Conducting Usability Testing (intermediate)

Introduction

Why Usability Testing

1. See if we meet user expectations
2. Match our business decisions to real-world use
3. Engineer out the flaws
4. See how successful users are with their tasks
5. Find out if we are on the right track
6. Get user reactions feedback

Make sure someone else test it, as we can get engrossed in our design after a while. It's a way to correct early in the process. There is no rule about how much how often, but any time we're not sure about our design, we should do it then.

Types of Usability Testing

- Down the hallway testing (grabbing anyone and get them to use the design = quick and dirty)
- Low fidelity (paper-prototyping) (=wizard of oz = smoking mirror test) - Early on
- Mixed-fidelity (paper and clickable PDF) (make sure they see a button, menu, it's about expectation)
- High-fidelity (Fully functioning prototype - Axure for example) - later on

Problems with Down the hallway / "quick and dirty"

Formal testing = + set up and lab + formality + time + cost + big event
- Sloppy recruiting
- Forced observations
- Poorly defined tasks
- Number of users
- Inaccurate findings (friend might want to please you or be the wrong audience)

4 steps:

1. Create Test Plan (recruit users, define test scope, identify objectives, establish metrics)
2. Facilitate Test (observe users, identify issues, identify solutions, interview users)
3. Analyse Test Data (assess user behaviour, analyse user click path, identify problem areas, assess navigation)
4. Create Test Report (review video footage, identify design issues, identify best practices, design recommendations)

Is Simply watching enough? Yes but keep agenda out of your observations, never prompt the user, listen and don't lead, follow the user or you might end up with dirty data.... this is what you really want to avoid!

UT is largely a qualitative activity. We usually have a small sample of users.

Quantitative:

- Time on task
- Success/failure rates (number one metric captured in UT) = ease of use

Do they get the task (success rate) do they not (failure rate), do they get part of it (partial success)

- Effort: e.g. number of clicks / perception of progress

Qualitative:

- Stress responses (looking at their face, change in body language)
- Subjective satisfaction (a questionnaire about how well they think they did)
- Perceived effort or difficulty

Why metrics:
1. Make usability recommendations concrete
2. Ground team in reality (probability thinking)
3. Help iterate and validate design concpets
4. provide objectivity to design debates
5. Guide fact-based design decisions

Warning: Common remote UX testing tools do not provide usability metrics capture

Summative: large sample size, to capture metrics, can be 50 for example, can be online (remote) but in that case they are not good to capture metrics (high fi and remote)
Formative: to see how users behave, quick insight, smaller more agile group, like the 'famous 5' users (low fi and more informal)

It's best to do a blend of the two.

What we are measuring:
- **Behaviours** (task performance, speed, efficiency, goal fulfilment, expectation matching)
- Opinions (how it looks, thoughts and opinions)
- Data (visitors, pages, documents)

What people think they do is actually different from what they really do.

Software used: User Testing.com is a good one for remote testing, silverback, morae and loop11 are the main ones.

Usability issues:
- Not seeing something
- Going in the wrong direction
- Thinking it's correct when it isn't
- Missing a "rule"
- Confusion
- Errors
- Highlights

Think aloud technique:
- Tell them: Test the interface not the user
- Ask them to externalise thoughts/feelings
- Verify researcher mental model
-Take notes (structured or unstructured)
- Capture on video- analyse later or distribute

Time on task
It used to be the king of usability metrics, but it's not necessarily true anymore, it depends on what you are testing (context).
It used to be: 3 (3 click rules) 3 strikes and you're out, but now it's more about perception of progress (you have signs about where it is next all along, it might take longer to get there, but you're ok about it)

Others metrics: number of confusions, number of errors
Actual vs Expected difficulty
Severity rating (impact vs percentage of users who had the issue)
We don't want the user to think they are working for us.

Mobile UX metrics:
1. Does the interface maintain flow? Number of interruptions
2. Is navigation apparent at all times? Note visibility of controls
3. Tablets: does the interface sustain orientation changes: note transition disconnect when user switches

Developing Your Test Plan

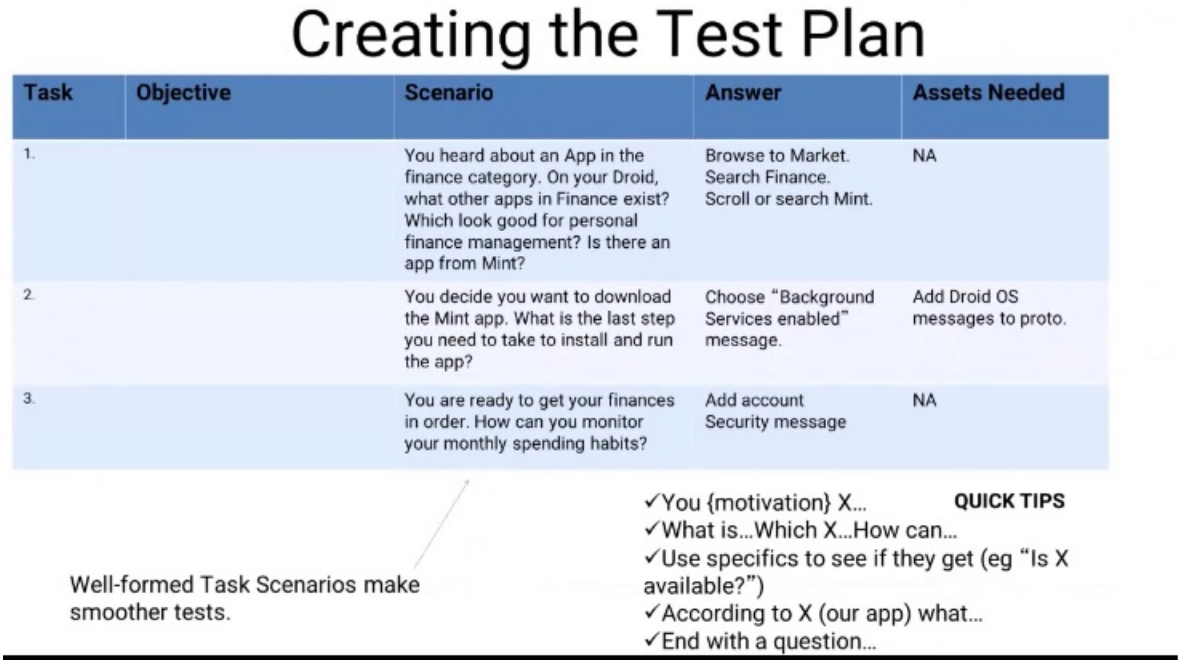
Preparing a Test Plan

It is important to spend time on your test plan to make sure what you are testing will be valuable or useful.
Factors to consider:
1) Are you clear about the business objectives behind each task?
2) Do you have time to manage input from others in order to get early buy-in?
3) Does everybody understand what a test plan is and what the goal of the test is?
- Decide what areas to concentrate on
- Determine potential usability issues
- Write a Test Plan
- Determine what task will be tested (takes between 1h and 90minutes, 2 hours being a maximum. About 12 tasks, 10 to 15 tasks and about the same number of users is a good number.)

Your test plan should include:
- Objective
-Test Task (scenario - good to end it with a question)
- End state (answer)
- Assets needed (for managing time up to the test)

Scenario: this what the user sees, it needs to be well formed so writing it is important. It's not a user story. Well-formed Task Scenario make smoother tests.

So: Who will you test? (recruit) - What will you test - What style of test will you do?



Recruiting considerations:

- Screen on demographics (money they make, where they live...)? (Often not what we want to sort on)
- Screen on psycho-graphics e.g. cognitive background (familiar with this type of app)
- Get a mix of user types (not only actual users, but also prospective users)
- Use a professional recruiter

Need to screen the users to make sure we have a good mix.

Sequencing task:
The first task should be an orientation task, then run your scenario from easy to difficult with difficult tasks in the middle of the Test Plan.
Tip: After each user test, ask the user if the tasks were realistic - this will help give you corrective feedback. Writing good test tasks takes time and practice.

Usability Testing and Moderating Skills

Why do we test by scenario?
- More realism with real world behaviour.
- To ensure a smooth and seamless experience for the test participant.
- So we don't give away what we are testing.
- So **we don't lead the witness,** we need to be a bit obscure and not use the vocabulary of the product.
- Scenario simulate real-life the closest.

Conduct a pilot test:
- Always conduct a live fire test the day before to iron out any issues.
- For mobile testing: always test reception in the room or the area you will be testing in (with the target device)
- Make sure lightning is ok for the screen of the device.

Why think aloud?
Asking users to think aloud helps you track their thought patterns. This becomes the running commentary on task performance (which you observe and take notes on).

We are observers, we do not talk. Even if the user is asking us something, if we do not respond they will usually move on. We can say ok, or hanhan, but not good or bad. Need to stay neutral. You can use reverse questions (is this what you were expecting?).

What we measure:
1) Success Rate as percentage of task completion (100% 50% 0%)
2) number of errors per task
3) number of confusions per task.
These are the main metrics we capture on every usability test. Metrics give you a common fact-based description of user/task performance upon which to make informed design decisions.

Suggested Sample Sizes for research:
Corporate usability research:
- Surveys (phone and web) = 240 / 1000+
- Focus Groups = 15 -20 (depends on audience segments involved and goals of study)
- Usability Testing = 10 - 15 participants
- Field Studies = 15 - 40 participants
- Card Sorting = 15 - 30 (higher is better since card sorting uses the statistical method of cluster analysis)
Academic Usability Research:
Samples are usually larger depending on size and scope and research objectives (e.g. 15 users per segment or 40-100 users in a usability test)

Mobile UX testing strategy:
Preparing for the Test: record make/model and build of phone software. Pre-test, ask users about website vs app difference. Phone charger - tell users to bring one (or have extra on backup) Have a plan B (device, location).

Facilitating the Test: Stationary or in context? If native device, testing with distractions (texts, ringer)? Tell users you may have to calibrate camera mid-test and show them what to expect.
Sharing with stakeholders: How will you record and share? Screenshot capture? one way mirror or 121?

Conduct paper prototyping

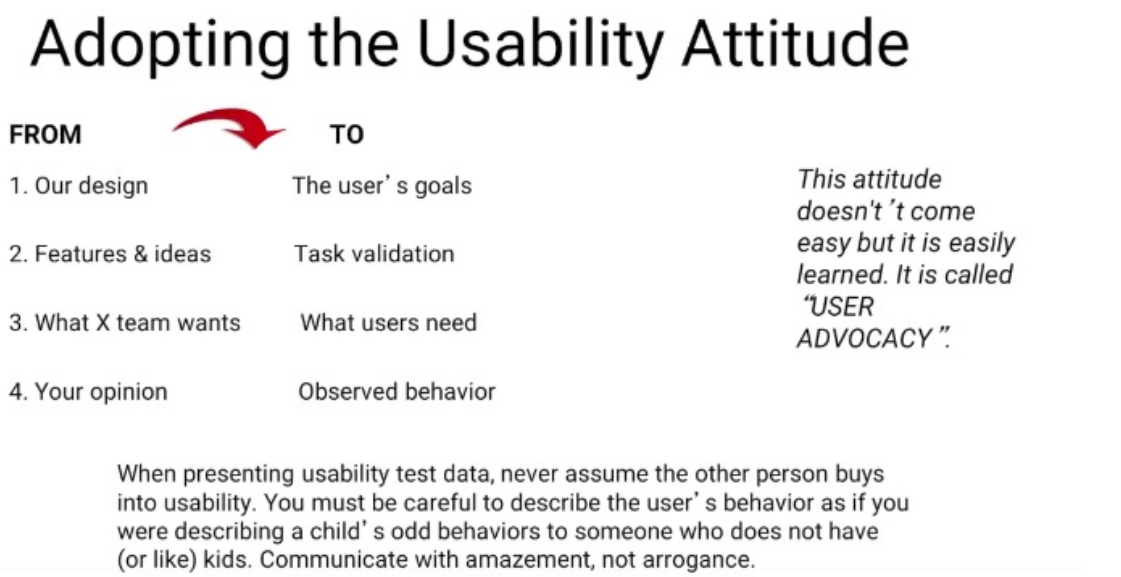
Print out UI elements for all state changes testings.
Tell user you will be the computer and feed her screens or changes.
Have user use a pen or finger to point to area on the protoype. For ipad: use PDF prototype.

Low-Fi Prototyping Tools:

Build interactivity into your prototypes (for mixed fidelity testing):
Axure, Flairbuilder, protoshare, balsamiq, mockingbird
For mobile:
Keynotes KungFu, Keynotopia, Proto.io

Reporting On Your Test Plan

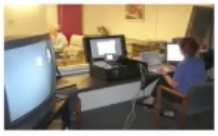
Ensuring Rigor in Testing



Factors	Ways to ensure rigor in usability test
Credibility (internal validity)	✓Write scenarios to reflect the user's perspective. ✓Use think-aloud protocol to get user's direct impressions. ✓Use direct observations of user actions (versus indirect reports such as surveys). ✓Check out conclusions or observation with the user through protocols such as teach-back, guided recall, or post-scenario interviews.
Transferability (external validity)	✓Select participants based on a set of criteria that defines the end-user of the product. ✓Write scenarios that involve realistic tasks. ✓Use platforms and equipment typical of what end-users have. ✓Use actual or realistic prototypes of packaging and documentation that will be supplied with the product.
Dependability (reliability)	✓Use an observation team with multiple perspectives. ✓Retest to verify improvements. ✓Use an external consultant to review process and findings.

Source: Hughes (1999)

Moderating Skills



- Proper screening and recruiting
- Moderation with rapport
- Curtailing non-verbal responses
- Managing user interaction
- Observation without hidden agendas

"Usability Testing is not a clinical thing. When you show the user you are relaxed, they will also relax (and tell you more!) ".

use a local moderator is you don't have the culture of the user where you are doing the usability testing.

Stay in Control as moderator

- ✓ Screen and recruit properly
- ✓ Give clear instructions (think aloud)
- ✓ Moderate with rapport
- ✓ Observe key issues
- ✓ Wait to see patterns emerge
- ✓ Mind your Non-verbal responses

INSTRUCTIONS TO THE USER:

1. Thank you for helping us.
2. Purpose of evaluation (here to get your feedback on design. No right or wrong answers).
3. Use these tasks (Test Packet)- 17 pt font task scenarios Word doc for user) to find the answer on the site.
4. Think Out Loud so I can follow along (tell me what you are thinking and feeling).
5. Go on to next task when ready (move at your own pace when you feel you found the answer).
6. Questions afterward when you are done. "If have questions won't be able to answer but will do my best at end to answer them".

Then don't answer them. Just keep saying "You helped us identify what we were looking for". or "I am not sure if that is the right answer or not".

Avoiding User Emotion Traps

Avoid User Emotion Traps

Trap:	Solution:
Wants to know if they are doing a good job	Say "OK" or "You're helping us identify what we are looking for". Repeat.
Wants to get answers right	Say "There are no right or wrong answers" or "You helped us identify what we were looking for".
Stubborn or quiet	Ask user to "Remember to Think Out Loud so I can follow along".
Waiting for permission to move on to next task.	Tell user to move at her own pace during test. Indicate before test in instructions.
Mumbles or speaks unclearly	Ask user to speak louder. Repeat.
Panicking for help	Let them struggle, within reason. Say nothing or just say "OK". Repeat.

Test Prep checklist

Item	Action	Date Due	Owner	Completed
Define Business Objectives and Pre-screener Needs	Usability Test Business Objectives gathered and Pre-screener Needs			
Determine user task expectations	Confirm user needs (confirm right tasks) Set up 3-6 30 min. phone interview or field visit			
Create test tasks and test plan	Create tasks that support the business and test objectives			
Create pre-Screener	Pre-screener sent to Business Groups			
Create Screener	Screener for recruiter (confirm right person)			
Pre-Recruit	Use contacts and existing relationships to leverage recruit			
Recruit users	Get 15 users (10 for test)			
User attends facility	Recruiter confirms attendance			
Conduct test (Pre-test questionnaire)	TEST			
Conduct test (User uses site to complete tasks)	TEST			
Conduct test (Post-test questionnaire)	TEST			
Test report (Posted online)	Analysis and Test write-up			
Design recommendations and support follow-up	Design mock-ups and prioritization of issues			
Highlight video editing	Edit and Compile Video Highlights			

Conduct Dry Run day before

Record a blank to ensure audio/video working

Bring back up device

Task Scenarios from your Test Plan. Below:17 pt "User Packet" for the user with tasks...

Reporting Tips

Remember to:

- Prioritize findings
- Report actual and observe behavior
- Avoid micro-usability
- Stay focused on business objectives/ technology constraints
- Suggest appropriate design recommendations

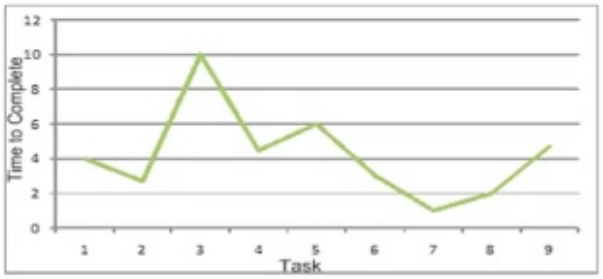
The more usability testing you do, the more confident you will be at interpreting findings and your analysis.

Report Writing Advice

1. Keep it short (10-15 pages)
2. Use screen shots to communicate findings
3. Include positive findings in your reports
4. Include an executive summary
5. Include user quotes– verbatim
6. Avoid harsh words or tones
7. Distinguish user opinion vs. observer opinion
8. Keep descriptions relevant
9. Include analyst or competitive analysis where relevant
10. Try to report out immediately after the test– as an email then followed by a report

Example Test Report

Time on Task

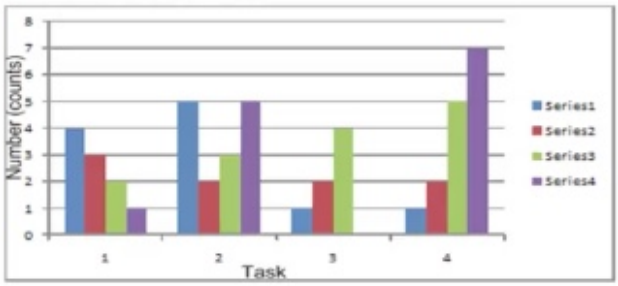


Tasks

Task 1: Find the animal in the toy store.
Task 2: Where are the stores located in New York?
Task 3: What is the name of the president?
Task 4: etc.

(2) Comments Add a comment

Errors/Confusions



TOTALS:

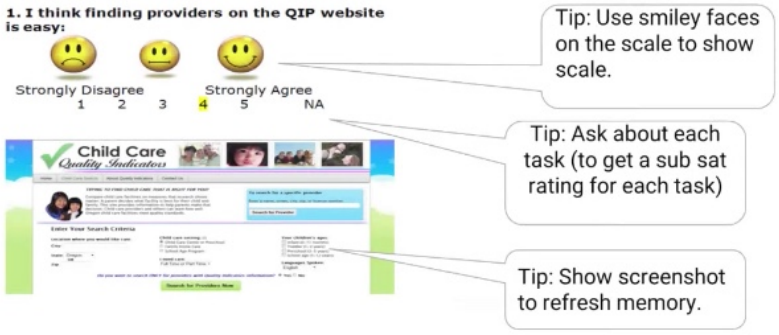
POSITIVE	15	CONFUSION	8
NEGATIVE	7	ERROR	12

(2) Comments Add a comment

5 point scale

Use a 5-point 'Likert Scale' for measuring subjective satisfaction.

Example (perceived difficulty being elicited)



Measuring Satisfaction

1. It helps me be more effective.

1 2 3 4 5

Strongly agree Strongly Disagree

2. It is easy to use.

1 2 3 4 5

Strongly agree Strongly Disagree

3. I easily remembered how to use it.

1 2 3 4 5

Strongly agree Strongly Disagree

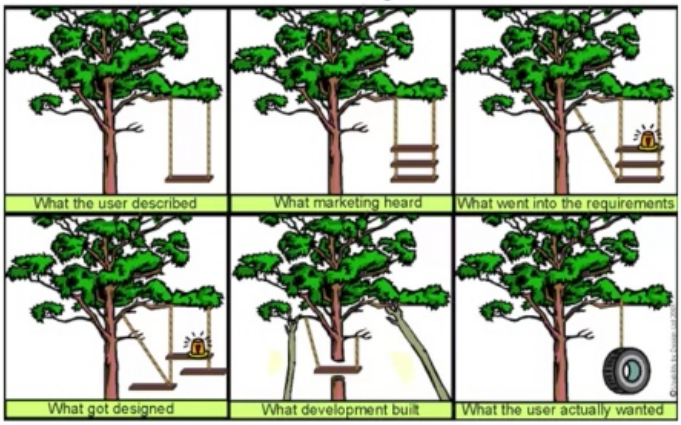
Subjective Satisfaction:

9/10 users felt the site was easy to use
7/10 users thought the home page was cluttered
5/10 users felt it was easy to get around the site

The Politics of Usability

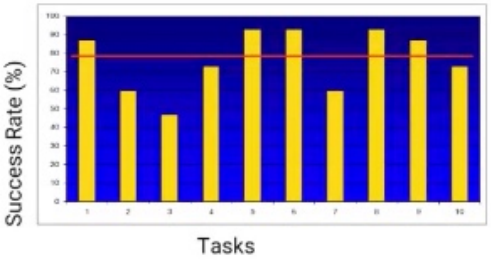
1. **Build Alliances**
2. **Avoid confrontations**
3. **Pick your battles**
4. **Position yourself as an ally**
5. **Build trust by being completely open**
6. **Document and sell successes to management**

Source: Ralph Molich, 2000



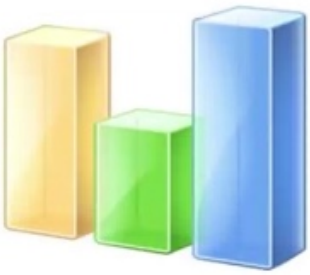
See Book: The Politics of Usability: A Practical Guide to Designing Usable Systems in Industry by Lesley Trenner and Joanna Bawa.

Metrics Capture is Vital



1. Make usability recommendations concrete
2. Ground teams in reality (probability-thinking)
3. Help iterate and validate design concepts
4. Provide objectivity to design debates
5. Guide fact-based design decisions

Communicate Clearly



- ✓ Watch for the feelings of others
- ✓ Be aware that some people might take the criticism personally (even culturally avoided in Asia)
- ✓ Be flexible with your recommendations. If there is another way to solve the problem– be open to that.
- ✓ If you are not sure, say so.
- ✓ Separate your observations from your thoughts.
- ✓ Be aware of your own bias and agenda (check agendas at the usability lab door).

Tip: Be careful with qualitative data– it is less understood than quantitative. Be sure to provide a balance of statistics (quantitative) and observation (qualitative) since many more people understand numbers over insights.