

FIT3175 - Usability

UX Evaluation Methods

Week 4 Lecture P2

Copyright Warning

Commonwealth of Australia Copyright Act 1968

Warning

This material has been reproduced and communicated to you by or on behalf of Monash University in accordance with section 113P of the Copyright Act 1968 (the Act).

The material in this communication may be subject to copyright under the Act. Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.

Learning objectives

Evaluation in UCD processes

- Aims and considerations

Usability Testing

- Why test with users?
- Test planning
- Testing methods

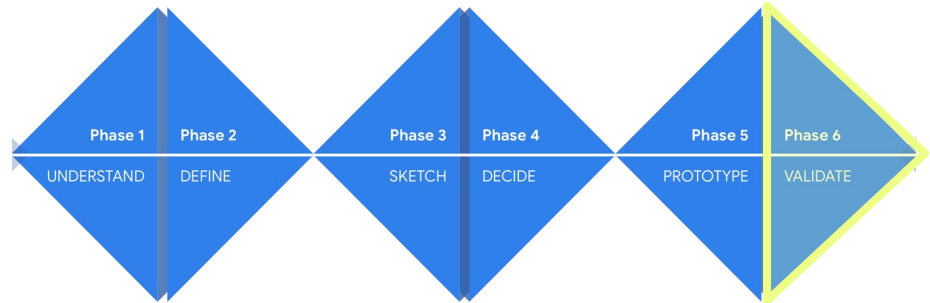
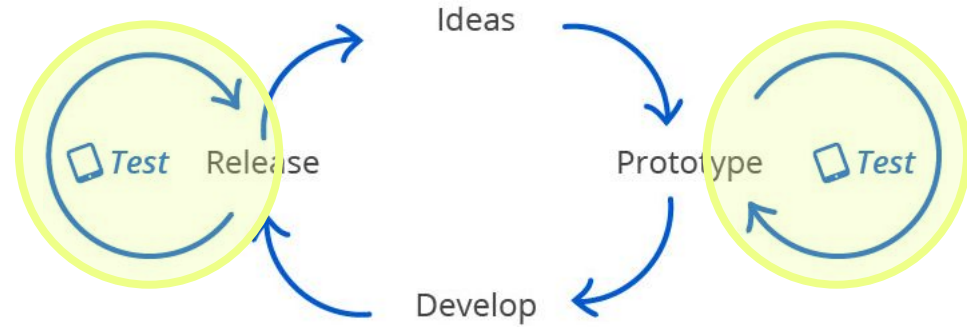
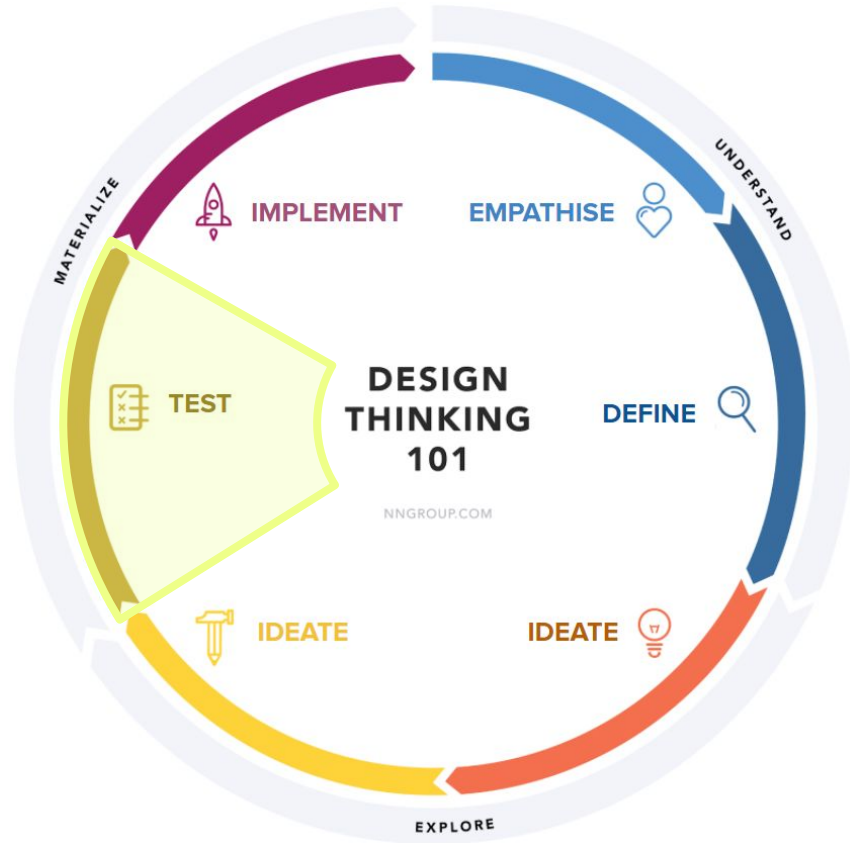
Expert Reviews

- Why test with experts?
- Nielsen Heuristics
- Heuristic Evaluation method

Discussion Forum 2, Assignment Project Stage E + F

Evaluation in UCD Processes

Recap: UCD Processes



Usability evaluation aims

Evaluation involves using different methods to test and assess a product, and solicit feedback.

- Evaluation can be conducted for new or existing products.
- Evaluation can be conducted at different stages of design.

The purpose is to check and ensure if a product meets the user needs, if there are any issues, and whether users like the product.

- Evaluation done early allows ideas to be validated
- Evaluation done later allows implementations to be validated

Evaluation aims

How would you know the following during a design process?

- Are the options and controls visible to the user?
- Does the user know how to interact with the UI elements?
- Is the navigation and workflow appropriate?
- Does the product do what the user expected it to do?
- Is the functionality correct from the user's perspective?
- What are the good and poor features?

Extensive testing is a key to successful design

- Improves our understanding of real users
- Identify the issues and addressing them.
- Improves user experience and increases user satisfaction

Recap: Data gathering techniques

Evaluation not only involves testing, but collecting data that can be analysed.

Consider the following data gathering techniques from Lecture 2:

Surveys

Structured Interviews

Semi-Structured Interviews

Unstructured Interviews

Experience Sampling

Diary Studies

Ethnography

Contextual Inquiry


Focus Groups

Some data gathering techniques are better suited to evaluations performed at earlier or later stages of a project's design and development cycle.

Behavioural research

Discovery and Architecture	Early Design Phases	Complete, working Design
<ul style="list-style-type: none">● Benchmarks<ul style="list-style-type: none">○ Measuring metrics● Field studies● Card sorting<ul style="list-style-type: none">○ Users organising content topics● Tree testing<ul style="list-style-type: none">○ Users testing findability of organised topics	<ul style="list-style-type: none">● Prototype testing<ul style="list-style-type: none">○ Interaction○ Visual design○ Content● Intercept survey<ul style="list-style-type: none">○ Engaging with users in real-time	<ul style="list-style-type: none">● Benchmarks● Field studies● Intercept survey● Eyetracking<ul style="list-style-type: none">○ Eye movements● Analytics<ul style="list-style-type: none">○ Real-time behavioural data● Support feedback<ul style="list-style-type: none">○ Collected from customer support

Attitudinal research

Early Design Phases	Later Design Phases	Complete, working Design
		
<ul style="list-style-type: none">• Surveys• Focus groups• Diary studies• Interviews<ul style="list-style-type: none">○ In-person (lab)○ In-person (field)○ Remote	<ul style="list-style-type: none">• Surveys• Focus groups• Diary studies• Interviews• Participatory design<ul style="list-style-type: none">○ Involve users and stakeholders in design processes	<ul style="list-style-type: none">• Surveys• Focus groups• Diary studies• Interviews• Participatory design<ul style="list-style-type: none">○ Involve users and stakeholders in design processes

Controlled and contextual environments

Controlled Environment	Medium Contextual	Highly Contextual
<ul style="list-style-type: none">• Lab testing• Focus groups• Participatory design• Interviews	<ul style="list-style-type: none">• Remote usability tests• Intercept survey<ul style="list-style-type: none">○ Live site○ Physical location• Interviews	<ul style="list-style-type: none">• Field studies• Diary studies• Interviews

Usability Testing

Why test with users?

Usability testing was introduced in early 1980s, realising how important user needs are in the interface design.

- Cheaper to fix problems during the design cycle rather than after development.
- Evaluate whether a product meets its intended purpose and user needs relative to usability measures.

Reminder: ISO recommends defining at least 1 metric for each usability measure

- **Effectiveness:** the accuracy and completeness with which users achieve goals
- **Efficiency:** based on the resources expended to complete a task
- **Satisfaction:** the level of the comfort and attitudes towards the product

Planning a usability test

A usability test has many planning considerations to deliver reliable outcomes.

1. **Purpose of the test:** What are the user goals? What do you plan to test/measure?
2. **Participants:** Selection criteria? How will you recruit them? How many users?
3. **Test environment:** Location and equipment.
4. **Specific tasks and goals:** What do users need to do? Task selection and order.
5. **Testing methods:** What is the procedure? How will it be facilitated?
6. **Survey instrument:** How will you collect and analyse data?
7. **Administration:** Explanatory statement, consent forms, ethics approval.

How many participants?

Many users
per round



Few users
per round

Quantitative usability tests (**>20 users**)

Tree-testing (**50 users**)

Eyetracking heatmaps (**40 users**)

Card sorting (**~20 users**)

Interviews (**~10 users**)

Focus groups (**~8 users**)

High-fidelity prototypes (**~5-8 users**)

Very early prototype usability tests (**~2 users**)

Real users as participants

When recruiting participants, define selection criteria to find participants that are representative of real users.

Real users when the target group is known (e.g. employees within a company)

- Users currently using the product.
- Users who will later use it.

Representative users if real users aren't available (e.g. publicly accessible website)

- Relevant demographic groups of interest.
- Proxies with relevant knowledge of user behaviour and attitudes.

Participation in usability tests must be voluntary.

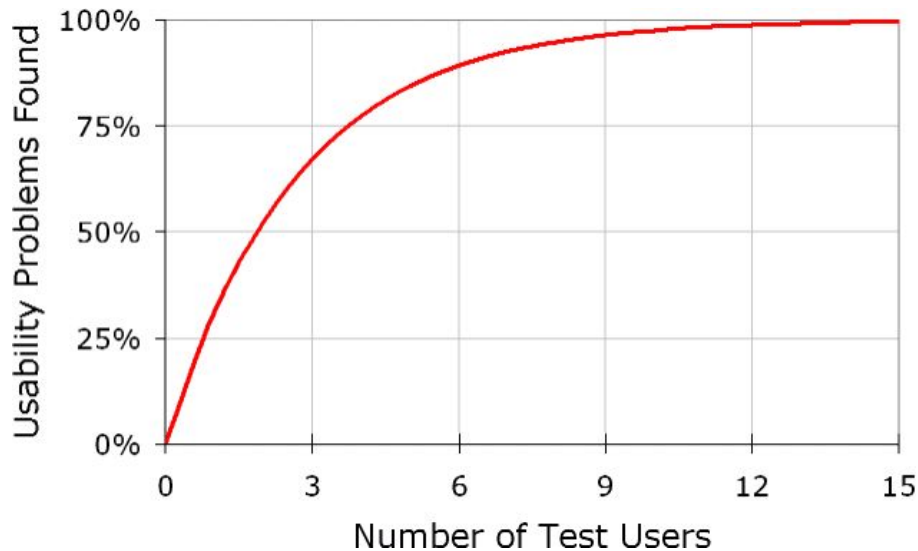
5% of users find 85% of problems

In **qualitative** studies we typically find:

- **0 users** = 0% of problems found
- **1 user** = 30% of problems found
- **5 users** = 85% of problems found
- **15 users** = 100% of problems found

So why not just test with 15 users?

- Tests are costly to facilitate.
- Some testing methods require significant time commitments.
- It is more effective to distribute your budget across many small tests.



Above: The addition of new users in testing increases the number of issues identified. However, not the diminishing returns as the number of users increases beyond 5.

Real tasks for testing

For complex systems, consider focusing on testing specific scenarios.

- Define user goals.
- Start with a simple task.
- Give users one task at a time.
- Follow your design's flow.
- Make tasks actionable.
- Set a scenario.
- Avoid giving precise instructions.
- Include up to eight tasks in a test.

Unnecessarily complex or lengthy testing process that burden a user can result in:

- **Inaccurate testing results**
- **Inaccurate qualitative data.**
- **Participant drop-off**

This is especially true for usability tests that are conducted remotely.


Users should be focused on the goals and tasks - not the testing process.

Explanatory statement and consent

An explanatory statement and consent form helps users to understand the research they are about to participate in.

It explains:

- The research
- The data collection
- Information storage
- Participant rights.

 **MONASH** University

EXPLANATORY STATEMENT

Project name: #####

Monash University Human Ethics ID: #####

Research team and contact details:

##### Chief Investigator	##### Co-Investigator	##### Project Manager
Phone:	Phone:	Phone:
E:	E:	E:

You are invited to take part in this study

What does the research involve?
...

Why were you chosen for this research?
...

Participant incentives
...

Consenting to participate in the project and withdrawing from the research
...

Confidentiality
...

Possible benefits and risks to participants
...

Source of funding
...

Storage of data

CONSENT FORM

Project name: #####

Monash University Human Ethics ID: #####

Chief Investigator: #####

I have been asked to take part in the Monash University research project specified above. I have read and understood the Explanatory Statement and I hereby consent to participate in this project.

I consent to the following:	Yes	No
Audio recording during the interview		
Taking part in a focus group of up to 10 people		

Name of Participant _____

Participant Signature _____ Date _____

Task description example

Task #	Task Description
1	Login Click on Login button to continue.
2	Data Entry Choose Daily Data Entry option from the menu.
3	Enter your daily record 1. Blood Pressure Choose Manual Data Entry option to enter your blood pressure values. Enter 128 for Systolic field Enter 84 for Diastolic field Enter 70 for Pulse field 2. Press NEXT button and select CAPD option under Type of Dialysis field. 3. Press SUBMIT button.
4	Check your daily report 4.1 Select Progress Reports option from the menu list 4.2 Select Daily Report option from the list. 4.3 Click on Back button.
5	Check your weekly report 5.1 Select Weekly Report option from the list. 5.3 Tap on calendar icon and select start date as 5-Sep-2016 5.4 Tap on calendar icon and select end date as 15-Sep-2016 5.5 Tap on Get Progress Report button
Tap on Back button of the phone to complete this part.	

Task analysis allows us to assess completion of goals.

- **Cognitive**
 - Mental processing such as problem-solving.
- **Hierarchical**
 - Sub-task breakdowns to analyse interactions.

Right: A task analysis table breaks down a goal into its component actions. Each action can be individually assessed.

Lab testing

A usability lab includes a testing room and an observation room.

- Controlled environmental settings where users are observed.
- Recording or monitoring users' performance doing typical tasks.
- Data is stored and analysed.
- Observing and measuring performance to identify and explain errors.



Field studies

Bring the usability testing and research to the participants in their environment.

- Observation in a more natural setting.
- Increases understanding of what users do and the impact of technology.
- It can involve interviews.

Most importantly, being there to observe, ask questions and record what users are doing and saying.



Think-aloud

An flexible and cost-effective method is to ask your to "think-aloud".

- Users verbalise thoughts while they perform tasks.
- Testers will listen for the clues.
- Begin with a simple practice task.
- Allow time for questions and comments at the end.

However, verbalising thoughts can create additional cognitive load for users!



Eye tracking

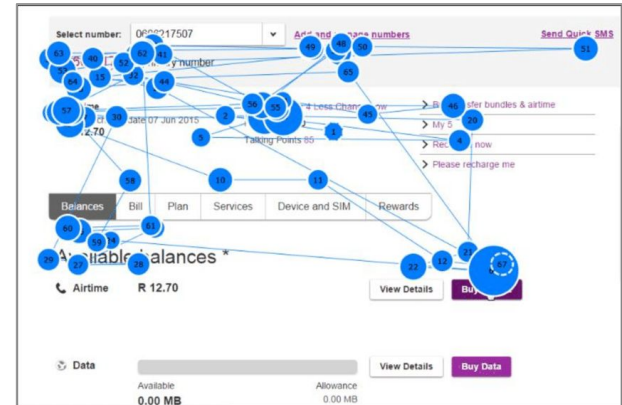
Measurement of user eye movement can be achieved in lab settings using cameras or specialised equipment.

- **Fixation** can be measured as the duration of a user's gaze at a specific location on a screen.
- **Seccade** tracks the movement path of the eyes as they navigate between fixation points.

Eye tracking tests can accurately validate user attention and interface scanning patterns.

Right-top: A **heatmap** visualises **fixation** locations and duration.

Right-bottom: A **gaze point plot** shows visual **seccade** patterns.

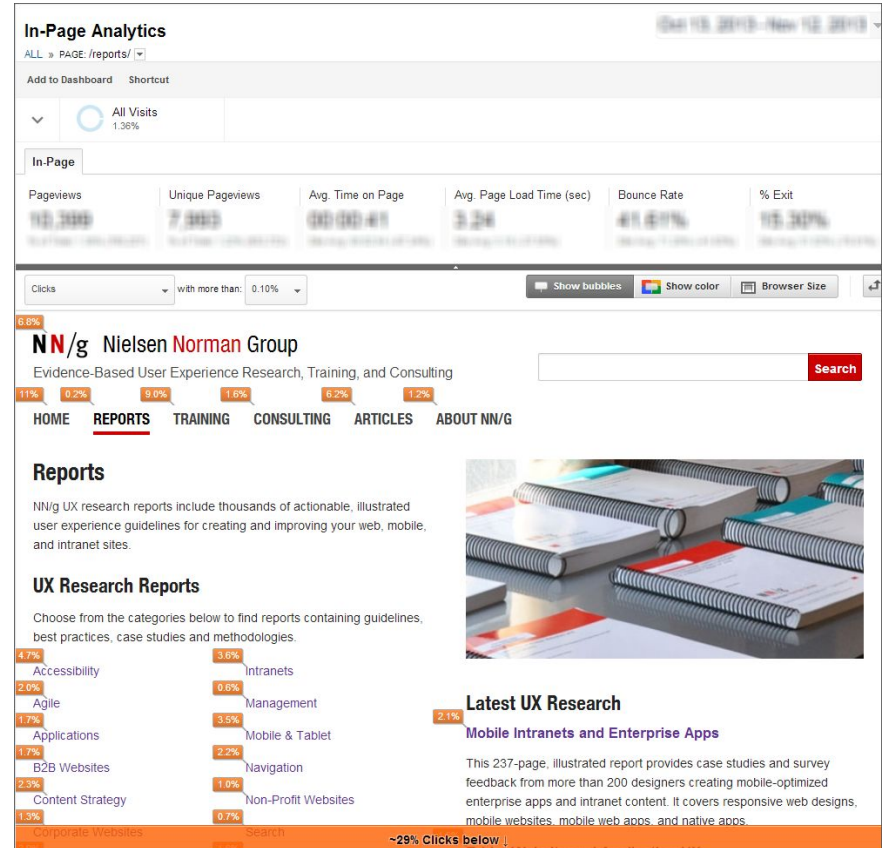


Analytics

Website analytics are commonly used as a marketing tool, but tracking of user behaviour can provide usability insights.

- Starting page and bounce rates
- Referral traffic (where users arrive from when entering a site)
- Desirable actions and paths
- Preferred platforms of users

Analytics can aggregate data from large groups of user with little added overhead.



A/B testing

Also known as **split testing**, 2 slightly different versions of a design are randomly shown to users.

- Capture analytics
- Look for success signals
- Measure success metrics

The design that produces the highest conversion rate is the winner.

The image displays two mobile app registration screens side-by-side, labeled A and B. Both screens feature a hamburger menu icon in the top left and a circular label with the variant letter in the top right. Screen A, titled 'Register form', has a clean layout with three input fields for 'First name' (containing 'John'), 'Last name' (containing 'Smith'), and 'Email' (containing 'johnsmith@mail.com'). Below these fields is a large blue 'Continue' button, and at the bottom is a link 'Sign in to my account'. Screen B, titled 'Hello! What's your name?', uses a more conversational tone. It has identical input fields for the same user data. However, the primary button is 'Create your account', and it includes an additional link 'I already have an account. Sign in to my account' just above the bottom link. The bottom right corner of the entire graphic features a decorative pattern of small dots.

A

Register form

First name
John

Last name
Smith

Email
johnsmith@mail.com

Continue

[Sign in to my account](#)

B

Hello! What's your name?

First name
John

Last name
Smith

Email
johnsmith@mail.com

Create your account

[I already have an account. Sign in to my account](#)

[Sign in to my account](#)

5-Act Interview

The evaluation process in a design sprint can be supported using an efficient 5-act interview structure:

- | | |
|----------------------------------|--|
| 1. Friendly welcome | Establish comfort and rapport. |
| 2. Context questions | Understand the user and relevant background. |
| 3. Prototype introduction | Explain the prototype and testing. |
| 4. Tasks and nudges | Ask your to perform tasks, with gentle guidance. |
| 5. Quick debrief | Open-ended interview questions |

A structured process can be replicated relatively easily to deliver more consistent results between different users and different testing rounds.

Sample post-testing questionnaire

Statements	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I was able to complete the set of tasks easily.					
The text was displayed in a way that was easy to read.					
I felt like the application was easy to learn.					
I found it easy to enter the text/numeric values manually.					
Information required to complete the task was easy to locate in the menu.					
The buttons to continue to the next screen were easy to find.					
The colour used in interface made text understandable and easy to read.					
The colour scheme used in interface was attractive.					
The number of steps required to get to the information I wanted was acceptable					

Severity scale for usability test findings

Categorising UX problems helps us understand and prioritise solutions.

- **High** Stopped users, frequent frustration or self-blame.
- **Medium** Users could complete the task, but with difficulty.
- **Low** Users only had minor issues, or cosmetic.
- **Bug** Experience not working as implemented or intended.
- **Good** Users easily completed the task or liked the experience.

Expert Reviews

Why test with experts?

It is important to understand that there are different types of experts that can be involved in an evaluation process.

- **Project stakeholders**

... whose interests are affected by project's outcome (managers, team members, clients, investors, end users, etc.)

- **Designers and developers**

... who understand the intent of the design and implementation feasibility.

- **Usability experts**

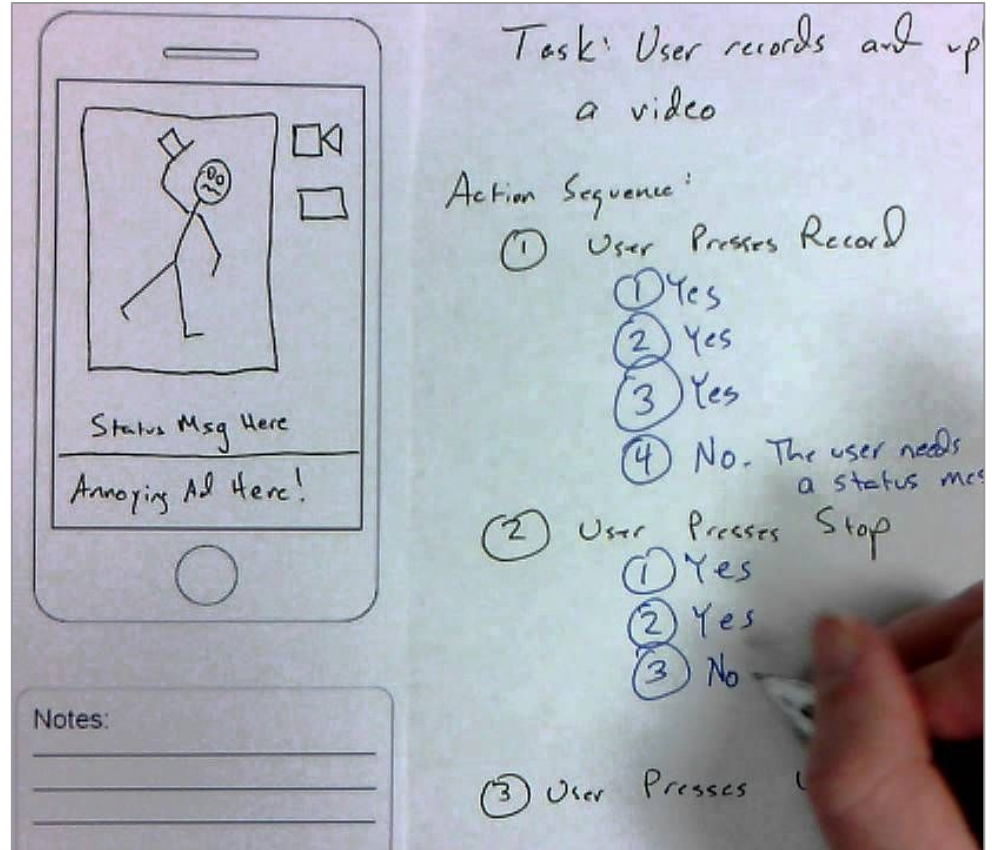
... who advocate for users, but also understand how to apply usability theory.

Cognitive walkthrough

An method without users, that can be used to validate early prototypes.

Designers walk through actions of a scenario's **golden path** and checks:

1. Will the user set the right goal?
2. Will the user notice the correct action?
3. Will the user associate the action with the goal?
4. Will the user see that progress is made towards the goal



Stakeholder review

In a design process, this review process allows you to verify the direction of the project with stakeholders.

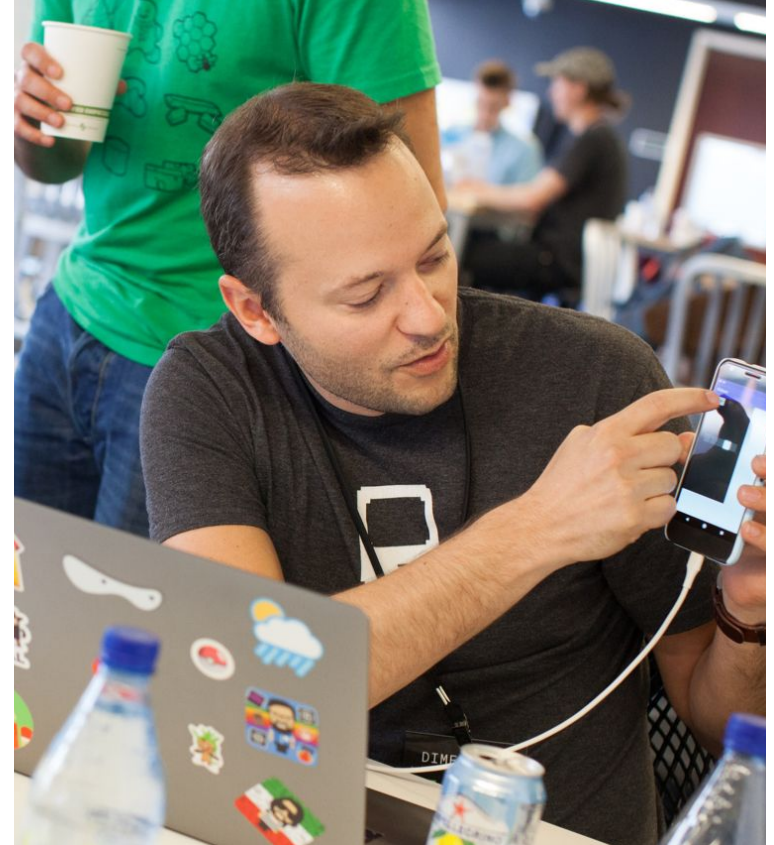
- Schedule as early as possible to ensure that stakeholders are available.
- If possible, conduct the review after user testing so that results and user feedback can be presented.
- Create a presentation with important prototype and feedback information.



Technical review

A similar review process can be conducted to validate technical feasibility of a project prototype.

- Schedule as early as possible to ensure that developers/engineers are available.
- Invite them to join a design sprint for 30 minutes to provide feedback.
- Invite them to participate in user interviews or usability studies to gain insight into usability issues.



Background: Discount usability engineering

"Usability Engineering at a Discount" by Jakob Nielsen proposes 3 important components of usability testing for teams with small budgets:

- **Simplified user testing**
 - Only recruiting a small number of participants.
 - Qualitative studies using the thinking-aloud method.
- **Narrowed-down prototypes**
 - Usually using cheap paper prototypes.
 - Focus on a single path through the product.
- **Heuristic evaluation**
 - Performed by experts using established usability guidelines.

Nielsen Heuristic Evaluation

Nielsen proposed that a team of **expert evaluators**, guided by **usability heuristics**, should evaluate user interfaces. Testing with real users still has a purpose.

Heuristic Evaluation is a method of finding usability problems in a design

- Assess against usability principles
- A formal list of principles is decided in advance
- Evaluators must have familiarity with the principles
- Only a small number of evaluators is required

Compared to user testing...

- Heuristic Evaluation is **easy to coordinate**.
- Heuristic Evaluation is **low-cost and fast**.

What are heuristics?

Jakob Nielsen is a User Advocate and co-founder of the **Nielsen Norman Group**.

While the word **"heuristic"** has a long history of use, this is what Nielsen has to say:

"...Heuristic means 'rule of thumb'. So basically, a very broad design guideline that applies to a wide range of user interfaces. Though not necessarily every single time."

- Jakob Nielsen



10 Usability Heuristics for User Interface Design

A list of principles proposed by Jakob Nielsen.

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose and recover from errors
10. Help and documentation

Each of the heuristics applies usability theory we have already learned!

The heuristics are the result of research and analysis.

Nielsen conducted a factor analysis of 249 usability problems, and these 10 heuristics have the maximum explanatory power across them.

Can you match each heuristic with another usability principle or rule?

10 Usability Heuristics for User Interface Design

Let's compare these to **Shneiderman's 8 Golden Rules** from **Lecture 4**:

- | | |
|---|------------------------------|
| 1. Visibility of system status | Short-term Memory |
| 2. Match between system and the real world | Conceptual Model |
| 3. User control and freedom | Keep users in control |
| 4. Consistency and standards | Consistency |
| 5. Error prevention | Prevent errors |
| 6. Recognition rather than recall | Short-term Memory |
| 7. Flexibility and efficiency of use | Universal usability |
| 8. Aesthetic and minimalist design | Short-term Memory |
| 9. Help users recognize, diagnose and recover from errors | Prevent errors |
| 10. Help and documentation | Universal |

Evaluator performance

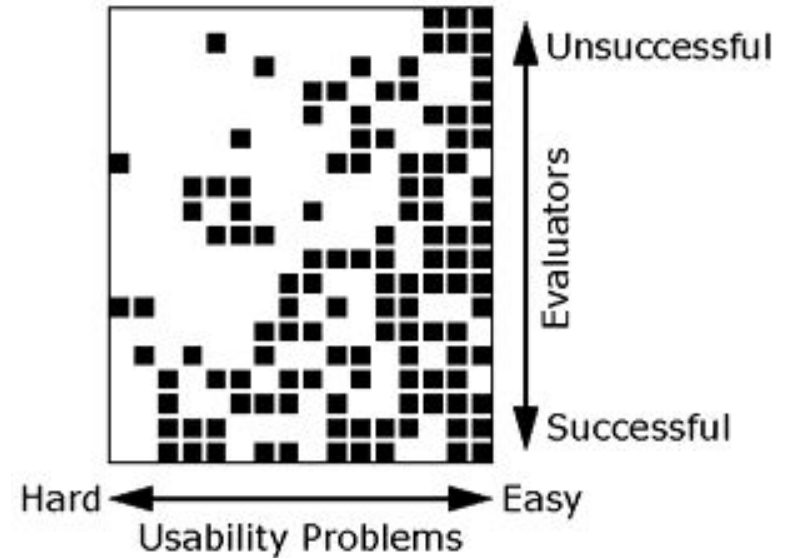
If an expert evaluation still requires experts, why not just recruit users for testing?

- You don't actually need many experts.
- Experts and user find different issues.

The optimal number of evaluators depends requires cost-benefit analysis.

Nielsen found in 6 of his projects:

- **1 evaluator** = about 35% of problems
- **5 evaluators** = about 75% of problems



Above: The results from 19 evaluators assessing an interface to identify 16 errors

Heuristic Evaluation process

The process of conducting a Heuristic Evaluation has a few suggested stages.

1. Select expert evaluators.
2. Define or select a list of heuristics to use for the evaluation.
3. Identify the goal of the system or select a specific task.
4. Understand the general interaction and scope of the system.
5. **Individually** perform tasks and consider heuristics.
6. **Individually** identify specific problems and create a list.
7. Collect and aggregate results from evaluators into a single list of issues.
8. Conduct a debriefing session to brainstorm redesign solutions.

Understanding the evaluation

Before starting a formal heuristic evaluation, it is important to brief the evaluators to ensure a shared understanding of the evaluation requirements.

Selecting and understanding heuristics

- Use existing heuristics (Nielsen's 10 Heuristics, Shneiderman's 8 Golden Rules)
- Define new heuristics - may be more if a specific usability focus is required.

Understand the system

- Basic familiarity with the system improves evaluator efficiency.
- In some cases evaluators may be assigned specific user tasks to attempt.

Understand the heuristic evaluation process and how to assess problems.

The severity of usability problems

3 factors help to determine the severity of a usability problem:

- **Frequency**

- How often did we encounter the problem?
- How many evaluators discovered this problem?

- **Impact**

- How much trouble does the problem cause for users?
- Will it be easy or difficult for the users to overcome?

- **Persistence**

- Is it a one-time problem that can be overcome once we know about it?
- Will the same problem appear repeatedly?

Market impact - the effect on the popularity of the product - can also be considered.

Severity scale for heuristic evaluations

Each factor (frequency, impact, persistence) can be rated using the following scale:

- 0** I don't agree that this is a usability problem at all
- 1** Cosmetic problem only: need not be fixed unless extra time is available
- 2** Minor usability problem: fixing this should be given low priority
- 3** Major usability problem: important to fix, so should be given high priority
- 4** Usability catastrophe: imperative to fix this before product can be released

Individual factors scores can be used to decide a single score **0-4** for the problem. This can be done by averaging or considering weighting of the individual scores.

Individual evaluation

Each evaluator inspect the interface alone. Evaluators will not communicate until the evaluations are complete to prevent evaluators from influencing each other.

A table can be used to keep track of the issues found.

Problem	Heuristic	Frequency	Impact	Persistence	Severity
1. Not all the fields in user sign up have tooltips when hovered using the mouse.	4. Consistency and Standards 10. Help and Documentation	2	1	4	3

Problems must be described in a specific way. A heuristic must be identified and severity factors considered to keep the assessment objective.

Collating results and debriefing

A heuristic evaluation should produce a single list of **unique usability problems**, each with its own severity rating to help prioritise work required to fix the issues.

Evaluators meet and collate results

- Compare individual lists and identify overlaps in problems.
- Produce a final list of only unique problems.
- Where a problem has been found by multiple evaluators, **average the scores**.

Conduct a debriefing session

- Evaluators meet and invite representative of the design team.
- Brainstorm possible design solutions for each problem
- Discuss positive aspects of the design.

Heuristic Evaluation Pros and Cons

Heuristic evaluation works well in many situations, but it's not a perfect solution.

Pros

- Inexpensive and fast.
- Can be repeated frequently!
- Obvious issues quickly revealed.
- "Dumb" issues can be fixed before any user ever sees it!

Cons

- Experts struggle to predict what users will actually do.
- "Experts" and "users" may assign different severities to issues
- Results dependent on evaluator skills, experiences, and biases.

Can you think of a way to help expert evaluators better understand the users' point of view before starting they evaluate an interface?

Discussion Forum 2

Discussion Forum 2

Test what you have learned in this week's lecture and tutorial by conducting your own **individual heuristic evaluation**.

1. Inspect and evaluate one of the **worst** apps on your phone.
2. Create a table for your individual evaluation results.
3. Apply **Nielsen's 10 Usability Heuristics for User Interface Design**.
4. Describe **5** different usability problems and the applicable heuristic violation.
5. Rate each problem and justify based on Nielsen's 3 main factors of severity.
6. Determine a single final severity score for each problem.

Full instructions and posting rules are available within the assessment activity.

Assignment Project

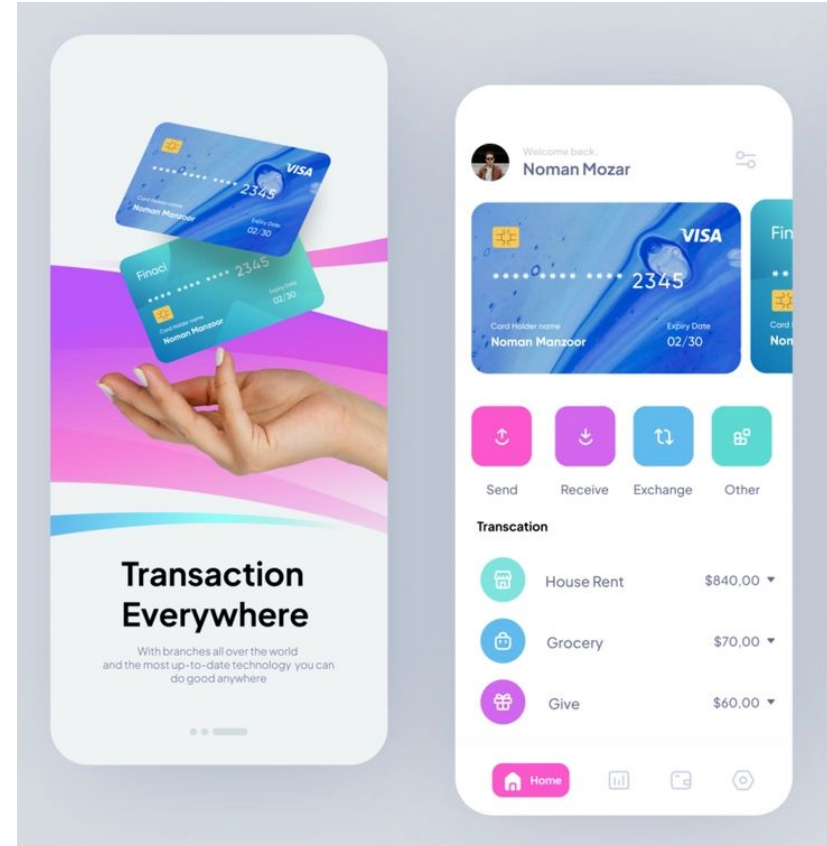
Stage E + F

Stage E - Prototype screens

Implement your mobile app concept as a high-fidelity prototype using **Figma**.

The total number of unique screens in the prototype should be equivalent to 3-4 times the number of group members.

- **2 members = 6-8 unique screens**
- **3 members = 9-12 unique screens**
- **4 members = 12-16 unique screens**



Stage E - High-fidelity prototype

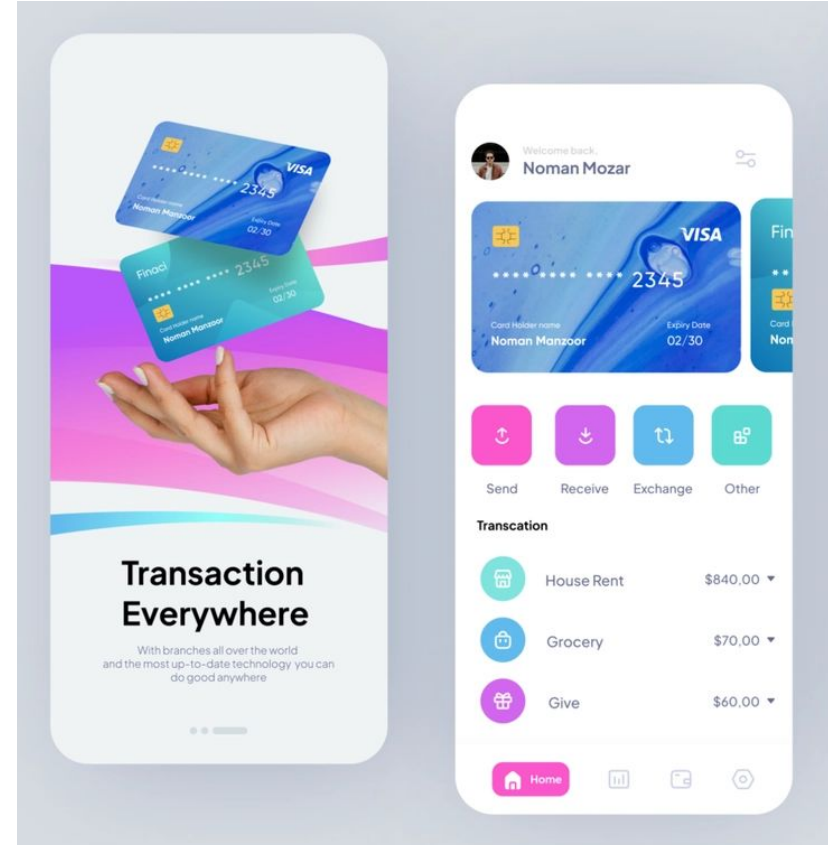
The design of each screen should reflect a close-to-final version of the app's design, which will include:

- Careful application of visual design
- Inclusion of realistic screen content.

Your group may use online resources:

- Images (icons, photos and illustrations)
- Existing Figma UI kits that have been modified to suit the group's needs.

All online resources must be referenced.



Stage E - Additional features

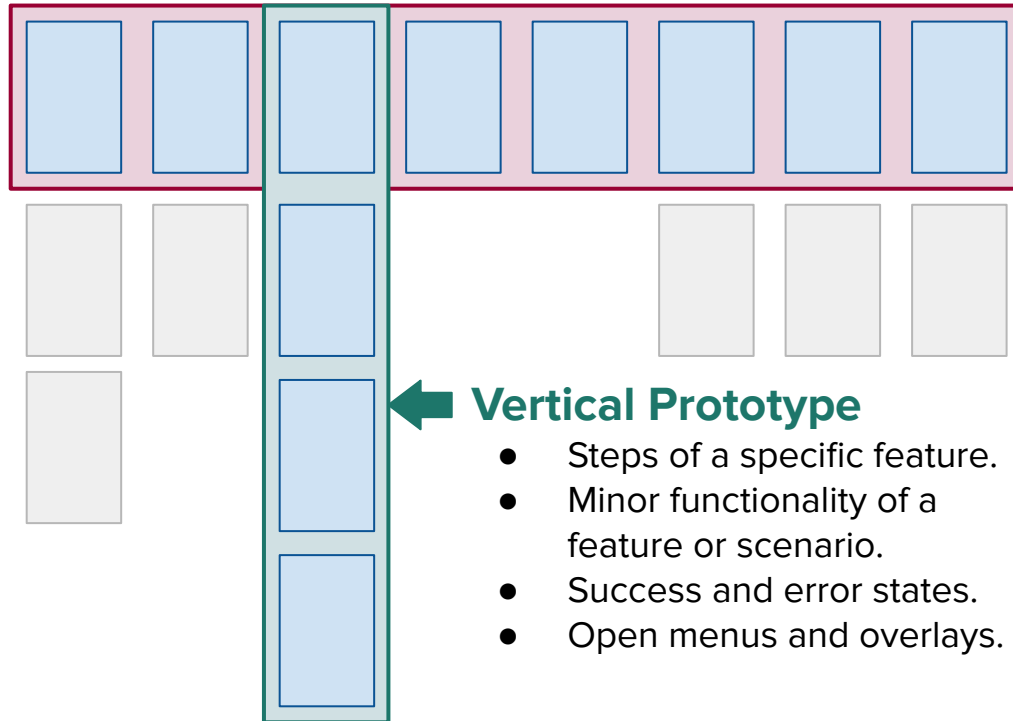
Each group member must incorporate 1 new feature from the following list:

1. Using the device's camera to capture or scan
2. Perform a search within the collection
3. Perform a search on the Internet
4. Use a map to show the location of something
5. Assign a rating or write a review
6. Create and update a user profile
7. Share content as a link or image
8. Earn and keep track of daily points
9. Get recommendations based on usage
10. Send messages to other users

Each group member chooses 1 new feature only, or designs a new implementation of one of the above features not already present in the previous Stage D submission.

Stage E - Prototype depth and breadth

The prototype should present a balance of breadth and depth in the features shown.



Horizontal Prototype

- Depicts the range of features.
- Major sections of an app.
- Screens that depict different major functionality.

Vertical Prototype

- Steps of a specific feature.
- Minor functionality of a feature or scenario.
- Success and error states.
- Open menus and overlays.

It is important for groups to divide workload evenly between members in order to maximise efficiency.

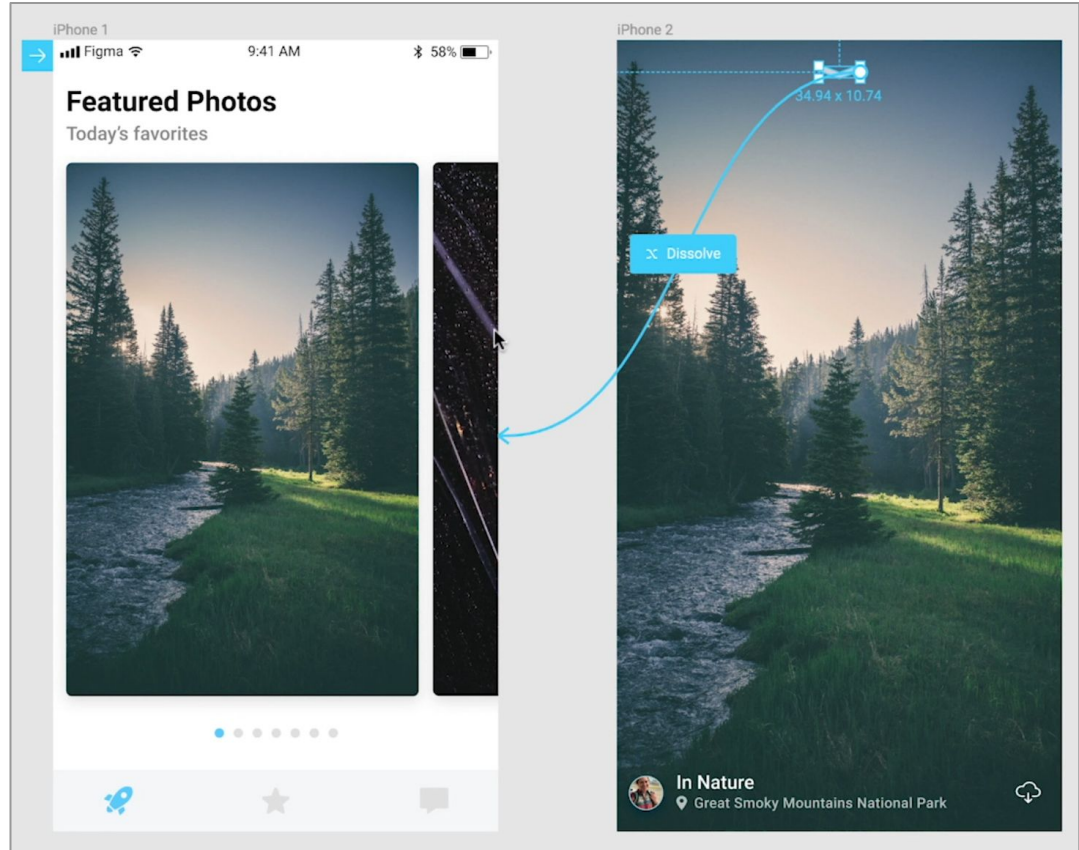
Work collaboratively in Figma to ensure better consistency.

Stage E - Navigation flows

The prototype must include working navigation flows that allow every screen to be accessed during testing using Figma's click interactions.

Note: Overlay pop-ups can be implemented but are not mandatory for this stage.

Advanced data/logic-driven interactions and animations will not be assessed.



Stage E - Group Heuristic Report

Explain how Nielsen's 10 Usability Heuristics has been applied in your prototype, or how screens could be improved with application of a heuristic. The number of heuristics explained must be equal to 2 times the number of group members.

Present each heuristic on a new page with the following:

- The name of the heuristic
- An image of a screen from your prototype that relates to the heuristic
- Clear highlighting of interface elements relevant to the heuristic
- A brief explanation of how the heuristic has been successfully applied or is lacking in the screen design.

Note: Severity ratings are not required. Do not undertake a full evaluation process.

Stage F - Individual in-class presentation

Prepare a short 3-minute presentation that briefly introduces your concept and explains the 3 most important usability features that you have contributed to your team's project.

- Introduce your project concept
- Briefly describe your persona
- 3 important features you have implemented or are currently working on.
- Explain how 3 different design principles, guidelines or heuristics are applied to support these features.

You will need to prepare a slide deck with visuals to support the spoken delivery of your presentation. **You must attend your allocated tutorial class to be assessed.**

Next session

- Accessibility principles and techniques
- Accessibility issues and audits

Reminders

- **P3 replacement classes this week (Australia Day university holiday)**
 - Wed 26 Jan classes replaced with Thu/Fri classes
- **Stage D is due on Friday this week**
 - Low-fidelity prototype exploration
 - Low-fidelity application prototype
 - Theory-based justifications