hallway Testing tips (continued):
  - Review and practice. Run through the test script multiple times and inform team members exactly what you are looking to identify during the test. It is key to inform team members about the duties of their role.
  - Use greeters. Greeters play an important role during hallway tests, namely identifying and recruiting the test participants. It is important to have outgoing greeters to get people involved in the testing sessions, some personalities are better suited for this task.
  - Be mindful of time. The optimal time for an individual hallway test is 10 minutes. Focus and interest tend to wane if you keep them any longer.
  - Explain the purpose to the participant. Let the tester know the ultimate goal of the hallway test. They have agreed to the parameters you have set for the test, do your best to keep them focused on the task at hand.

UX Testing: hallway testing

- Concrete Hallway Testing tips (continued):
  - Reward your volunteer. Participants are volunteering part of their day to aid your product, you should reward them for their kindness. Simple gestures can be effective, like handing out pens or candies.
  - Look to improve. Always be aware of how you can improve your testing processes. A team debriefing session should be a requirement after hallway testing sessions.
  - [MAKE VIDEO GO NOW (First Fridays)]
  - Guerilla Testing Technique

Usability Testing: Scripted Tests

UX Testing: scripted tests

- A scripted test is the most controlled of the test types, and is recommended for testing if the user can find certain feature (or how long it takes to do so).
- They produce more quantitative data, but can also generate qualitative data as well, depending on the how tight or controlling the script is.
- Before designing a script for a scripted test, you must make a critical decision: whether the tests will be moderated or unmoderated.

UX Testing: moderated tests

- Many UX designers believe that the payoff of moderated testing significant if you have the time available.
- A moderator can help probe the participant to delve deeper, creating data that is more complete – plus they can keep users on track and clarify any confusion.
- Also, user reactions and even body language can provide useful data as well, but only if someone is present to document and interpret them.

UX Testing: moderated tests

- Moderated tests are usually conducted in controlled environments, like UX testing laboratories:

UX Testing: moderated tests

- Due to many factors, moderated testing is not recommended for all tests.
- It can be very useful in the following situations:
  - Early stages in the development process: Specifically in the prototyping phase, where features may be incomplete or not even work, a moderator can help answer questions and explain the unclear parts.
  - An advanced, complicated, or high-level product: As with a prototype, if there is a great chance for confusion or misinterpretation, a moderator will help keep things on course.

UX Testing: moderated tests

- Even moderation proponents admit that moderated tests have their drawbacks, specifically convenience.
- Moderated tests require a knowledgeable moderator, their time, and usually a specified location, as opposed to remote usability testing.
- Coordinating the schedules of moderated tests can be problematic, and only one can be done at a time, unless more moderators are hired.
- More importantly, moderated tests can take participants out of their comfort zone, so special care must be taken to avoid the various kinds of biases.

UX Testing: types of bias

- It's important to understand the subtle biases that creep into moderated (and unmoderated) usability tests:
  - Hawthorne Effect: "You and all those people online and in the next room are watching my every keystroke, I'm going to be more vigilant and determined than I ever would be to complete those tasks – I'll even read the help text."
  - Task-Selection Bias: "If you've asked me to do it, it must be able to be done." This state of knowing is unnatural.
  - Social Desirability: Users generally tell you what they think you want to hear and are less likely to say disparaging things about people (seen and unseen) and products. Users tend to blame themselves.
  - Availability: If a user has two hours during the day to volunteer for a study, it limits the users you are testing to those who are available. (Also problematic for the jury system in the USA.)
  - Honorariums: If the honorarium the user receives is the sole motivator, the quality of the data can be questionable. In unmoderated studies (e.g. Mechanical Turk), users will cheat to receive the honorarium.

UX Testing: types of bias

- Types of bias (continued):
  - Note Taking: "I see you wrote something down after I did that so I must have done something wrong." Users who are aware the moderator is taking notes may become more self-conscious about the actions they are taking.
  - Recency and Primacy Effects: The Recency Effect is the tendency to weigh recent events more heavily than earlier events. Conversely, weighing events that happened first more heavily is called the Primacy Effect. Users typically perform worse on their initial tasks.

UX Testing: unmoderated tests

- While moderated testing allows for instantaneous give-and-take feedback, there are advantages to letting users interact with a product in its natural environment.
- You can get maximum value for minimum cost when the tasks are written as clearly as possible.
- Users are encouraged to think out loud, and you record their onscreen interactions (when you can).
- When the test is done, you can then use the video clips that are most insightful and use them to inform design changes.

UX Testing: unmoderated tests

- Unmoderated testing benefits include:
  - Time savings: Simultaneously testing hundreds of participants. You can also test multiple products at once, including competitors.
  - More natural product usage: Remote usability testing allows participants to remain in their natural environment, their use of the product will more closely resemble real-world scenarios.
  - Cost savings: Costs are usually quite low since you don't need to pay for moderators or equipment setup. Unmoderated tests are also scalable depending on the testing tool used.
- [MAKE VIDEO GO NOW (remote-testing)]

Usability Testing: Natural Tests

UX Testing: natural testing

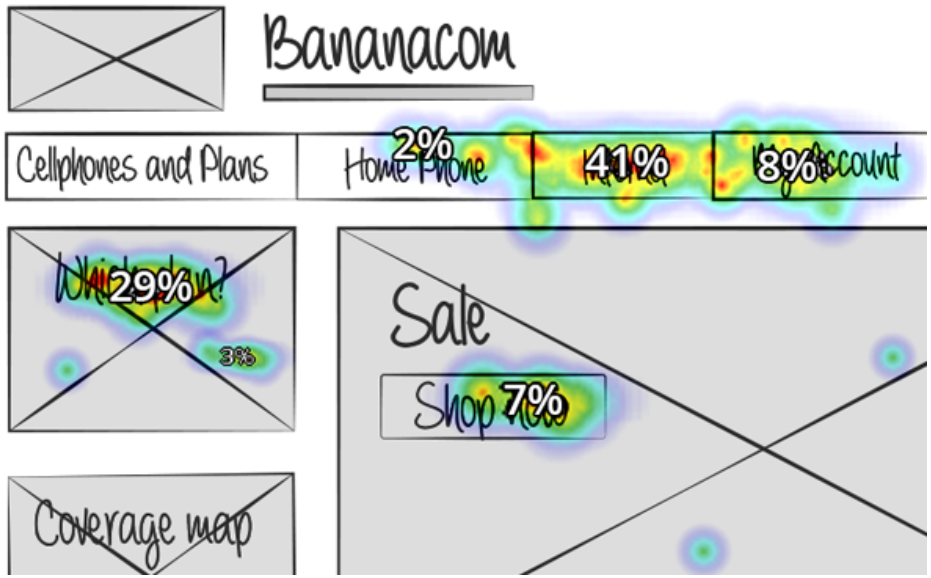
- Tests in which people use the product naturally (without a script) are the closest you will get to seeing how your product might perform "in the wild."
- Natural and near-natural tests minimize the amount of interference from the observer.
- Because the goal is to minimize interference from the study, natural tests are usually conducted remotely and without a moderator.
- The most common natural tests: A/B testing, first click testing, field/diary studies, and eye-tracking.

UX Testing: A/B Testing

- In an A/B test, different groups of participants are presented with two choices or variations of an element.
- It is generally a scientific test, where only one variable differs, while the rest are controlled.
- Mostly conducted to test if a certain layout, placement, or messaging results in better results.
- A/B testing is considered a natural test because users are not notified nor provided a set of tasks.

UX Testing: first click testing

- It turns out that for any given task, a user's success rate is 87% as long as their first click is correct.
- If their first click is not correct, the chances for success fell to below 50%.
- This is a "near-natural" test because users are still assigned tasks.
- But, these tests are usually unmoderated and ran remotely in the comfort of the user's home.



UX Testing: field studies

- A field study provides data you can't find anywhere else by letting you observe users in their own environment.
- Field studies provide three main benefits:
  - Terminology and processes: In an interview setting, a user may not be aware of how they behave or how they would talk about a product in a natural setting.
  - Context: Users aren't always aware of how external factors, like timing for example, affect their decisions. Field studies mark the times and environments of the user, and their impact can be seen during the analysis of the data.
  - Similarities and Differences: By observing how the user interacts with different products, you can start to notice similarities and differences, which will flesh out your data to enormous degrees.
- The biggest downside is primarily the cost of organization and time required.

UX Testing: diary studies

- A less-involved study of a user in their natural environment is the diary study.
- Participants are asked to keep a diary and account for their experiences with a type of product or system.
- A diary study captures the expectations, mindsets, moods, and social contexts that affect the user experience.
- A diary study might reveal that a bad mood or criticism read on the web impacted the user's assessment of the product experience, independent of the product itself.

UX Testing: diary studies

- Like all others, diary testing too has drawbacks:
  - Significance of participant: The quality of results will depend on the quality of the participant. The participant's self-awareness, self-expression, and writing skill can all skew the results.
  - Training sessions: While it may sound like the participant acts independently, the truth is that a thorough training session is necessary to ensure the participant understands exactly what is expected before starting.
  - Analysis: The analysis of an entire diary is time-consuming, especially if it is hand-written.

UX Testing: diary studies

- To help counter the downsides, you can follow a few best practices:
  - Provide contextual and open-ended questions: Contextual questions like, "What prompted you to use the app?" give you direct insight.
  - Let users decide how to record themselves: Text, online photo galleries, voice recording, even Twitter can all work. It also helps the process feel more natural and makes participants less self-conscious.
  - Keep size in mind: The diary (whatever form) can be as small or large as needed. On paper, space for forty entries can be overwhelming, while ten might be more encouraging.

UX Testing: eye tracking tests

- An eye tracking test tracks a user's eye movement, and more to the point where specifically they are looking at at very high temporal resolution.
- Eye-tracking is useful to gather massive amounts of quantitative data.
- Some general conclusions we already know from eye-tracking experiments:
  - Users are predictable: As we can see by the eye tracking patterns above, people follow similar trends, allowing us to plan our visual layouts for the masses.
  - Users search pages differently depending on goals: eye patterns differ depending on why users are searching. Browsing and searching for something in particular have two different modes.
  - Users are drawn to visuals: Visuals like thumbnails or vibrant colors will attract a user's attention more than plain text.
  - People ignore ads: People will neglect ads habitually, so online advertisers will have to work harder.

Usability: HelloSign Case Study

The HelloSign starting point

- HelloSign is an iOS app that enables users to scan, sign and send documents from their phone using the camera.
- It was a mediocre success upon release, but after usability testing to a solid five star success after a redesign.
- Here we will look at some elements of the usability study used to improve the application.

HelloSign: Overview

- Hellosign app has four primary sections: authentication, welcome, document creation and document editing.
- The biggest made to the app were to the authentication and welcome screens.
- This seems like a very modest intervention, but these are the first points of interaction for all users.
- Let's first look at these (and other section) to see how the original app worked.

HelloSign: Authentication and Welcome

- The authentication and welcome screens are critical moments in the initial experience of the product.
- The app was designed to complement the website, so users were assumed to be somewhat familiar with the product.
- Had this been designed as a standalone app, authentication would have been a secondary option, rather than a requirement.

HelloSign: Document Creation

- The document creation process consists of a camera with guides to position the document in the frame.
- The designers we looked to the native iPhone camera screens and to Instagram.
- Why is this?

HelloSign: The Usability Test

- UserTesting.com was used for the initial usability test.
- Users record videos as they go through whatever pre-specified tasks assigned to them.
- Users were also asked several follow-up questions of each user, such as, "What was the most frustrating part of your experience?"
- For HelloSign, a test was designed that took users through a typical use case: creating, editing and sending a document.
- Follow-up questions queried users on ease of use, areas of difficulty and areas for improvement.

HelloSign: The Review

- Each video was reviewed twice – the first pass was merely to identify any glaring issues and identify tester style.
- On the second pass, specific problems were noted.
- During testing, most users don't articulate problems they're having, but the problems will be fairly obvious from their behavior.
- A problem is obvious when the user does any of the following:
  - pauses for a few seconds when trying to complete a task;
  - stumbles and has to backtrack in their steps or has to undo an action;
  - expresses audible frustration (a sigh or grumble);
  - takes a longer route to achieve a goal than expected; or
  - fails entirely at a task.

HelloSign: The Redisign

- The following problem areas were discovered:
  - Creating pages was too repetitive a process. After taking each photo, the user had to tap "Add page," take another photo and then repeat. This was tedious, and some testers remarked that the process felt unnecessarily repetitive, while others expressed audible frustration.
  - Users could not edit a date after adding one. One tester wanted to add a past date to a document. While a "date" object could be added to the document, it showed only the current date and could not be edited. This was confusing and unnecessarily restrictive.
  - Markers for aligning the document during scanning needed refinement. Users had trouble lining up the document with the boundaries on the camera screen, with some expressing frustration and with many giving up.
- Moreover: a significant amount of confusion was experienced by many users immediately at the sign-in/up and welcome pages.

HelloSign: Fixing Authentication

- Dumping the user right into the sign-up screen, with only a "Sign up" button as a state indicator, was confusing.
- During testing, most users seemed to expect this screen to let them sign in, not register.
- The revised screen added a step, forcing the user to explicitly select "Sign up" or "Sign in," making it a conscious decision:

HelloSign: Fixing Authentication

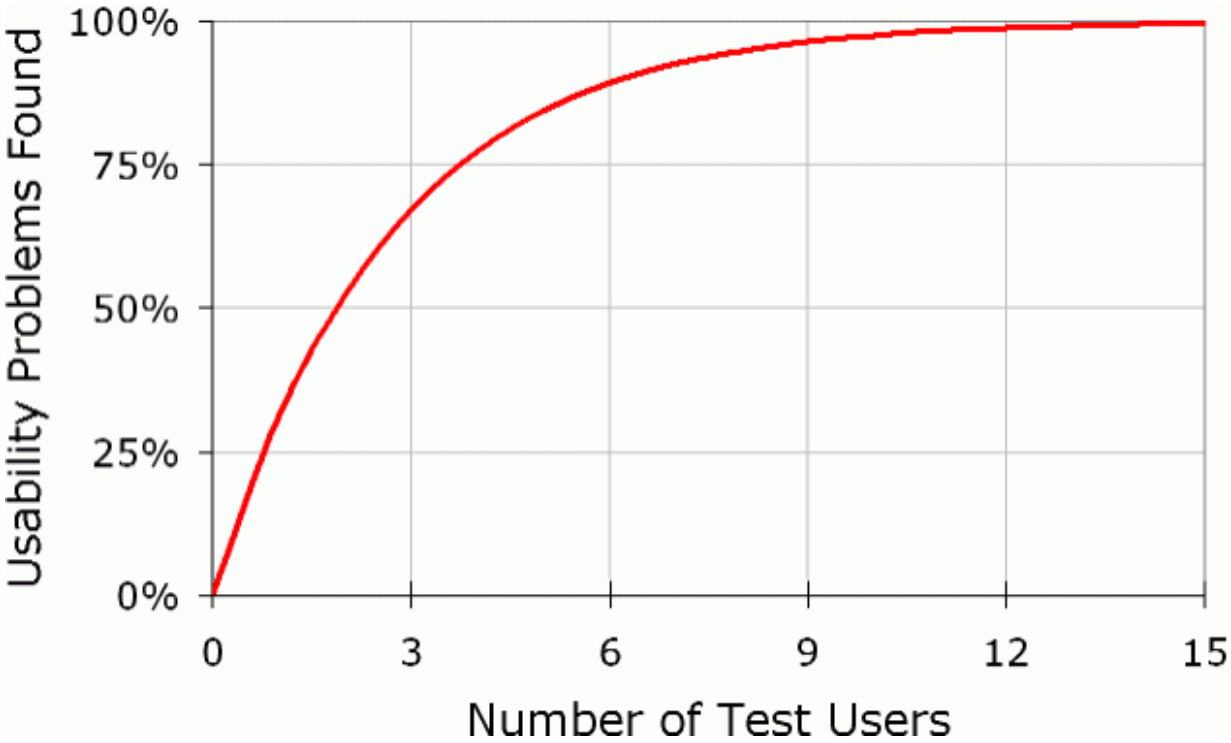
- After selecting one of the two options, users were brought to one of two nearly identical screens
- The only difference was in the labeling of the button: "Sign up" or "Sign in".
- In the revision, the layout was simplified and Google authentication was added.
- Forcing the user to choose a path cleared up any confusion.

HelloSign: Fixing the Home Screen

- The home screen was also heavily revised
- The original version de-emphasized Help and generally felt "clunky and heavy-handed."
- The revision brought Help to the forefront and highlighted the Scan action (the primary purpose of the app).

HelloSign: More is Not Always Better

- As mentioned before, testing interfaces on many users is not necessary in order to identify issues – even 8-10 users can be enough.
- This curve was first shown by Jakob Nielson and Tom Landauer[1]:



[1] Nielsen, Jakob, and Landauer, Thomas K.: "A mathematical model of the finding of usability problems," Proceedings of ACM INTERCHI'93 Conference, 1993.

HelloSign: Takeaway

- Design is a highly iterative process, and all of the intuition in the world won't identify gaps in your product.
- As designers, we are too familiar with our own work to be able to spot where it fails.
- The only way to truly improve a design is to test it on real users and watch how they interact with it.
- Testing with a live app uncovered problems that turned a mediocre effort into a five-star product – with only a little work.

HelloSign: More is Not Always Better

- Some observations:
  - Zero users is not enough.
  - As soon as you collect data from a single test user, your insight skyrockets and you have learned almost a third of all there is to find.
  - The difference between zero and even a little bit of data is astounding.
  - When you test the second user, you will discover overlap with the first one – but people are different and you will still learn something.
  - As you add more and more users, you learn less and less because you will keep seeing the same things again and again.
  - After the fifth user, you are wasting your time by observing the same findings repeatedly but not learning much new.
- Summary: Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford.

Usability: Project Examples

The AirMouse usability test

- I now want to spend a few minutes revisiting the usability tests performed for an HCI Course Project.
- It is a good example of what I expect from your usability testing exercises.
- AirMouse studied how to emulate the behavior of a physical mouse using the 3D hand model generated by a Leap Motion device.
- Machine learning techniques were used to recognize the "mouse grab" pose and a small set of mouse gestures.
- It was evaluated with a usability test to assess user opinions and by a set of timed tests to collect an objective measure of the performance and an estimation of the learning curve.

AirMouse: The Testing Protocol

1. Timed test: 5 colored squares are presented on the screen, the user has to follow the instruction on the screen and click the right square. The application records the total time spent to complete the 16 instructions.
2. Task oriented test: The user is asked to complete a series of routine tasks, the supervisor is allowed to help the user by giving some tips.
3. Timed test again.
4. Free Use: The user is left free to use device to navigate Google Maps; the suggested task is to find his house on the map (by dragging the map / zooming in and out).
5. Final timed test.

AirMouse: The Timed Test

- [Make webpage go now]

AirMouse: The Tasks

- (a) Open Firefox browser
- (b) Open a new tab (plus button in the tab bar)
- (c) Open the website in the first panel
- (d) Scroll down until a red button is found
- (e) Click the red button (it will open Google News)
- (f) Click the second news
- (g) Close the tab
- (h) Open a new tab
- (i) Swap the third and the fourth panel
- (j) Close Firefox

AirMouse: Example Questions

1. **Task completion required too much effort** *

   Il completamento dei compiti ha richiesto troppo sforzo
   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

2. **Task completion required mental concentration** *

   Il completamento dei compiti richiede concentrazione mentale
   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

AirMouse: Example Questions

3. **I was able to complete the tasks very quickly** *
   Ho completato i compiti molto rapidamente
   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

4. **The device is accurate** *
   Il dispositivo è accurato
   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

AirMouse: Example Questions

5. **Actuation required a lot of force** *

L'utilizzo richiede molta forza fisica
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

6. **The operation is very smooth** *

Il funzionamento è molto fluido
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

AirMouse: Example Questions

7. **I felt very uncomfortable during operation** *
   Sono stato scomodo durante l'utilizzo
   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

8. **The system is very easy to use** *
   Il sistema è molto semplice da utilizzare
   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

AirMouse: Example Questions

9. **The gestures and hand poses are difficult to remember** *

Le "gestures" sono difficili da ricordare
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

10. **I learned to use the system better after a little while** *

Sono stato in grado di utlizzare il sistema dopo poco tempo
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

AirMouse: Example Questions

14. **Most people will learn to use the system quickly** *
La maggior parte delle persone impareranno a usare il sistema rapidamente
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

15. **Overall, I am satisfied with the system** *
In generale sono soddisfatto del sistema
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

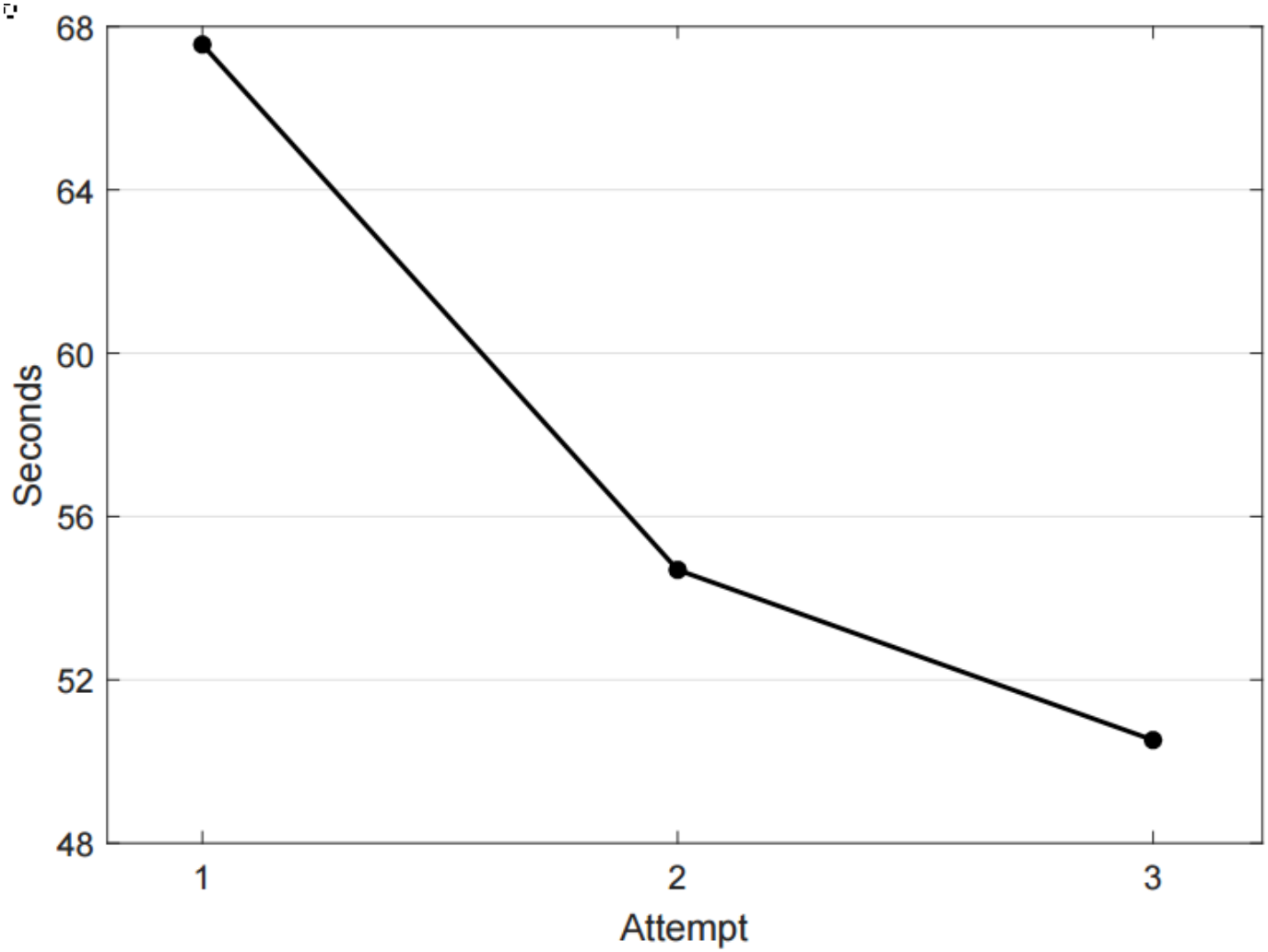
AirMouse: The Qualitative Results

- The following table summarizes the answers to the task-oriented questions:

N	Question	Mean rating	σ	% Agree
1	Task completion required too much effort	3.2	1.6	19%
2	Task completion required mental concentration	3.6	1.3	19%
3	I was able to complete the tasks very quickly	4.2	1.5	44%
4	The device is accurate	4.9	1.4	69%
5	Actuation required a lot of force	2.1	1.2	6%
6	The operation is very smooth	4.6	1.5	56%
7	I felt very uncomfortable during operation	2.5	0.8	0%
8	The system is very easy to use	4.5	1.2	50%
9	The gestures and hand poses are difficult to remember	1.7	0.8	0%
10	I learned to use the system better after a little while	5.7	1.3	81%
11	The system is illogical or inconsistent	1.3	0.5	0%
12	Position and posture are comfortable	3.9	1.2	25%
13	Help given during the test is very important to understand how the device works	4.9	1.2	50%
14	Most people will learn to use the system quickly	5.3	1.3	75%
15	Overall, I am satisfied with the system	5.7	1.0	81%
16	I like the similarity with a physical mouse	5.1	1.8	63%

AirMouse: The Quantitative Results

- The following plot summarizes the quantitative results:



AirMouse: Takeaway

- The AirMouse Usability Study is an excellent example of controlled evaluation of an interface idea with known problems.
- It demonstrates both the limits of intuition and the value of usability testing with real users.
- See the full code and usability tests (including questionnaire) in the github repository:

https://github.com/buddino/AirMouse

Summary

Summary

- In this lesson we saw a broad overview of the critical issues in usability testing.
- We saw the main advantages and disadvantages of scripted versus unscripted interaction tests.
- And we also saw the advantages of moderated versus unmoderated tests.
- UX testing is an extremely dynamic topic – necessarily so due to the fast pace of innovation.
- For us, or for anyone with limited resources and modest goals, hallway testing is probably the best compromise.
- You should use SEQs as the basis for collecting quantitative usability data.
- There are many, many tools and services available to help collect and analyze usability data (see Resources section).

Summary

- There is no magic formula for usability testing.
- It is part science (a small part) and part (a large part) art.
- It is effective already with only 5-10 users.
- The important things to do are:
  - Have clear ideas about what you want to obtain (i.e. craft your tasks and questions to your needs).
  - Test often and with relatively few subjects to optimize results.
  - Use SEQs with a 1–7 Likert Scale.
  - Summarize results both qualitatively and quantitatively.

Resources

Resources

- Here I have gathered some links to tools and in-depth information.
- Tools:
  - Optimal Workshop offers a range of UX evaluation products covering most of the types of testing discussed today.
  - UserTesting.com offers remote testing support with video.
  - Some unmoderated testing tools
- An excellent starter kit for usability testing (includes forms and checklists for interviewers and interviewees).
- Tips on moderating UX testing sessions
- A/B Testing in depth
- Guide to field studies
- Guide to diary studies