Human-Computer Interaction

CS 449 | 649 – F26

Week 11: Nov 23 – Nov 27

High-fidelity prototyping evaluation:

- **Heuristic evaluation**
- **Cognitive walkthrough**
- **Usability testing**
- **Design critique**

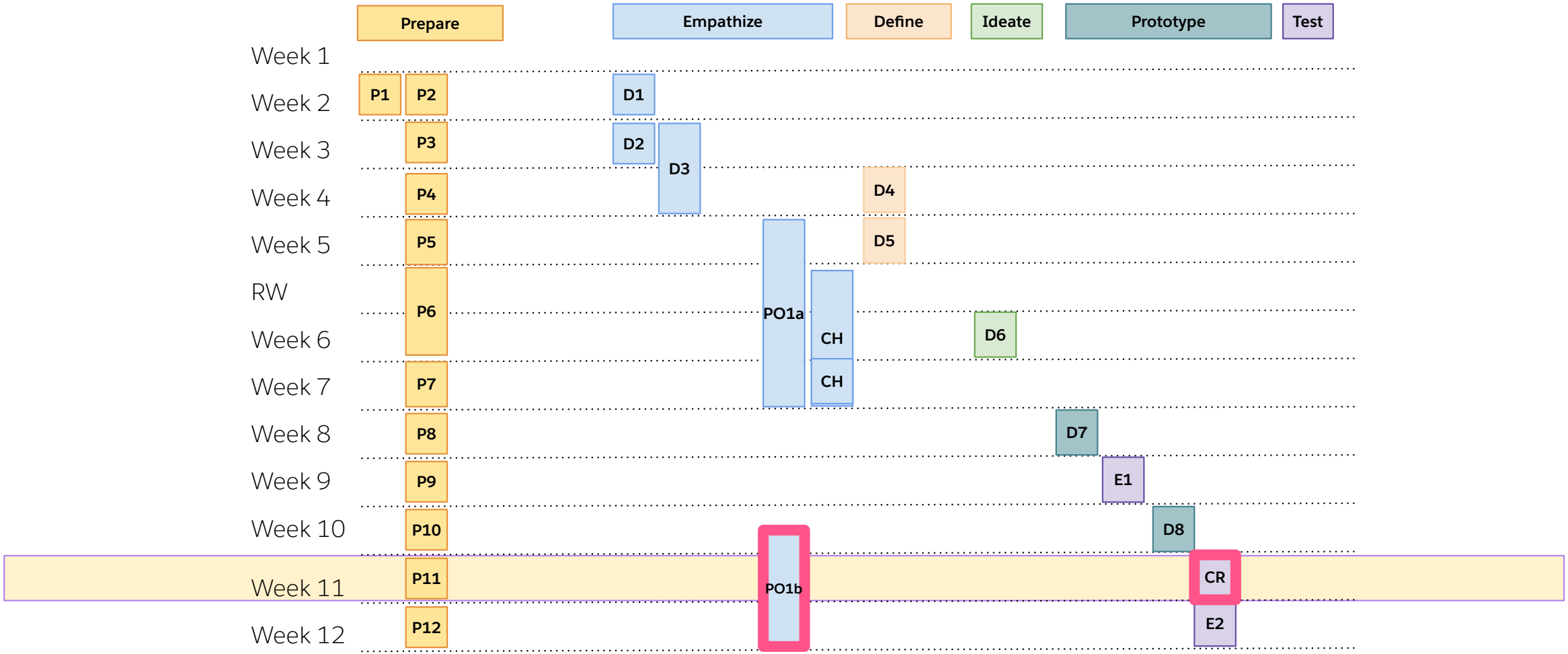
UNIVERSITY OF WATERLOO | DAVID R. CHERITON SCHOOL OF COMPUTER SCIENCE
WATERLOOHCI

* Updated Nov 23 2026

Week 10 Recap

Visual design and high-fidelity prototyping

Project Deliverables Build on Each Other

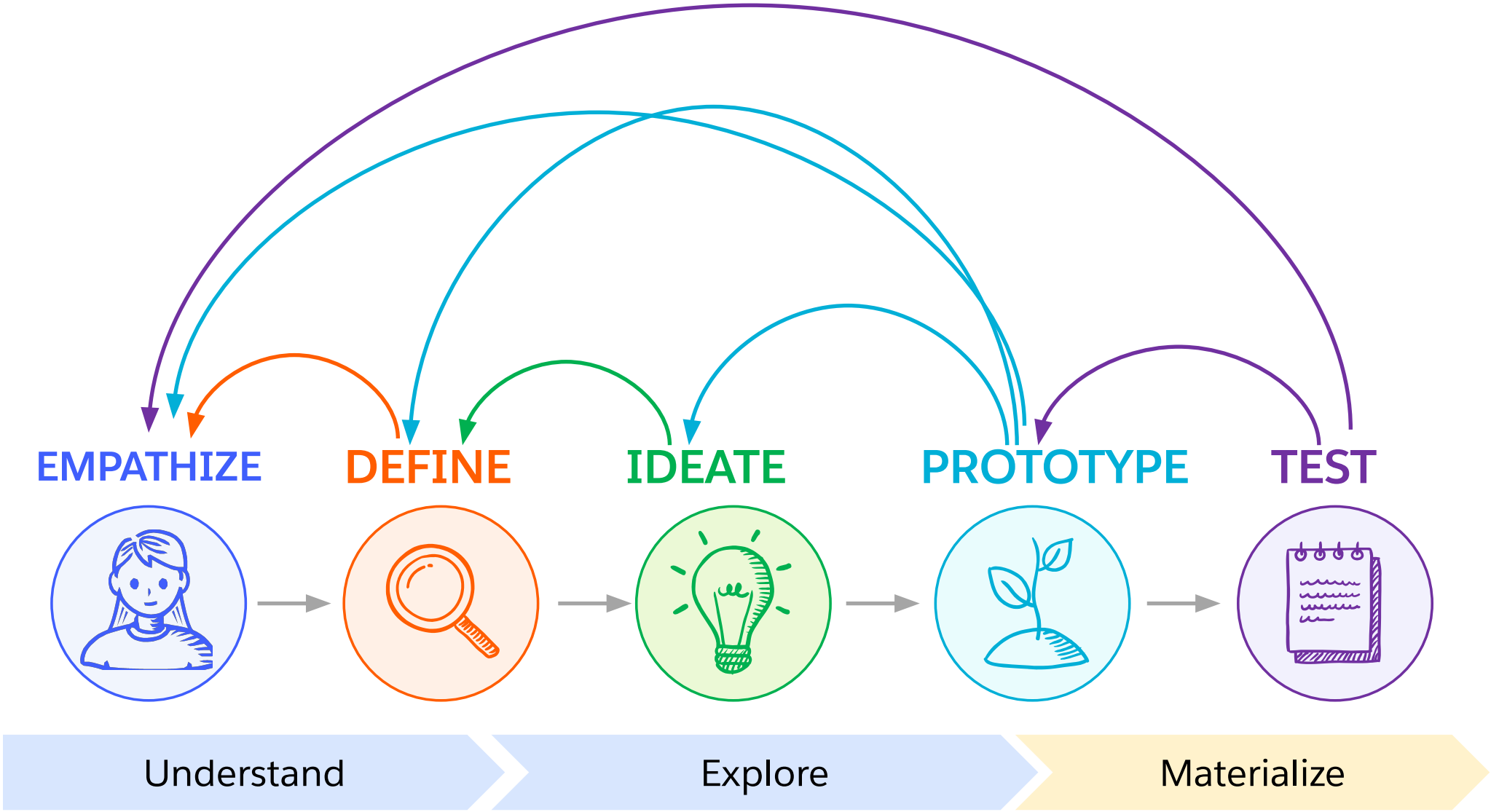


Week 11 To-Dos

- ✅ Complete design notebook entry for D8: High-Fidelity Prototyping
- ✅ Conduct 1-2x PO1b Interviews
- ✅ Complete a T2 Team reflection entry
- ✅ P11: Prepare for Design Critique

Evaluating High-Fidelity Prototypes

Heuristic Evaluation, Cognitive Walkthrough, Usability Testing, Design critique


EMPATHIZE
DEFINE
IDEATE
PROTOTYPE
TEST
Understand
Explore
Materialize

UX Evaluation Types

There are different kinds of evaluation you can run.
These aren't entirely mutually exclusive:

User Evaluation: People representative of your target personas perform tasks with interface ... e.g. usability test, A/B testing, controlled experiment

Expert Evaluation: Experts use established guidelines to evaluate interface
... e.g. heuristic evaluation, cognitive walkthrough

Technical Evaluation: A system runs automated tests on the interface
... e.g. accessibility report

UX Evaluation Types

	Heuristic Evaluation	Cognitive Walkthrough	Usability Testing
Who	One or more professionals evaluators (e.g., someone with design knowledge)	One or more professional evaluators (e.g., someone with design or product knowledge)	People representative of your target personas
What	Look for issues via heuristic categories (e.g., Neilson's 10)	Perform a list of specific tasks	Use the system in a realistic situation
Why	Identify problems and assess the systems usability	Evaluate the learnability of a systems from the perspective of first time use	Gain knowledge from those likely to use the system

📖 Sepideh Yazdi: Cognitive walkthrough vs. heuristic evaluation vs. usability testing (2022)

Heuristic Evaluation

a.k.a. “Expert Evaluation”

UX Evaluation Types

	Heuristic Evaluation	Cognitive Walkthrough	Usability Testing
Who	One or more professionals evaluators (e.g., someone with design knowledge)	One or more professional evaluators (e.g., someone with design or product knowledge)	People representative of your target personas
What	Look for issues via heuristic categories (e.g., Neilson's 10)	Perform a list of specific tasks	Use the system in a realistic situation
Why	Identify problems and assess the systems usability	Evaluate the learnability of a systems from the perspective of first time use	Gain knowledge from those likely to use the system

📖 Sepideh Yazdi: Cognitive walkthrough vs. heuristic evaluation vs. usability testing (2022)

Heuristic Evaluation

A small set of evaluators (experts) examine the interface and judge its compliance with recognized usability principles (the "heuristics")

... also referred to as "**Expert Evaluation**"

It requires multiple experts to do the evaluation

... an expert means someone who is familiar with **usability principles**

... no "actual users" are involved

In theory can be done at any stage of design

... most effective at later stages, such as high-fidelity prototype

Can cost less and take less time than testing with “actual users”

📖 Interaction Design Foundation: What is Heuristic Evaluation?

Heuristics are a “rule of thumb”

They are broad design guidelines to apply to a variety of interfaces…

If you violate these principles, you should do a lot of testing to validate that your decisions is effective…

Intended to apply across all types of technology (e.g. website, video game, etc.), but rules can be adapted to the medium…

Heuristics have been the same for 25+ years because they are broad…

📖 Interaction Design Foundation: What is Heuristic Evaluation?

Jakob Nielsen's 10 Usability Heuristics

1 Visibility *of* System Status

Designs should *keep users informed* about what is going on, through appropriate, timely feedback.



Interactive mall maps have to show people where they currently are, to help them understand where to go next.

2 Match between System and the Real World

The design should speak the users' language. Use words, phrases, and concepts *familiar to the user,* rather than internal jargon.



Users can quickly understand which stovetop control maps to each heating element.

3 User Control *and* Freedom

Users often perform actions by mistake. They *need a clearly marked "emergency exit"* to leave the unwanted action.



Just like physical spaces, digital spaces need quick "emergency" exits too.

4 Consistency *and* Standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. *Follow platform conventions.*



Check-in counters are usually located at the front of hotels, which meets expectations.

5 Error Prevention

Good error messages are important, but the best designs carefully *prevent problems* from occurring in the first place.



Guard rails on curvy mountain roads prevent drivers from falling off cliffs.

6 Recognition Rather Than Recall

***Minimize the user's memory load* by making elements, actions, and options visible. Avoid making users remember information.**



People are likely to correctly answer "Is Lisbon the capital of Portugal?".

7 Flexibility *and* Efficiency of Use

Shortcuts — hidden from novice users — may *speed up the interaction* for the expert user.



Regular routes are listed on maps, but locals with more knowledge of the area can take shortcuts.

8 Aesthetic *and* Minimalist Design

Interfaces should not contain information which is irrelevant. Every extra unit of information in an interface *competes* with the relevant units of information.



A minimalist three-legged stool is still a place to sit.

9 Recognize, Diagnose, *and* Recover from Errors

Error messages should be expressed in plain language (no error codes), precisely indicate the problem, and constructively suggest a solution.



Wrong-way signs on the road remind drivers that they are heading in the wrong direction.

10 Help *and* Documentation

It's best if the design *doesn't need* any additional explanation. However, it may be necessary to provide documentation to help users complete their tasks.



Information kiosks at airports are easily recognizable and solve customers' problems in context and immediately.

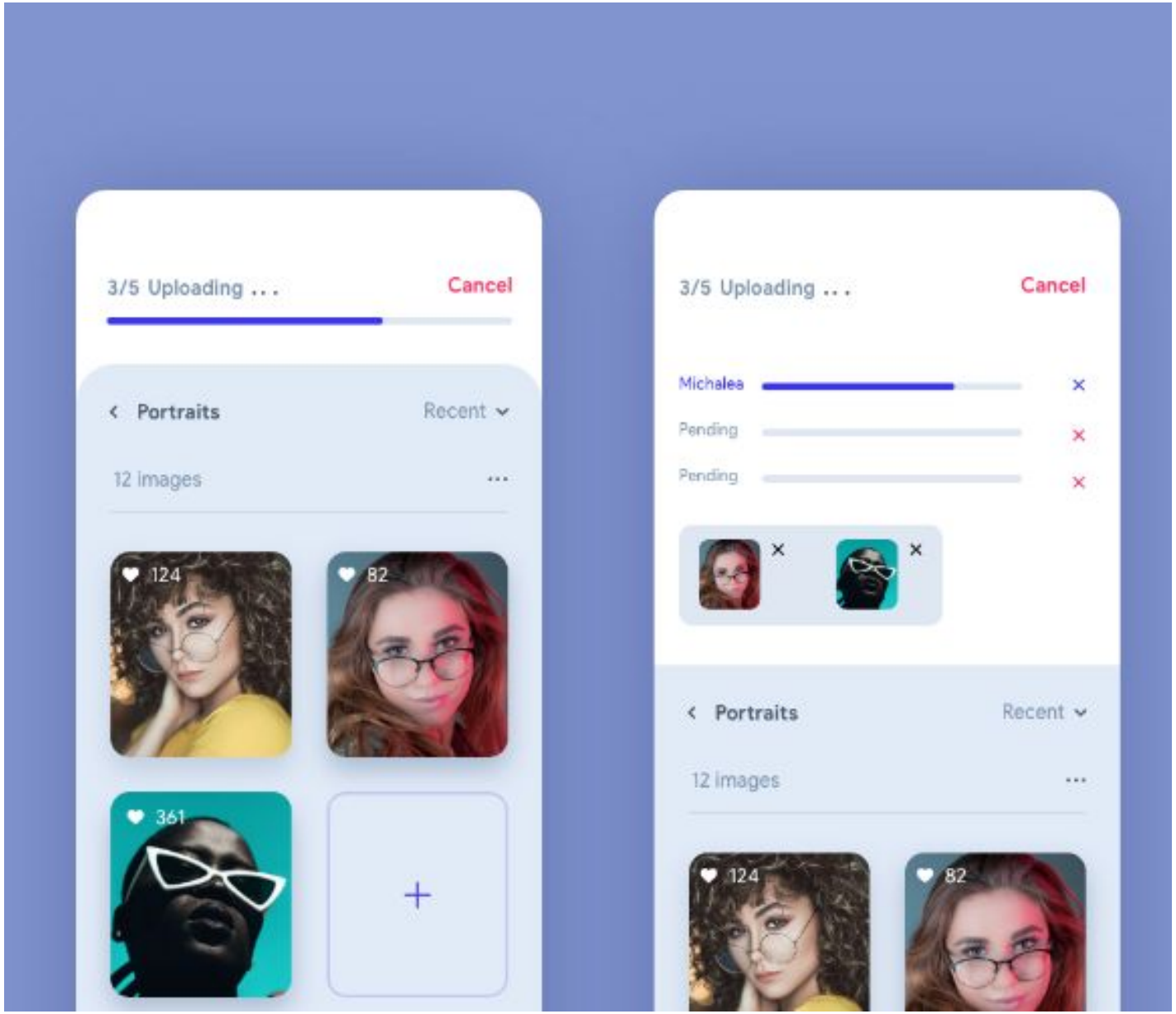
📖 Nielsen Norman Group: Usability Heuristics Summary

1 Visibility *of* System Status

Designs should *keep users informed* about what is going on, through appropriate, timely feedback.



Interactive mall maps have to show people where they currently are, to help them understand where to go next.



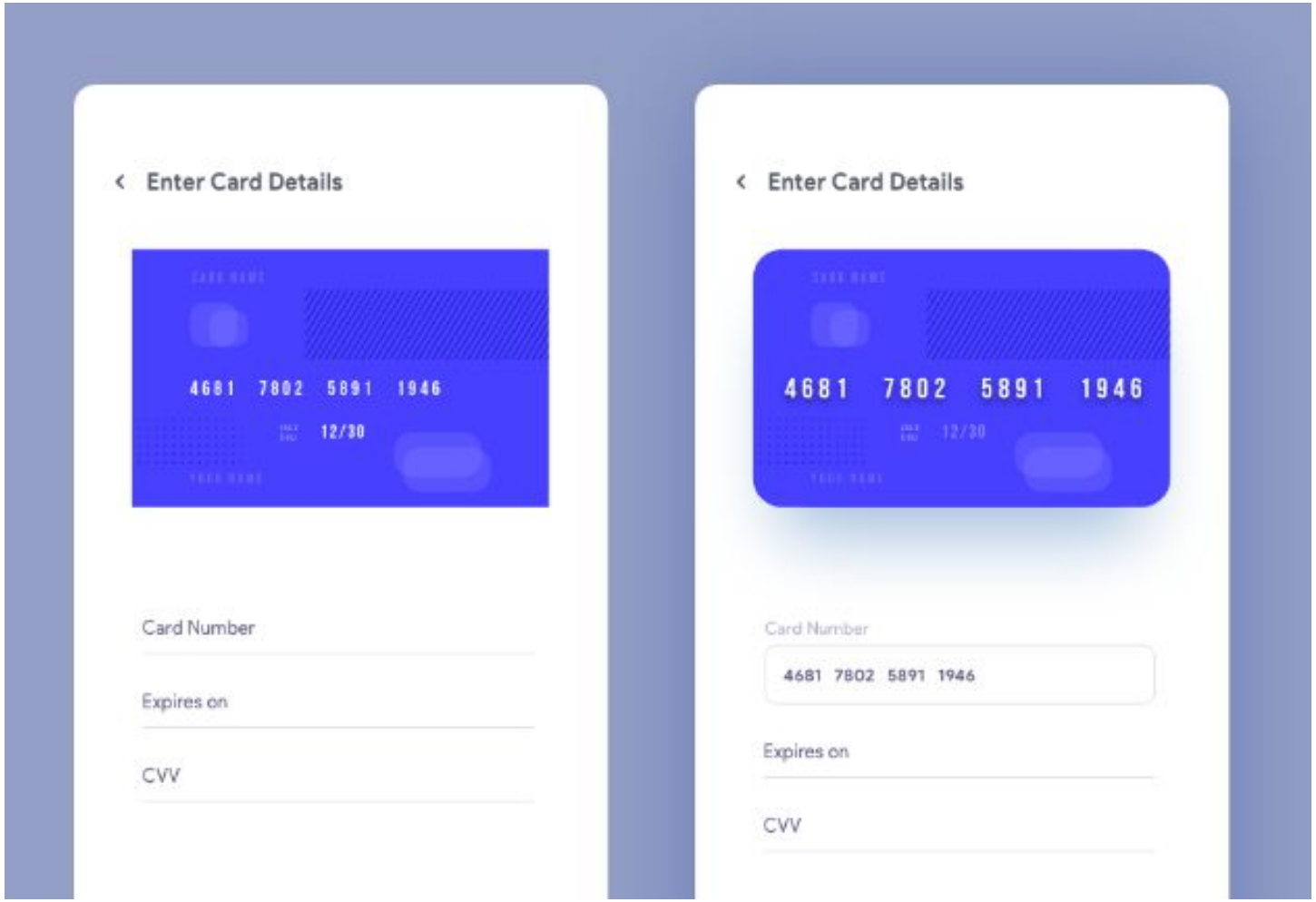
📖 Akshay Devazya: 10 usability heuristics with examples (2019)

2 Match between System and the Real World

The design should speak the users' language. Use words, phrases, and concepts *familiar to the user*, rather than internal jargon.



Users can quickly understand which stovetop control maps to each heating element.



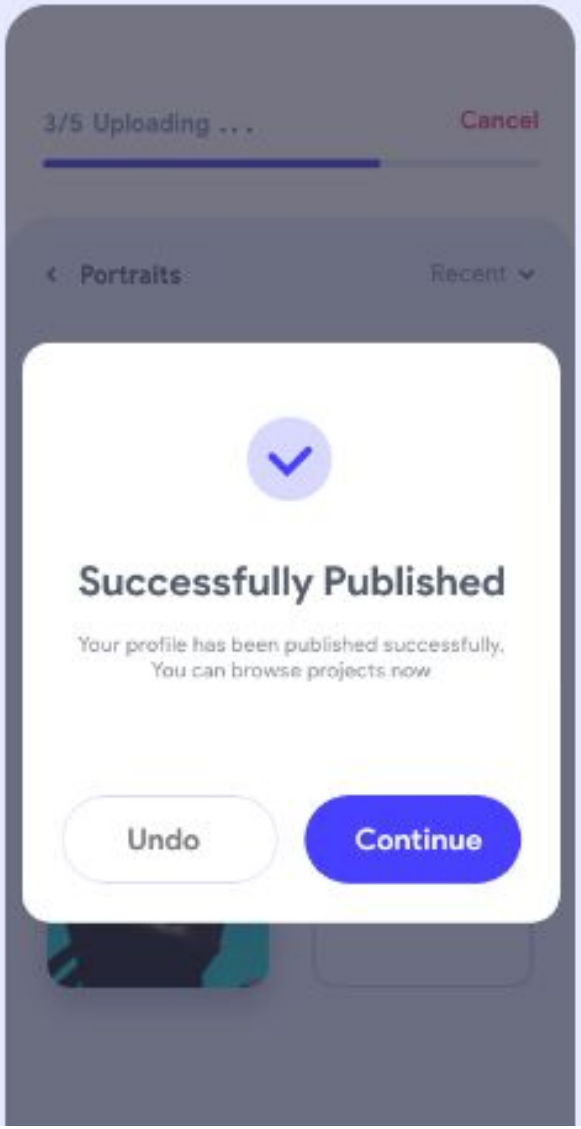
📖 Akshay Devazya: 10 usability heuristics with examples (2019)

3 User Control *and* Freedom

Users often perform actions by mistake. They *need a clearly marked "emergency exit"* to leave the unwanted action.



Just like physical spaces, digital spaces need quick "emergency" exits too.



📖 Akshay Devazya: 10 usability heuristics with examples (2019)

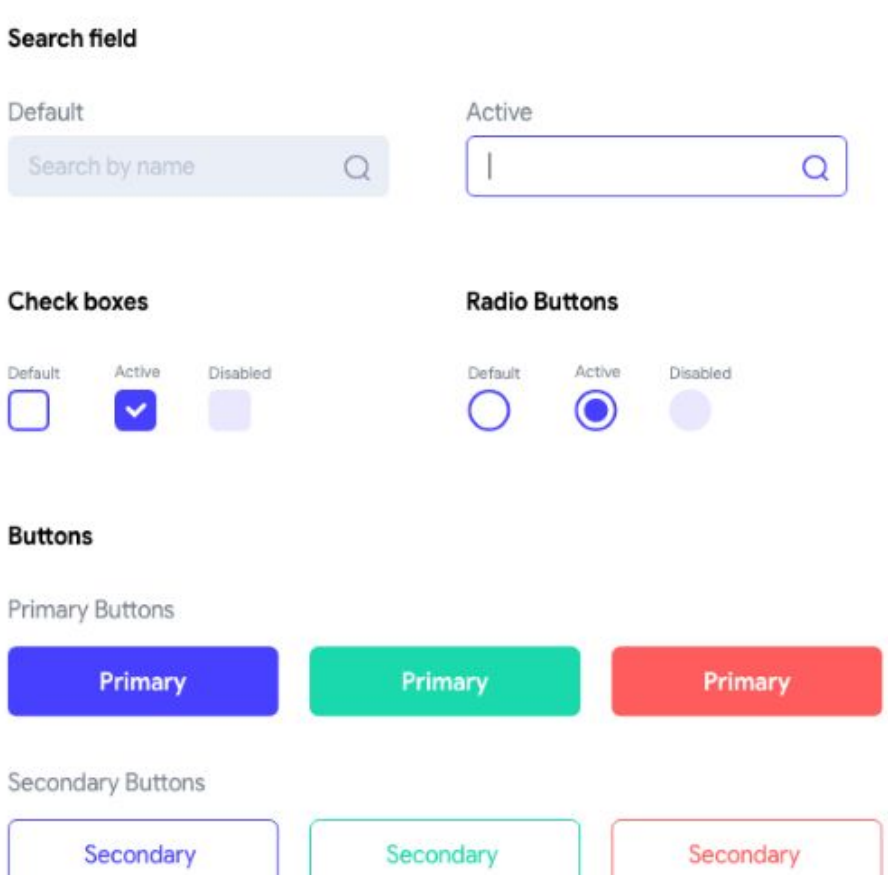
4 Consistency *and* Standards

Users should not have to wonder whether different words, situations, or actions mean the same thing.

Follow platform conventions.



Check-in counters are usually located at the front of hotels, which meets expectations.



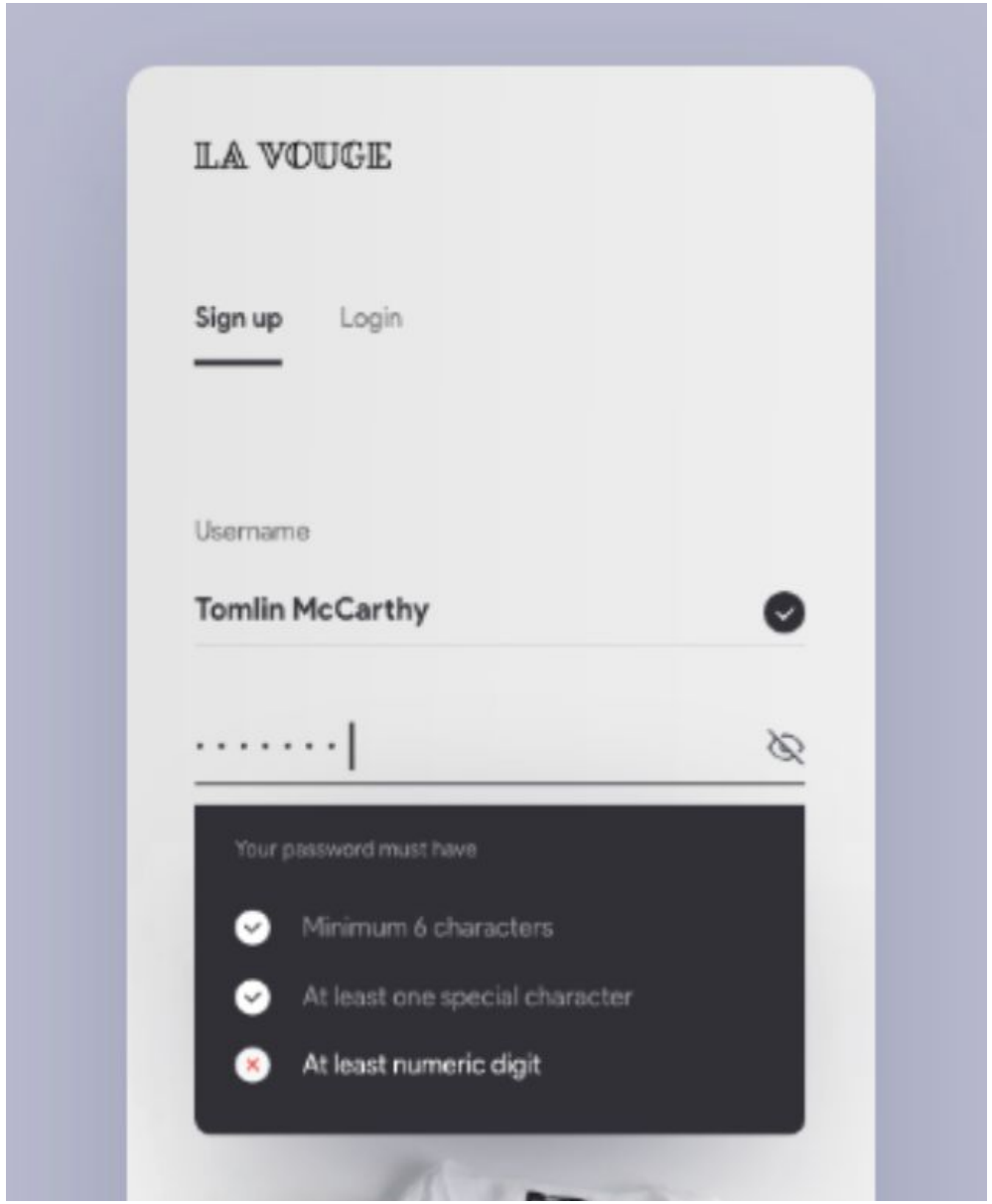
📖 Akshay Devazya: 10 usability heuristics with examples (2019)

5 Error Prevention

Good error messages are important, but the best designs carefully *prevent problems* from occurring in the first place.



Guard rails on curvy mountain roads prevent drivers from falling off cliffs.



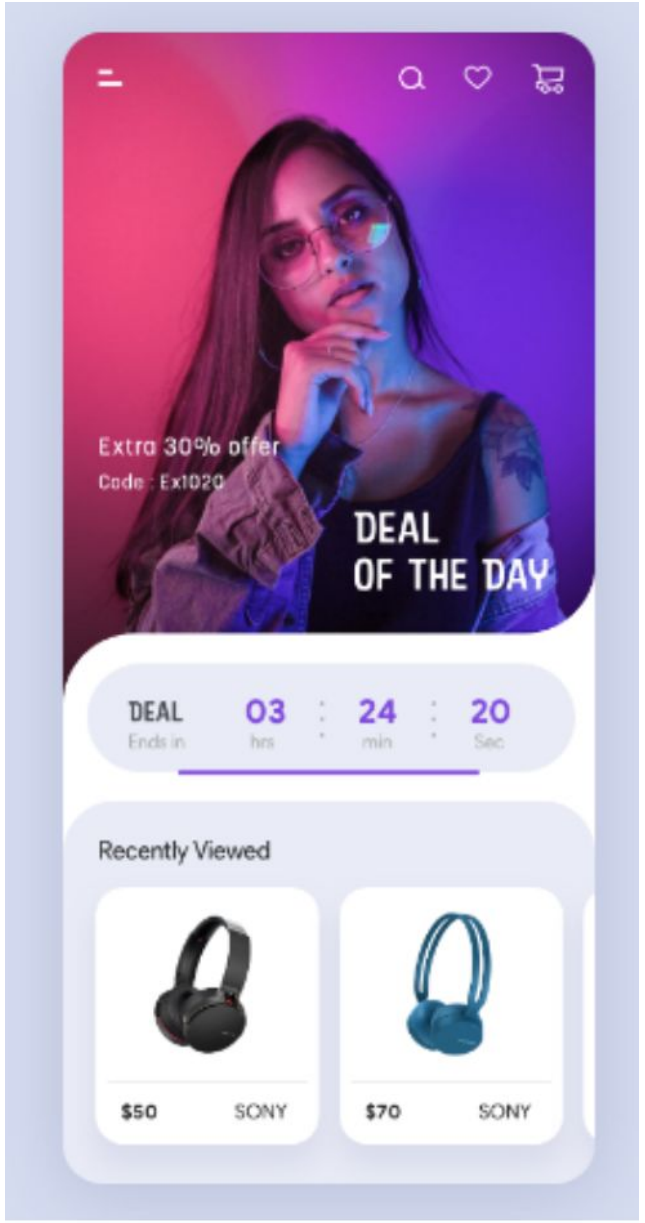
📖 Akshay Devazya: 10 usability heuristics with examples (2019)

6 Recognition Rather Than Recall

Minimize the user's memory load **by making elements, actions, and options visible. Avoid making users remember information.**



People are likely to correctly answer “Is Lisbon the capital of Portugal?”.



📖 Akshay Devazya: 10 usability heuristics with examples (2019)

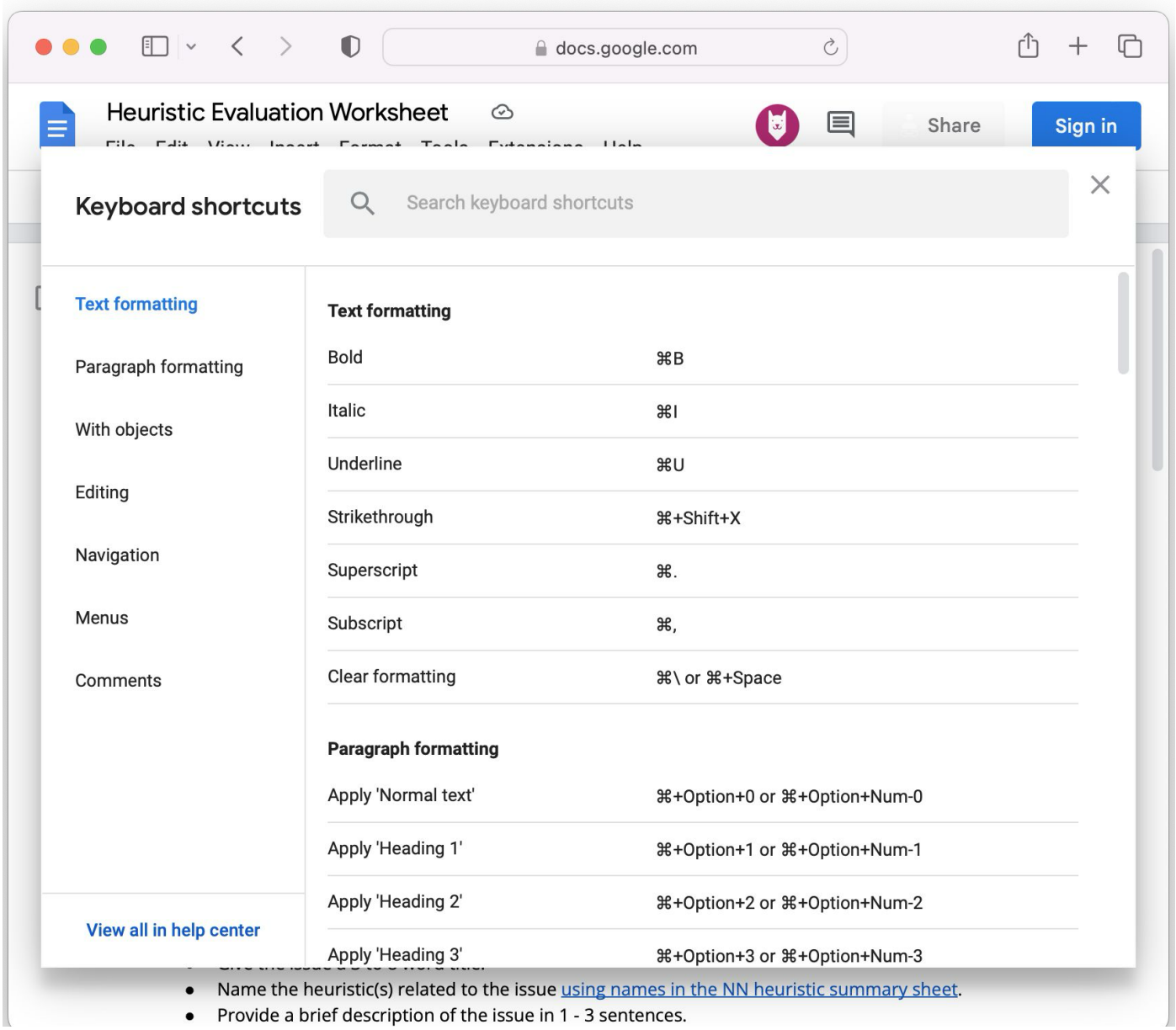
20

7 Flexibility *and* Efficiency of Use

Shortcuts — hidden from novice users — may *speed up the interaction* for the expert user.



Regular routes are listed on maps, but locals with more knowledge of the area can take shortcuts.



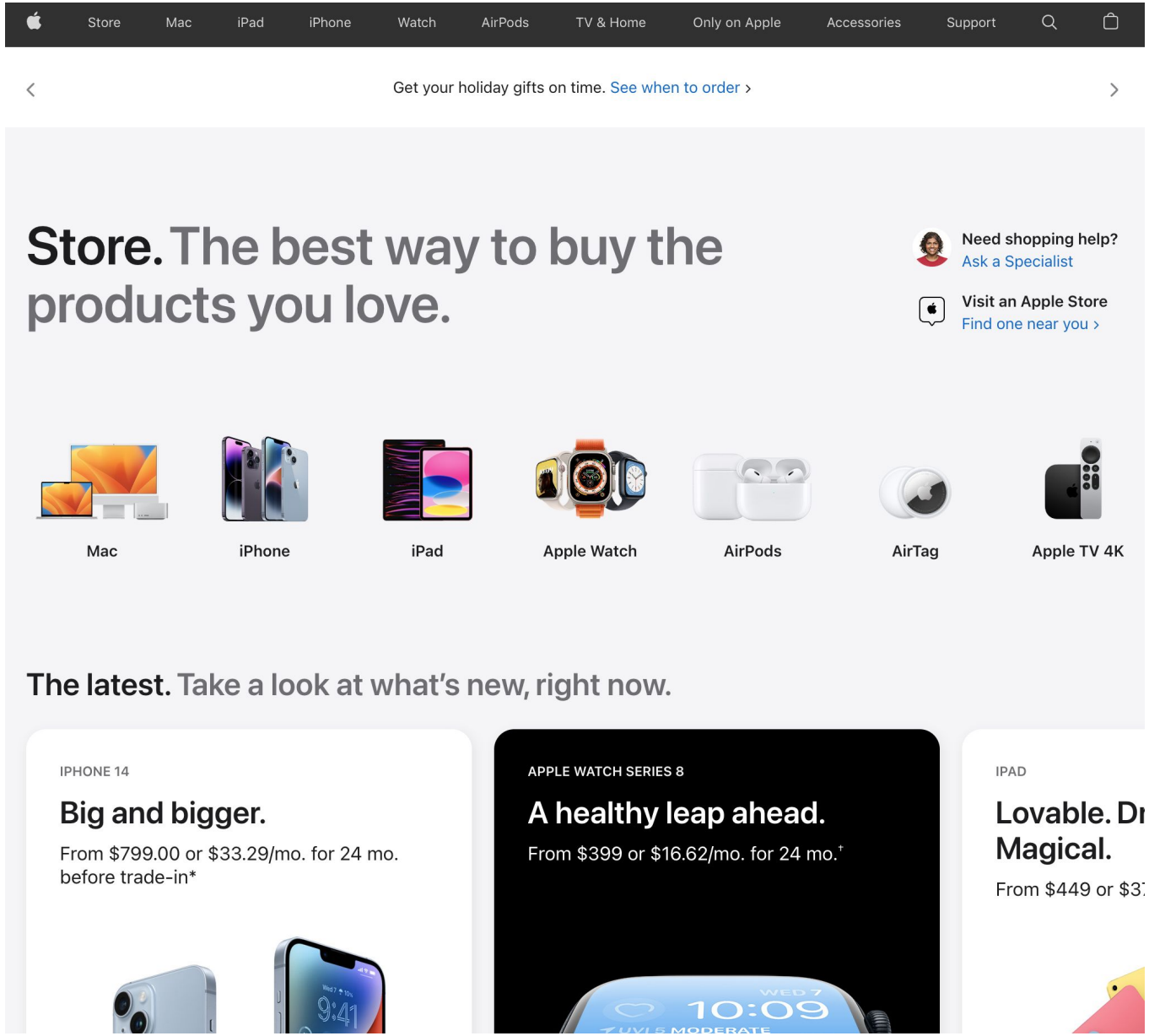
📖 Akshay Devazya: 10 usability heuristics with examples (2019)

8 Aesthetic *and* Minimalist Design

Interfaces should not contain information which is irrelevant. Every extra unit of information in an interface *competes* with the relevant units of information.



A minimalist three-legged stool is still a place to sit.



📖 Akshay Devazya: 10 usability heuristics with examples (2019)

9 Recognize, Diagnose, *and* Recover from Errors

Error messages should be expressed in plain language (no error codes), precisely indicate the problem, and constructively suggest a solution.



Wrong-way signs on the road remind drivers that they are heading in the wrong direction.



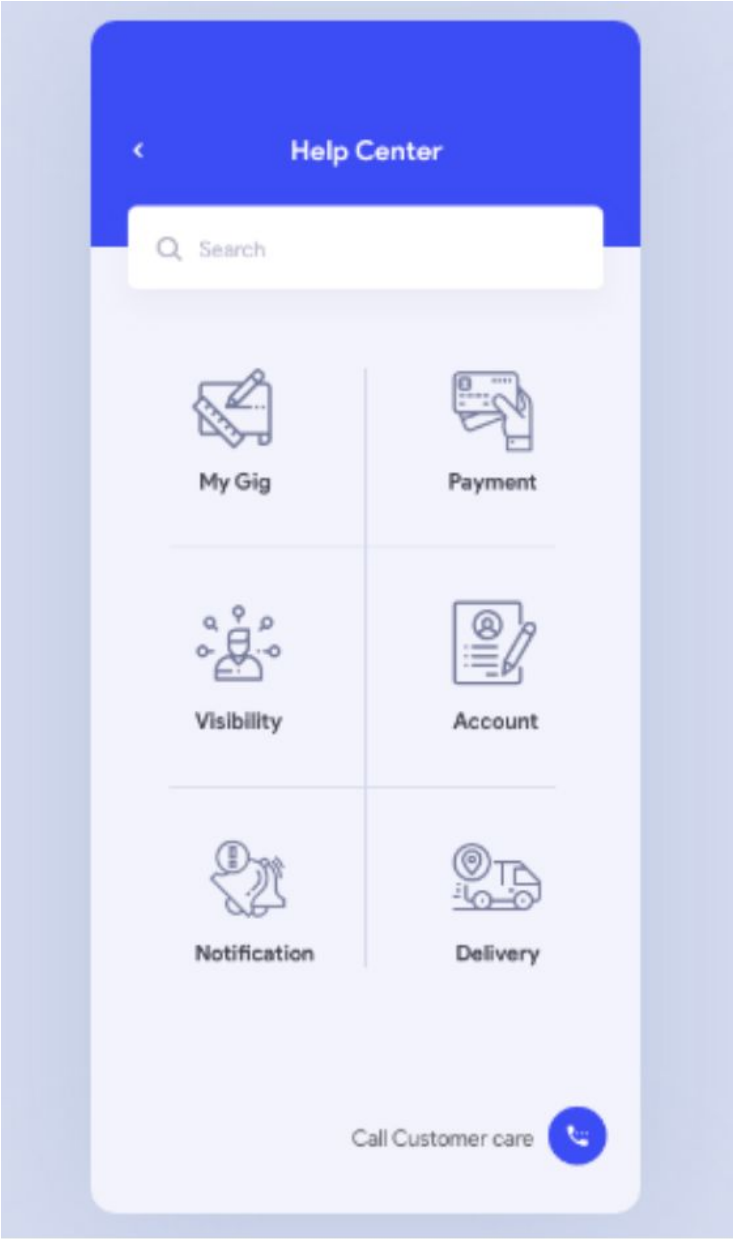
📖 Akshay Devazya: 10 usability heuristics with examples (2019)

10 Help *and* Documentation

It's best if the design *doesn't need* any additional explanation. However, it may be necessary to provide documentation to help users complete their tasks.



Information kiosks at airports are easily recognizable and solve customers' problems in context and immediately.

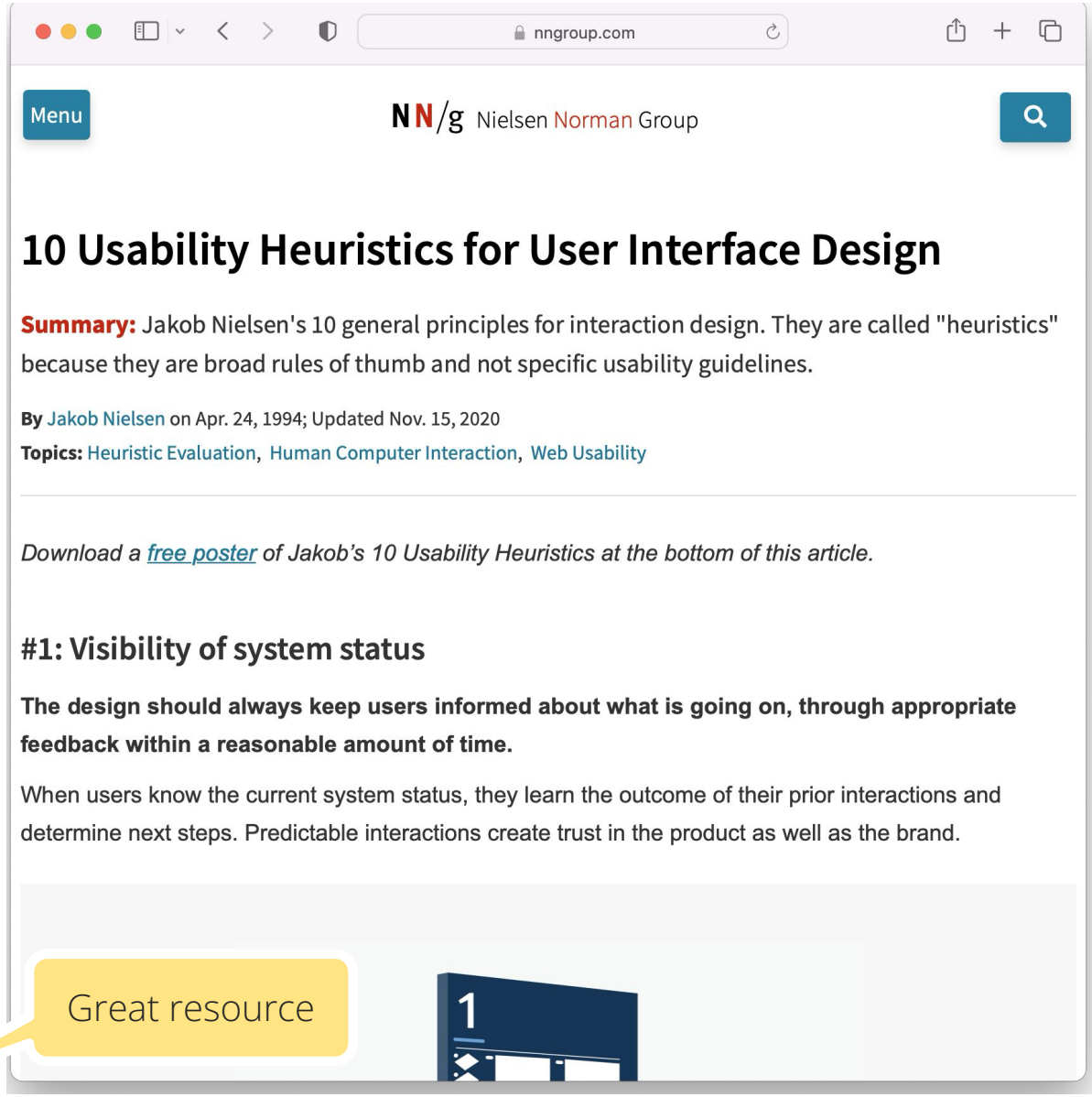


📖 Akshay Devazya: 10 usability heuristics with examples (2019)

Heuristics Resource

For *each* of the 10 heuristics:

... definition
... examples
... tips
... short explanatory video
... link to full article



📖 Jakob Nielsen (Nielsen Norman Group): 10 Usability Heuristics for User Interface Design (1994)

Heuristic Evaluation Methodology: Planning

Define what to test: which prototype, what scope (entire product, ***specific tasks***, etc.)

Provide background for evaluators:

... summary of product and related domain
... high level summary of personas emphasizing goals

Select 3 to 5 evaluators who have expertise in UX

Define the set of heuristics: e.g. Nielsen's 10 heuristics

Define a scale of issue severity codes:

[1] Cosmetic [2] Minor [3] Major **[4] Critical**

📖 Interaction Design Foundation: What is Heuristic Evaluation? • 📖 Kate Moran and Kelley Gordon: How to Conduct a Heuristic Evaluation (2023)

Methodology

The following are what the *evaluator* does

First walkthrough: use prototype and identify elements to examine closely

Second walkthrough: scrutinize individual elements using the heuristics and identify issues

For each issue:

- create an issue title
- list the related heuristic(s)
- describe the issue
- capture and annotate a screenshot
- assign a severity level code

Heuristic Evaluation Worksheet (Template)

CS 449/649

Heuristic Evaluation Worksheet

For: {{ team name, product name }}

By: {{ evaluator name }}

Background

{{ product background }}

Prototype and Tasks

{{ link to prototype }}

{{ description of tasks }}

Evaluation Instructions

Evaluate the prototype using Nielsen's 10 heuristics (see lecture notes for suggested resources).

When an issue is found:

- Give the issue a 3 to 8 word title.
- Name the heuristic(s) related to the issue using names in the NN heuristic summary sheet.
- Provide a brief description of the issue in 1 - 3 sentences.
- Paste a screencap showing the issue (crop as needed to focus on a specific area if that makes sense, use a simple annotation like red circle or arrow if possible).
- Assign a severity level by copying can pasting one of the labels below:
  [1] Cosmetic [2] Minor [3] Major **[4] Critical**

Issues Found

{{ issue title }}

{{ severity level }}

{{ heuristic name, heuristic name, heuristic name }}

{{ brief description of issue }}

{{ image }}

{{ issue title }}

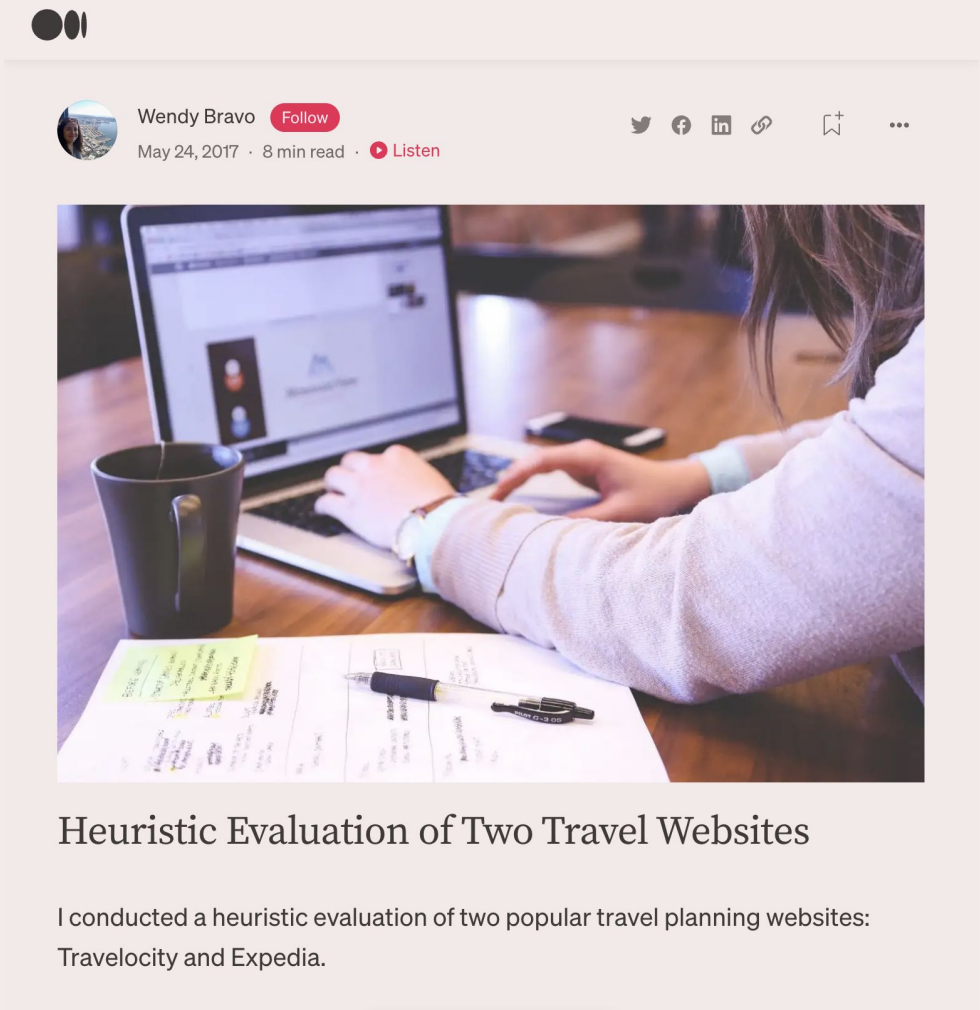
{{ severity level }}

{{ heuristic name, heuristic name, heuristic name }}

{{ brief description of issue }}

See **Resources** page on course website

Example - Heuristic Walkthrough of Two Travel Websites

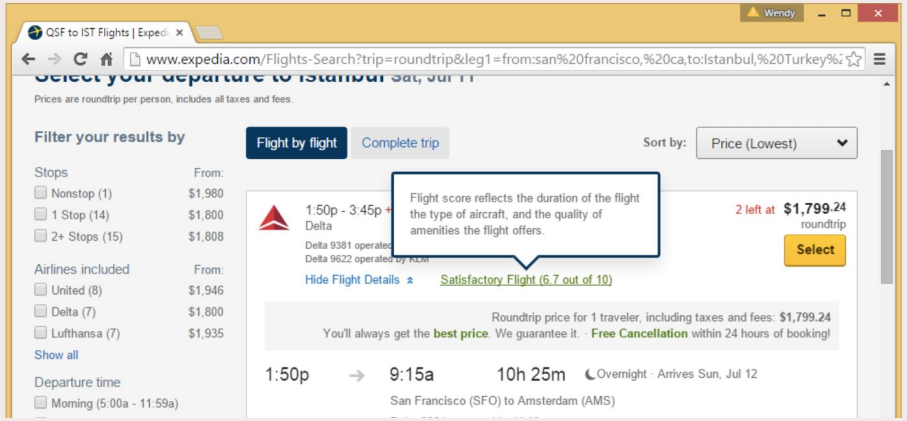


Second Issue/Violation: Insufficient information regarding their flight rating system.

Severity: 3

Heuristics violated: Help

Description: I noticed that my flight was rated as “Satisfactory Flight (6.7 out of 10).” I clicked to find out what they meant by that. There was some information offered, but it was too generic and not enough. I was left wondering which of the factors mentioned in the pop-up description lowered this flight’s score. This could be fixed by adding a “more info” link at the end of the description for users who want to find out more.



📖 Wendy Bravo: Heuristic Evaluation of Two Travel Websites (2017)

Cognitive Walkthrough

A task-based approach to identify potential usability problems

UX Evaluation Types

	Heuristic Evaluation	Cognitive Walkthrough	Usability Testing
Who	One or more professionals evaluators (e.g., someone with design knowledge)	One or more professional evaluators (e.g., someone with design or product knowledge)	People representative of your target personas
What	Look for issues via heuristic categories (e.g., Neilson's 10)	Perform a list of specific tasks	Use the system in a realistic situation
Why	Identify problems and assess the systems usability	Evaluate the learnability of a systems from the perspective of first time use	Gain knowledge from those likely to use the system

📖 Sepideh Yazdi: Cognitive walkthrough vs. heuristic evaluation vs. usability testing (2022)

Cognitive Walkthrough

Task-based approach to identify potential usability problems for first time use. It's meant to assess the learnability of the system.

Designers are trying to image the perspective of someone encountering the system for the first time

... if evaluator is familiar with design of product, they refer to personas

Based on assumption that people prefer to learn by doing rather than reading manuals or seeking help.

Evaluator is provided with exact steps to operate the interface

... it's about answering specific user-centred questions about each step

... it's not about discovering the steps to complete a task

Is especially useful for unfamiliar and niche features

... Its overkill for something that people do regularly (e.g., checkout flow)

📖 Interaction Design Foundation: How to Conduct a Cognitive Walkthrough

Tasks Become Processes

Create a “successful story” for each task and include step-by-step list of actions for the cognitive walkthrough evaluation.

Example task: "login to a website"

Convert task to **step-by-step process** describing how to use interface. These are built of off your UX flows:

1. Open browser
2. Navigate to site
3. Enter username in username field
4. Enter password in password field
5. Click the login button

📖 Interaction Design Foundation: How to Conduct a Cognitive Walkthrough

Cognitive Walkthrough Questions

"**Walkthrough**" the task process, examine the "**action**" needed to progress to the next step.

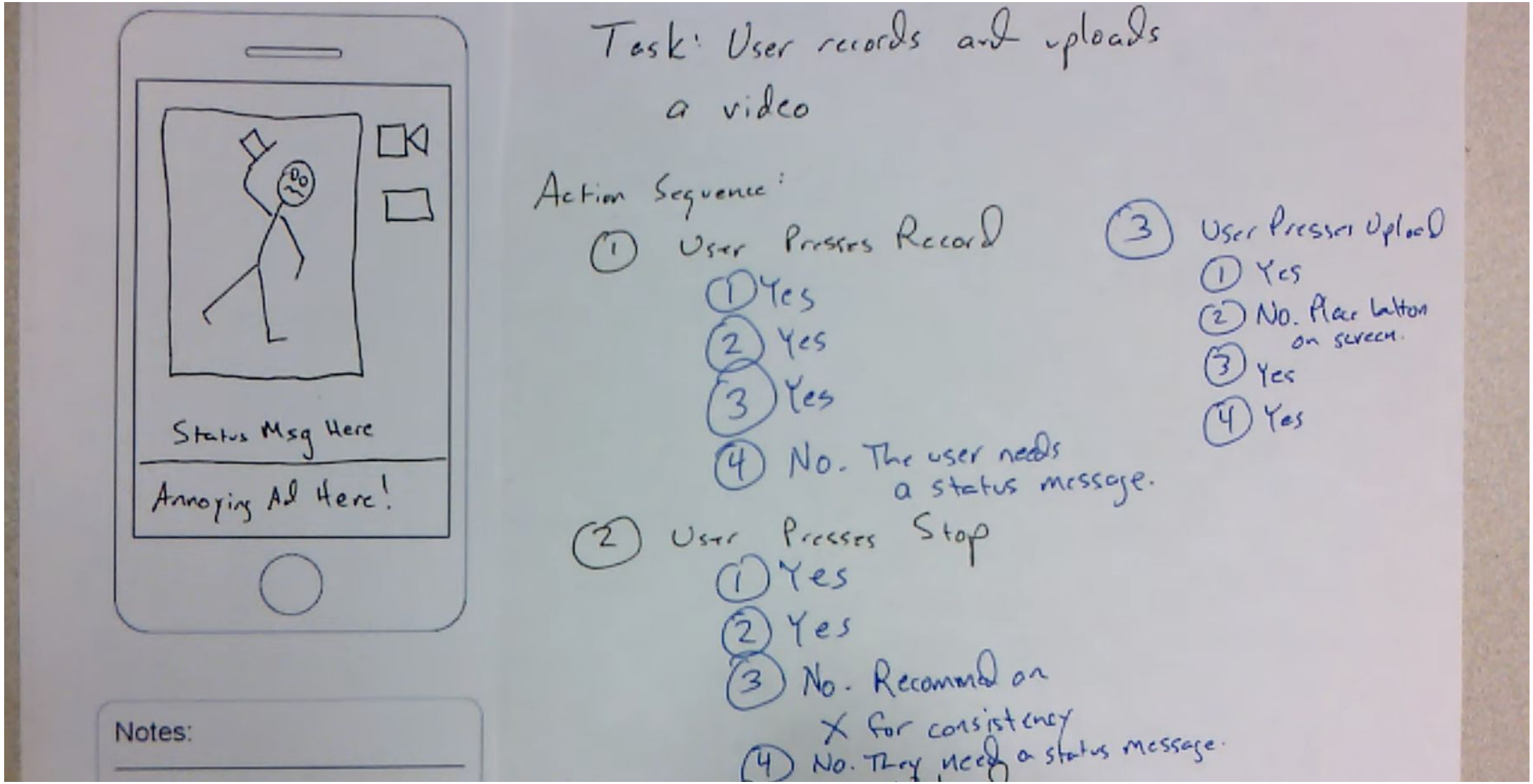
At each step, answer four simple questions:

1. Is effect of current action same as the user's goal?
2. Is action visible?
3. Will they recognize action as the correct one?
4. Will they understand feedback?

If answer is "no" to a question, provide a potential solution

Answer each question as though answers to other questions are "yes"

Cognitive Walkthrough Questions



Usability Testing

Is the UX learnable? … efficient? … memorable? … does it reduce errors? … does it meet expectations?

UX Evaluation Types

	Heuristic Evaluation	Cognitive Walkthrough	Usability Testing
Who	One or more professionals evaluators (e.g., someone with design knowledge)	One or more professional s evaluators (e.g., someone with design or product knowledge)	People representative of your target personas
What	Look for issues via heuristic categories (e.g., Neilson's 10)	Perform a list of specific tasks	Use the system in a realistic situation
Why	Identify problems and assess the systems usability	Evaluate the learnability of a systems from the perspective of first time use	Gain knowledge from those likely to use the system

📖 Sepideh Yazdi: Cognitive walkthrough vs. heuristic evaluation vs. usability testing (2022)

Usability Testing

A formal method for evaluating whether a design is learnable, efficient, memorable, can reduce errors, and meets expectations.

Evaluated with people representative of your personas.

Participants complete tasks with your design in a lab setting

Measures: task completion, satisfaction, etc.

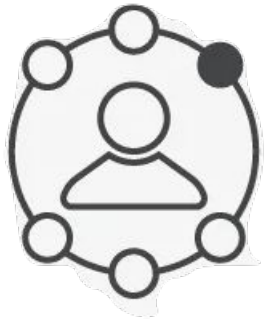
Protocols: think-aloud, co-discovery, conceptual model extraction

Report: usability problems, task measures, design suggestions

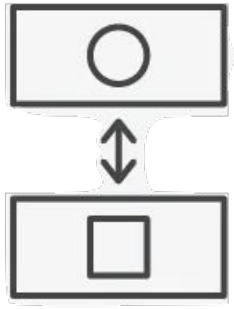
Four Approaches to Usability Testing

There are many different approaches to usability testing

The approach you choose depends on what you want to learn, your time, and your budget. Below are four examples, it's not exhaustive



Exploratory used for a quick review of your material, often used during task-based or scenario-based testing



Comparative compare two or more drafts or finished services to see which one the participants like best



Assessment observe participants' interaction with the service, can be used at any stage of development

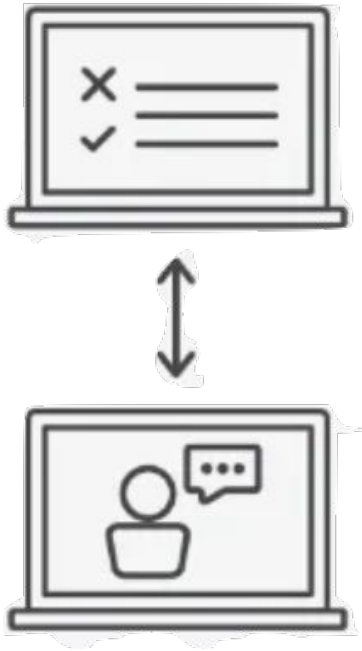


Validation test a version of the service, focusing on quantitative data, like completion rates and time taken to complete a task

📖 U.S. Department of Homeland Security: Approaches to Usability Testing

Remote vs. In Person Tests

Usability testing can be done **in person** or **remotely** ... consider what you want to learn and your available resources to choose the best option



Remote Testing
The test moderator and participatant are NOT in the same physical location.

Typically requires less resources and planning to conduct because there is no need for travel.



In-Person Testing
The test moderator and participatant are in the same physical location.

Easier for the moderator to capture physical and emotional responses to material being tested.

📖 U.S. Department of Homeland Security: Approaches to Usability Testing

In-Person Testing

Better insights into a participant's **non-verbal communication**

... Easier to recognize when to probe based on **body language**

... Able to recruit participants in lower socioeconomic areas that may not have access to technology

... It depends, some controversy

A **controlled environment** to limit distractions during the testing session

vs. Remote Testing

Participants may f**eel more comfortable** testing in their own space and using their own technology

... BUT, lots of cheating too!

... You can test participants **outside of your geographic location**

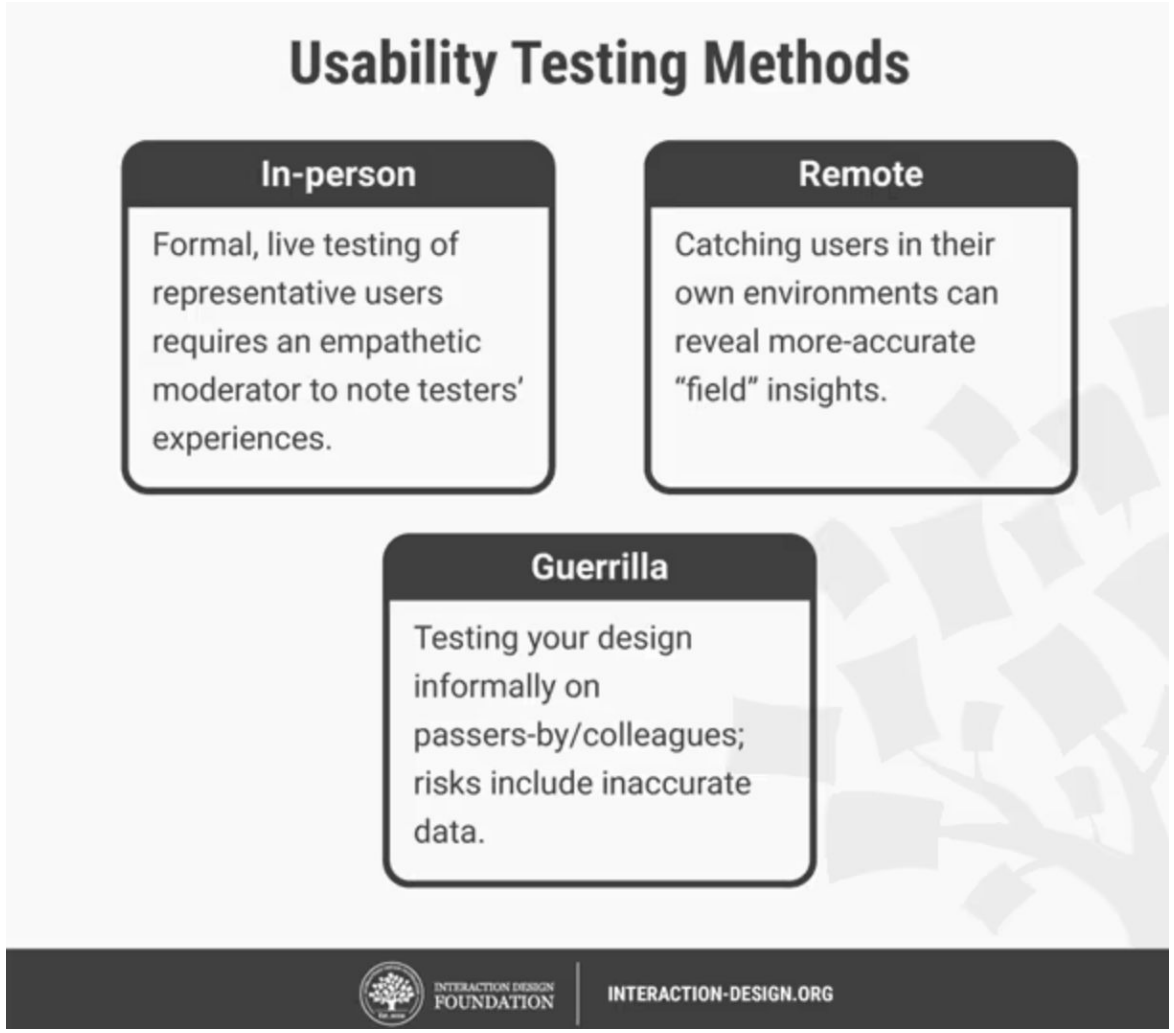
... Easier to recruit **participants** based on their **availability**

... Can be synchronous or async

More **flexible** with time and **budget friendly**

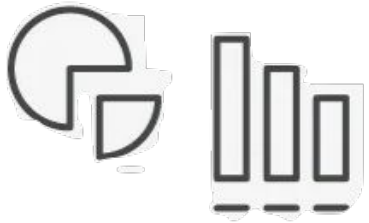
📖 U.S. Department of Homeland Security: Approaches to Usability Testing

Remote vs. In-person … vs. Option 3: Guerilla testing



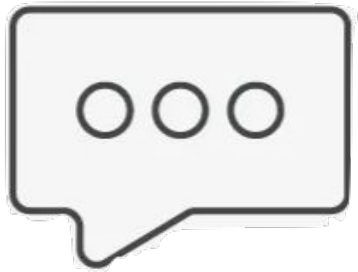
Generally, **Guerilla testing** is useful for some first impressions, but overall is not an effective method from my experience.

Quantitative vs Qualitative Data



Quantitative Data
focuses on collecting and using metrics to describe and/or quantify a user's experience.

Example: tracking how many users successfully completed an application accurately.

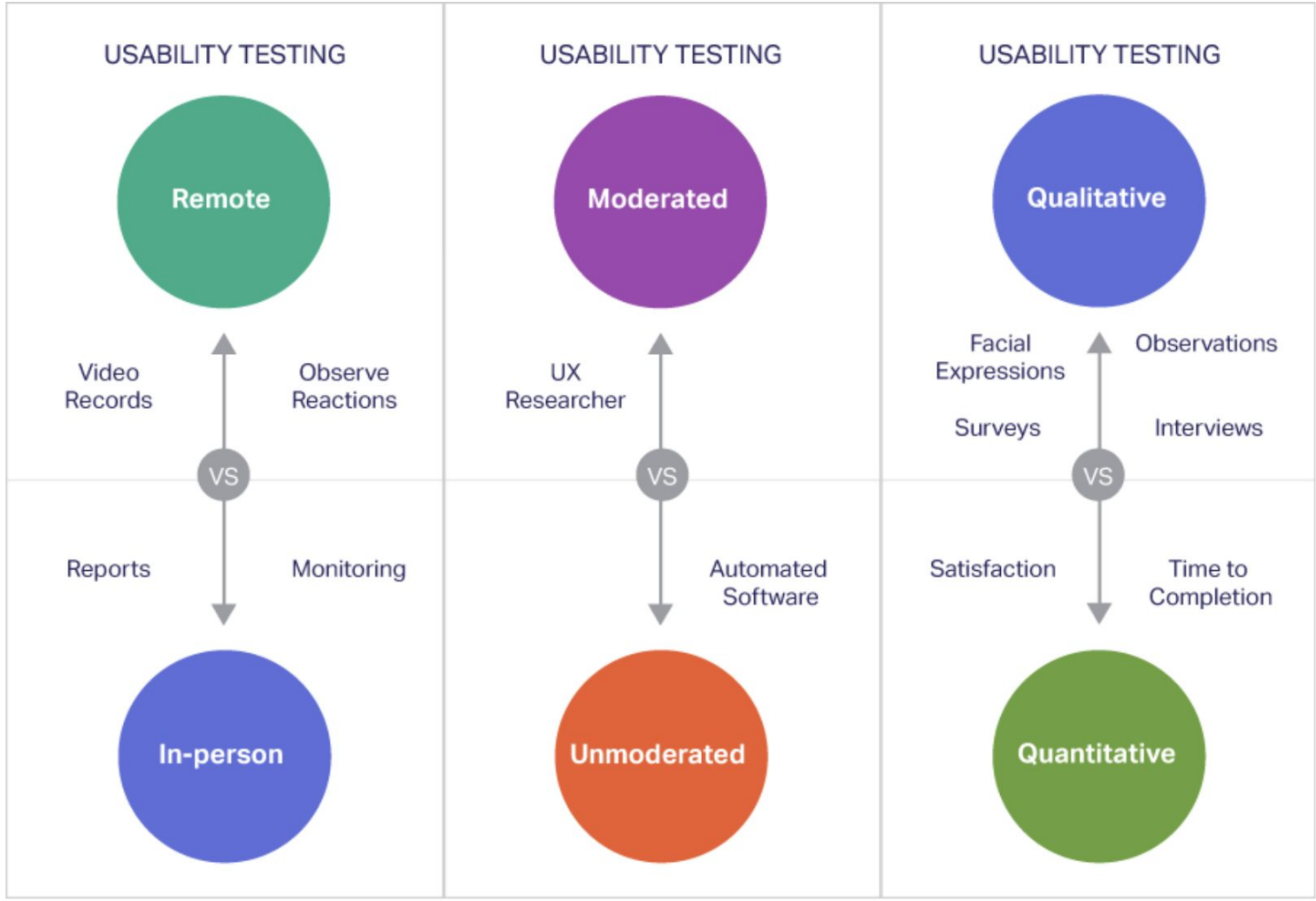


Qualitative Data
focuses on collecting and using insights, findings, and anecdotes shared by the participant to describe their experience.

Example: capturing a participant's pain-points by observing their frustrations while they test a service.

📖 U.S. Department of Homeland Security: Approaches to Usability Testing

Approaches to Usability Testing - Additional Dimensions



📖 U.S. Department of Homeland Security: Approaches to Usability Testing

How Many People Do You Need for a Usability Test?

5 participants will identify 84% of usability issues

- (according to Jakob Nielsen's seminal 1989 paper)

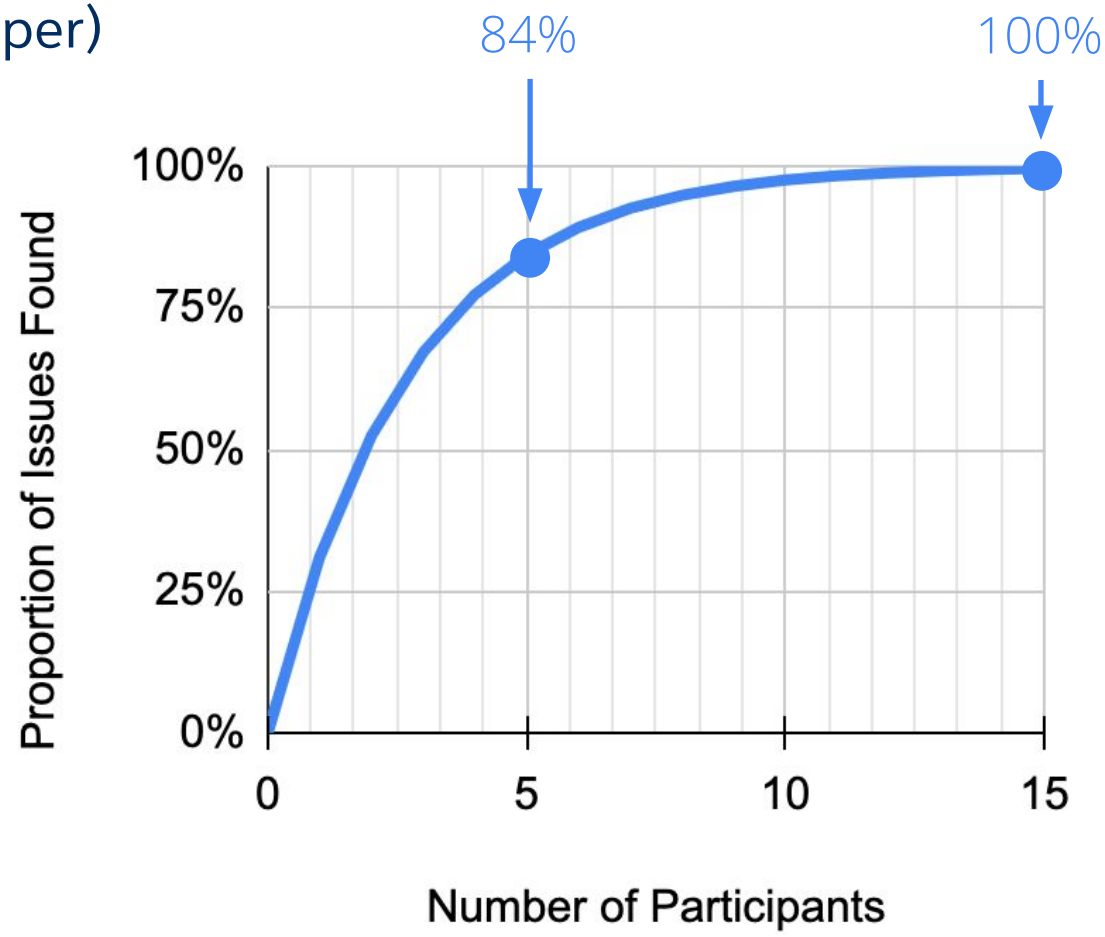
Modeled as:

$$1 - (1 - L)^n$$

where:

n is number of participants

L is proportion of usability issues discovered while testing a single person (typical value 0.31)



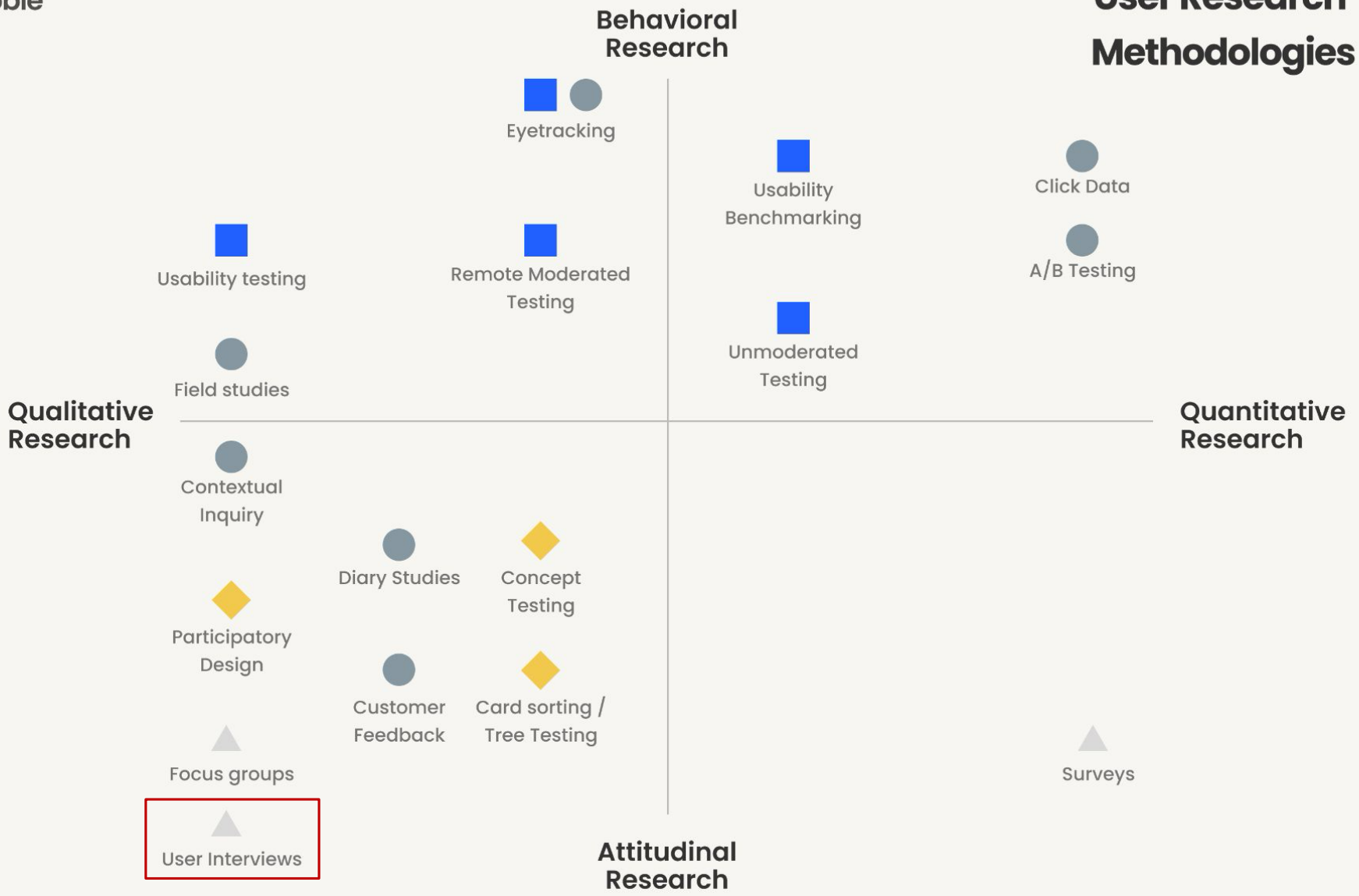
📖 Jakob Nielsen (Nielsen Norman Group): Why You Only Need to Test with 5 Users (2000)

More UX Evaluation Methods

A/B testing, Eye tracking, Card Sorting

hubble

User Research Methodologies



📖 Brian Byun: User Testing vs Usability Testing (2026)

Limited use of product · Using the product · Scripted use of product · Not using product

hubble

User Research Methodologies



📖 Brian Byun: User Testing vs Usability Testing (2026)

Limited use of product · Using the product · Scripted use of product · Not using product

Example : A/B Testing

Quantitative comparison of two or more versions of a interface

... also called "split testing", randomized trials

... Hypothesis is a specific interface design change will affect behaviour

Randomly split participants between *two versions of an interface*

... almost always done with "live" interface, high number of participants

Success is determined using 1 or more *quantitative metrics*

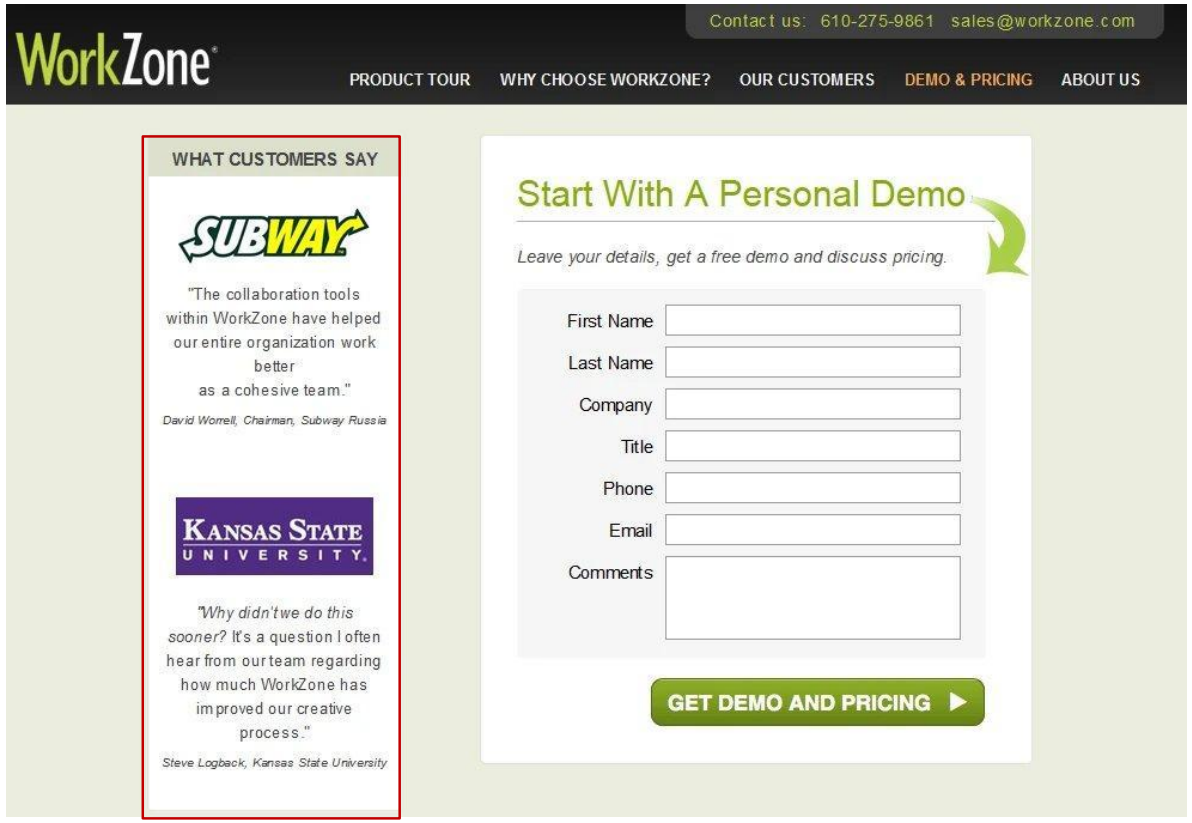
... e.g. clicks, conversion rate, sales, support requests, etc.

Highlighly dependent on *instrumenting* your application to collect data

... Modifying software code so that it can gather data for analysis (e.g., time, clicks), ... high coverage is difficult to achieve

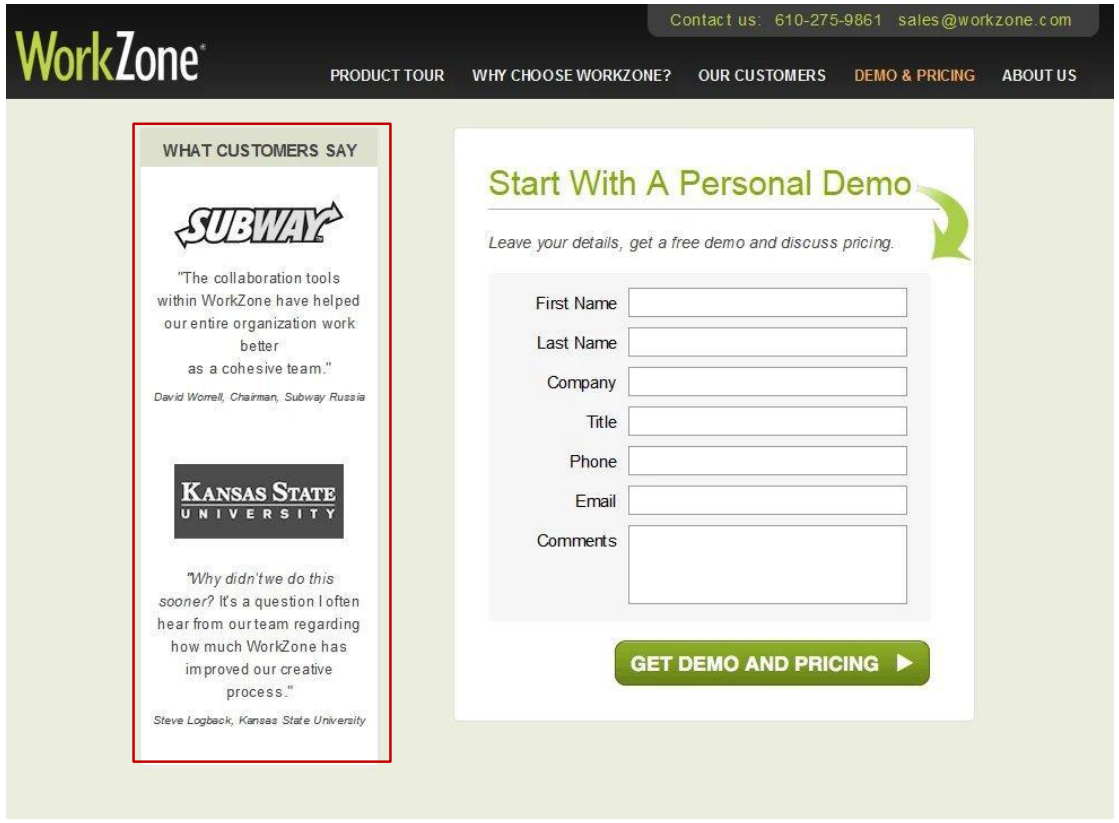
A

Prominent logo



B

Less prominent (b/w) logos *



📖 Astha Khandelwal (VWO): 7 A/B Testing Examples and Case Studies (2026)

50

A



B

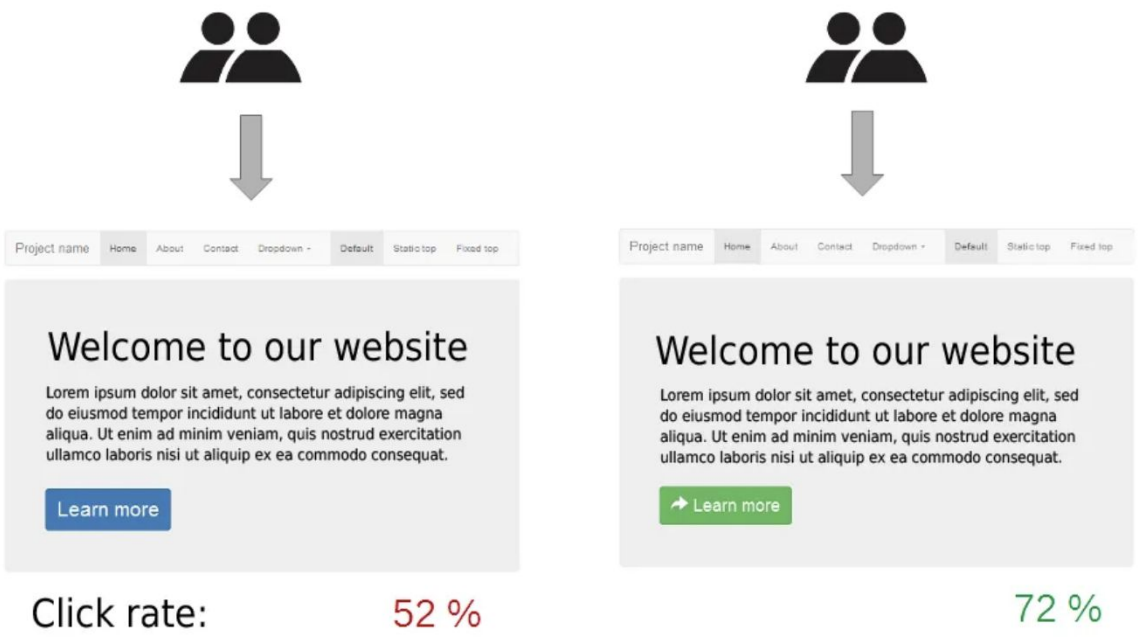


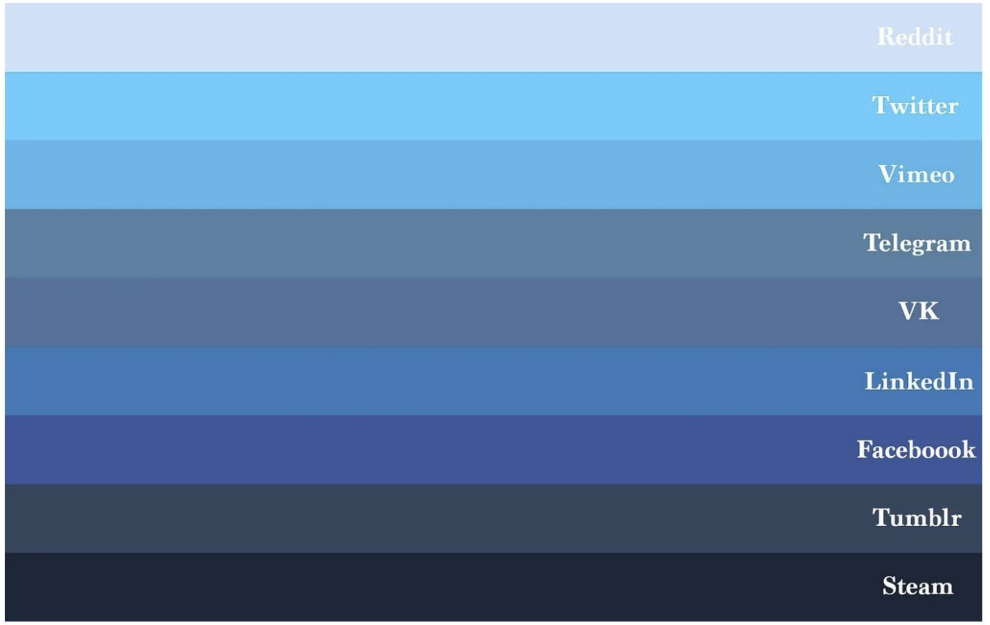
📖 Astha Khandelwal (VWO): 7 A/B Testing Examples and Case Studies (2026)

Example : A/B Testing

Data Driven Decisions: Google's '50 shades of Blue' experiment

Would you believe it if I told you that switching to a different shade of blue netted Google as additional $200 Mn in Ad revenue?





📖 Bharath Kumar Balasubramanian: Data Driven Decisions: Google's 50 Shades of Blue Experiment (2020)

Formal Experiments

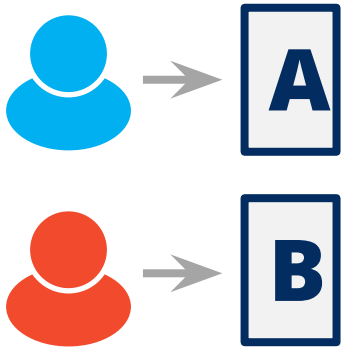
Hypothesis
e.g., *people click red buttons more than buttons with other colours*
... formally, there's a "null hypothesis" and "alternative hypothesis"

Independent variable (IV)
Factors that possibly “cause” changes in dependent variable
... e.g., button colour

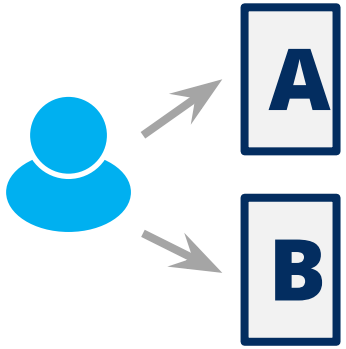
Dependent variable (DV)
A measurable outcome of a task
... e.g. number of clicks

Experimental Design

Between-subjects: *different people test each condition*
e.g. each group of participants tests different UX variant
- ... need to recruit many participants
- ... individual differences can confound results
- ... no learning, fatigue, or transfer effects



Within-subjects: *the same person tests all the conditions*
e.g. each participant tests all UX variants
- ... recruit fewer participants
- ... no individual differences across conditions
- ... learning, fatigue, and transfer effects are a threat
- ... need to counterbalance or randomize condition order



Statistical Analysis of Experiment Data

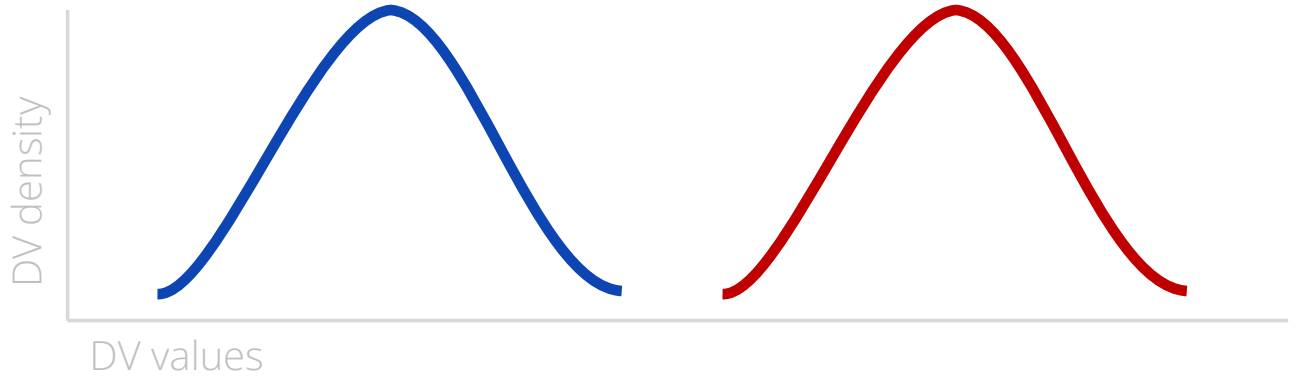
Descriptive statistics

- Measures of central tendency: mean, mode, median
- Measures of spread: variance, stdev, range, IQR, confidence interval

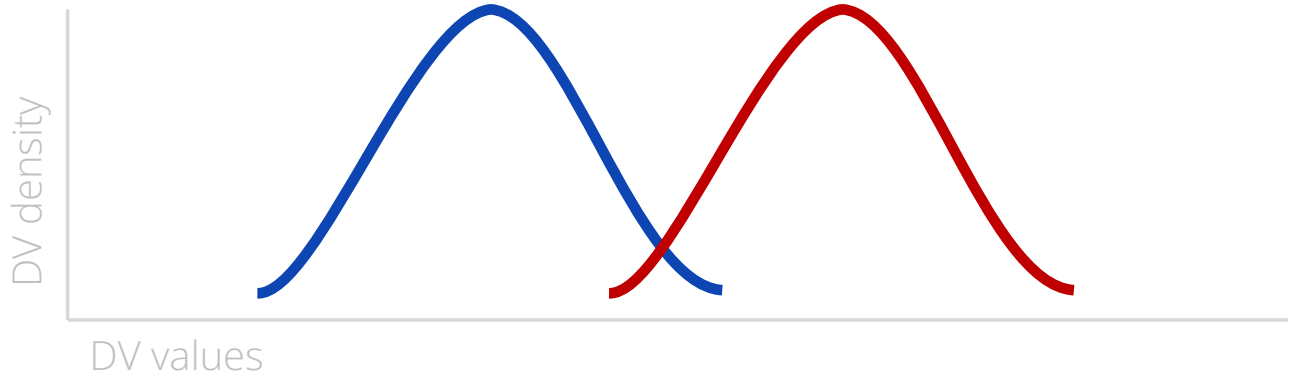
Statistical tests

- Evaluate variances explained by the IVs and those that can't
- Find probability of observed difference occurring by chance (p-value)
- Methods: t-test, ANOVA, ...

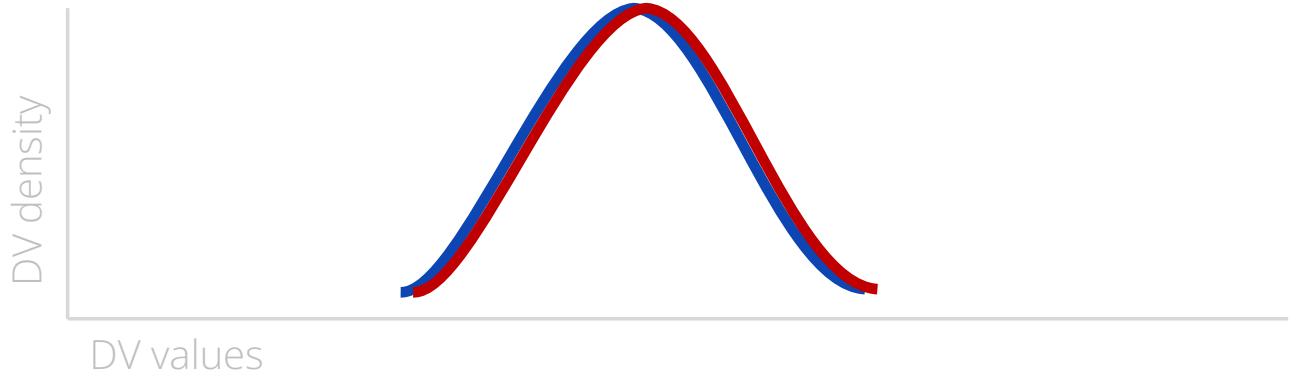
different



?



same



IV: **A** **B**

Experimental Validity and Bias

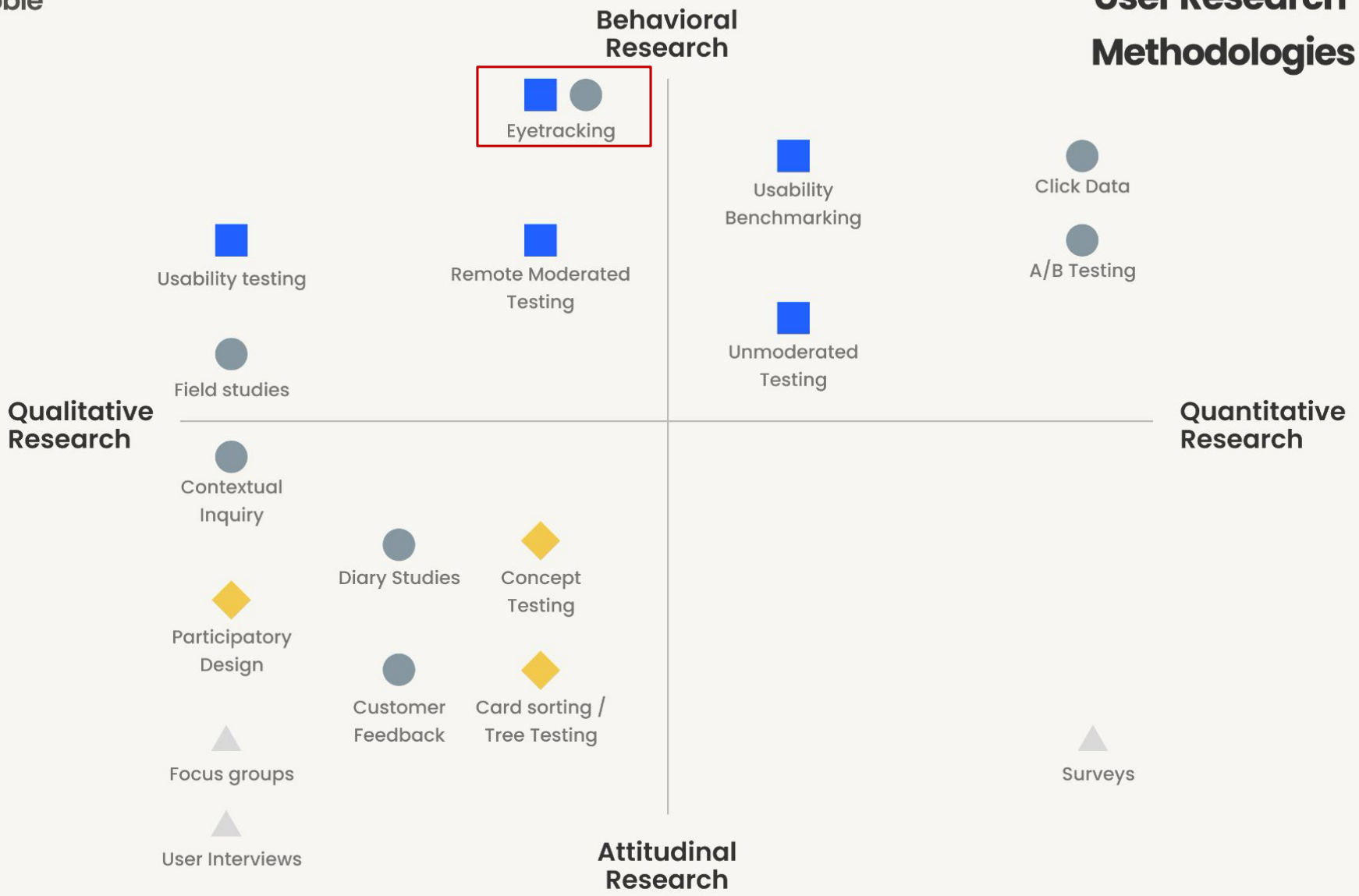
External validity, *i.e. realism*
... Generalization across situations
... Generalization across people

Internal validity, *i.e. integrity*
... How well the people, setting, and task are controlled
... Results are likely to replicate

Potential Experiment Biases
... Unreliable or different measurement instruments/methods
... Failure to account for learning effects, fatigue effects, or transfer effects
... Selection of participants
... Experiment facilitator behaviour

hubble

User Research Methodologies



📖 Brian Byun: User Testing vs Usability Testing (2026)

Limited use of product · Using the product · Scripted use of product · Not using product

Eye Tracking

Detects person's fovea fixations and movements between fixations

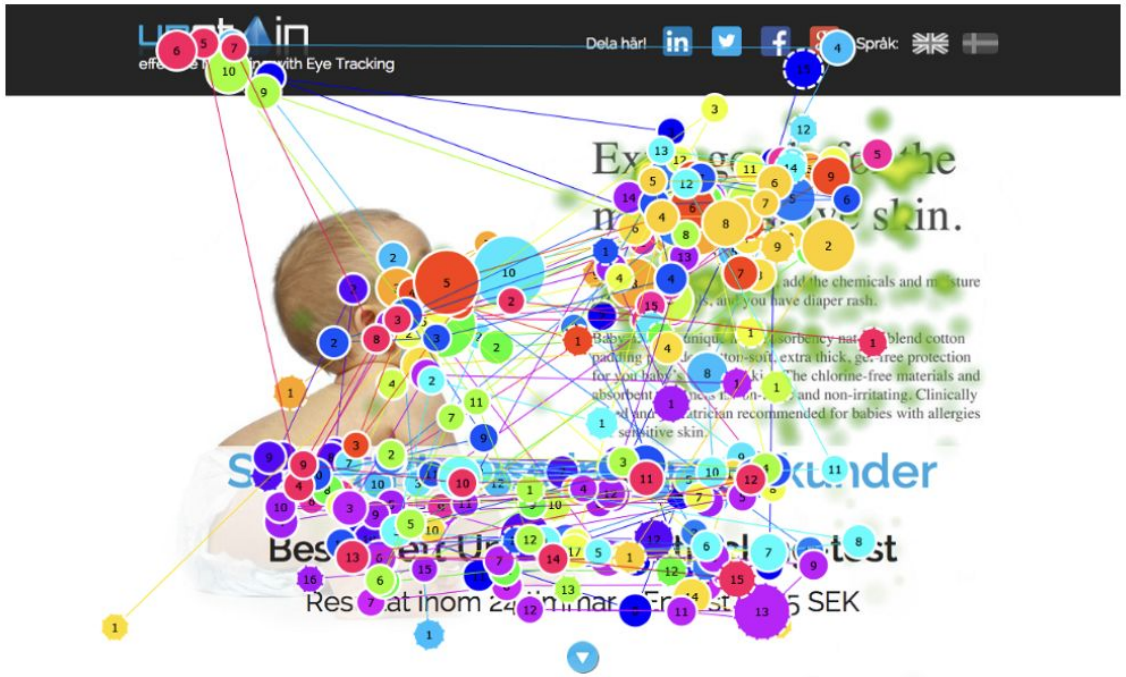
Show hard-to-articulate behaviour:

... Where participants expected to find certain elements
... Whether participants noticed a particular element
... Whether there were differences between personas groups
... What elements of the interface are distracting
... Efficiency of a design or guidance through a task
... Which content participants read in details, scan or ignore

Reported as a gaze plot for one participant and as heat maps for study overall.

Eye Tracking Results

Gaze Plot

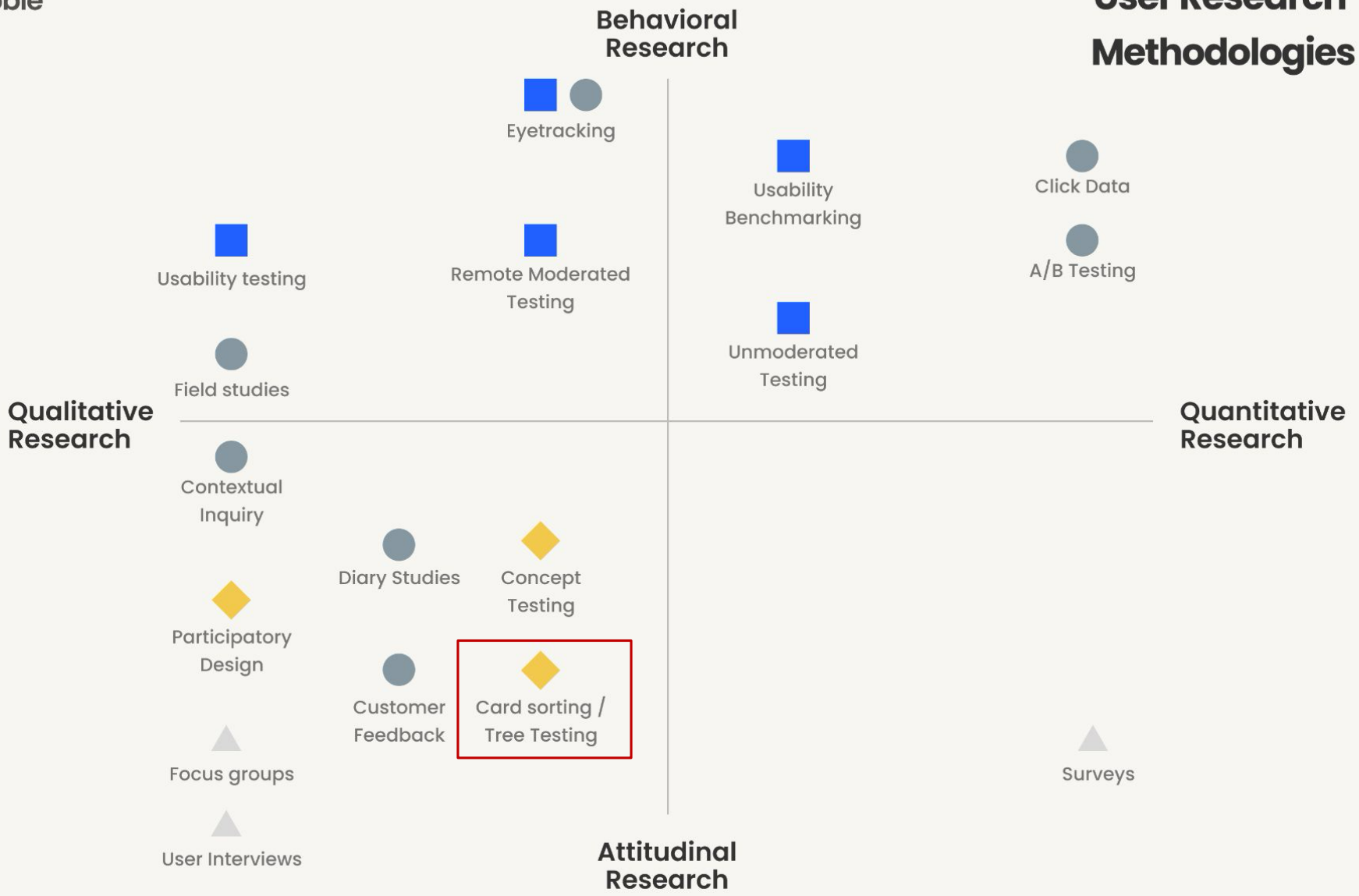


Heat Map



hubble

User Research Methodologies



📖 Brian Byun: User Testing vs Usability Testing (2026)

Limited use of product · Using the product · Scripted use of product · Not using product

Card Sorting

Card sorting is a UX research method used to understand how people categorize and organize information

... It primarily helps in designing or evaluating the information architecture (IA) of a website, application, or content

... It reveals participants' mental models, showing how they group and label content, which is crucial for intuitive navigation

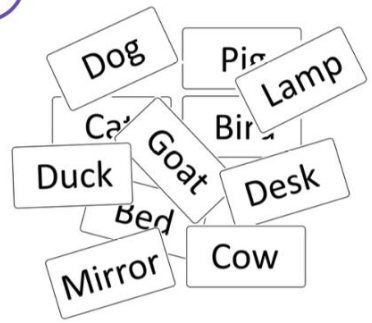
It can help with different elements of application design:

... Menus design, content strategy, product categorization, etc.

Example Card Sorting

Closed Card Sort:

1



Participants get a stack of cards

2

Animals	Furniture
Goat	Table
Dog	Mirror
Cow	Desk
Pig	Chair

Participants sort cards into groups the researchers created

Open Card Sort:

1



Participants get a stack of cards

2

Goat	Table
Dog	Mirror
Chair	Desk
Cow	

Participants sort cards into groups

3

Has 4 legs	Flat
Goat	Table
Dog	Mirror
Chair	Desk
Cow	

Participants label groups

📖 Digital.gov: Usability (2025)

Example Card Sorting

Open Sorting Pros

... more accurate reflection of a participants' mental model
... more flexible
... good for **generative** phase of a project

Open Sorting Cons

... less control
... wide array of answers that may not result in a discernible pattern
... bad for evaluative phase of a project

Closed Sorting Pros

... more likelike to see grouping patterns in results
... more controlled
... good for **evaluative** phase of a project

Closed Sorting Cons

... less flexibility
... maybe less accurate reflection of participants' mental model
... bad for generative phase of a project

📖 Cary-anne Olsen-Landis (IBM): Card sorting: a powerful, simple research method (2017)

Evaluating High-Fidelity Prototypes

Heuristic Evaluation, Cognitive Walkthrough, Usability Testing, Other Methods

Using Results of Usability Studies

1. Identify what needs to be tested and why
2. Find the perfect recruit.
3. Create a test plan and make clear objectives.
4. Where appropriate implement learning effects during the evaluation
5. Decide upon the measures metrics which will help you assess the results.
6. Define roles. Choose appropriate moderator and way to observe participants.
7. Run a pilot test before the real test.
8. Analyze and prioritize on the findings and discuss insights with stakeholders.
9. Introduce the required changes to your product design.

📖 S. Ivanov: What Is Usability Testing and How To Conduct One? (2021)

Design Critique

i.e., how to ask for it, how to give it

Asking for Critique

Project goals: clarify your value proposition, objectives, metrics

Audience: who **is** and who **is not** audience for your project?

Project stage: e.g., early ideation stage vs. high-fidelity prototype

What you're looking for feedback on: e.g., usability, visual identity, information hierarchy, … and what you don't want feedback on!

Week 11 Studio

CR Design Critique

P11: Prep for CR – Design Critique

Continue to develop your High-Fidelity Prototype in Figma (**D8 High-Fidelity Prototyping**). For the **Design Critique** session, your high-fidelity prototype needs to be partially interactive.

If applicable, download the **Figma Mobile App** onto your phone and open your Figma prototype on the phone before the design critique session.

Link to your Figma presentation in your design notebook.

See full **P11** instructions on course website

CR – Design Critique

The presenting team will give a 10 minute (sharp) presentation about the visual designs for their features and UX flow using Figma displayed on a laptop/tablet. Link to your Figma presentation in your design notebook.

Then, the critiquing team will have 15 minutes to provide feedback and suggestions for improvement. Following this, the teams will switch roles and the process will repeat.

In the design notebook entry, list at least 5 feedback (that you received from a design critique session) that your team decided to act on. Describe what your team has changed in your design based on those feedback, and include screenshots of your design changes.

See full **P11** instructions on course website

Looking Ahead to Week 12

HCI / UX research in industry Panel • E2: High-Fidelity Prototyping Evaluation