

# Evaluating Pervasive Systems

---

COMP5047 – Lecture 08

Anusha Withana

Think Aloud – Adapted from slides developed by Prof. Judy Kay

The School of Computer Science  
The University of Sydney

# Evaluation

- How would you find out whether your system
  - Would appeal to intended user?
  - And they will **really** use it!
- Remember “Don’t blame the user!”



[pinterest.com.au/pin/461407924294118013](https://pinterest.com.au/pin/461407924294118013)

# Why Evaluate?

- Designers can focus on real problems and the needs of different user groups
- Make informed decisions about the design
- Enables problems to be fixed before the product goes on sale

# Evaluation

- Primary goal is to improve the design of interface (artifact)
  - E.g. screen sketch, prototype, app, computer system or a component
- Involves collecting and analyzing data about real/potential use cases and users' experiences when interacting with a design artifact

# Two aspects

- Technical Evaluation
  - Does it work as expected?
  - Is it accurate?
  - Technical specification
    - Power consumption
    - Size / Weight / Materials
    - Safety
    - Lifecycle
- User aspects
  - The usability of the system
    - How easy it is to learn and use?
    - Works as intended
  - The users' experiences
    - How satisfying, enjoyable, or motivating

# Technical Evaluation

- Apple Watch Tech Specs



## Display

- **45mm**  
396 by 484 pixels  
1143 sq mm display area
- **41mm**  
352 by 430 pixels  
904 sq mm display area
- **Always-On Retina LTPO OLED display**  
1000 nits brightness
- Accelerometer  
*up to 32 g-forces with fall detection*  
–

## Connectivity

- **LTE and UMTS<sup>6</sup>**  
GPS + Cellular models  
[Learn more about available carriers](#)
- **Wi-Fi**  
802.11b/g/n 2.4GHz and 5GHz
- **Bluetooth 5.0**

## Power

- Built-in rechargeable lithium-ion battery  
*Up to 18 hours<sup>8</sup>*
- USB-C magnetic fast charging cable

[https://support.apple.com/kb/SP860?locale=en\\_GB](https://support.apple.com/kb/SP860?locale=en_GB)

# Technical Evaluation

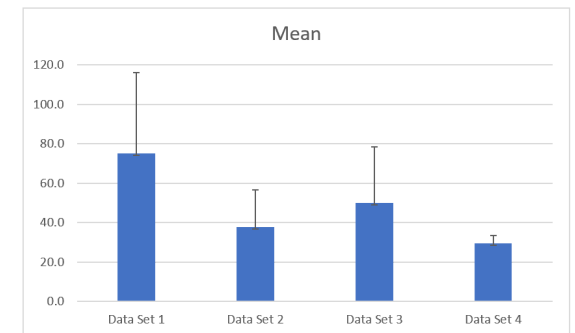
- Consider the case

You are developing a wearable sensor to measure heart rate,

- how do you convince people it works?
- what aspects/qualities/specs you should document?

# Technical Evaluation

- How do you test for accuracy?
  - Compare your system to the ***state-of-the-art***
    - Find out what is the state of the art for heart rate monitoring?  
E.g. ECG
  - Compare in a **controlled setting**
    - Same user group / same test setup
    - Same context
    - Same measures / technical approaches
  - Report **accurately / comprehensively**
    - Do not tamper with the data
    - Always report the margin of error (e.g. average and standard deviation)





# Technical Evaluation

- Consider the case

You are developing a wearable sensor to measure the elbow angle using a flex/bend sensor,

- What will be a possible state of the art?
- What will be a good context?

# Technical Evaluation

- Power consumption
  - Total power is not the collective power consumption of individual elements
    - Best to measure over time
      - You can measure current (I) and voltage (V) consumption. Power is  $V \times I$  in Watts (W)
    - Also under different activities
    - Nowadays usually reported in a readable formats
    - Think about your battery

## Power

- Built-in rechargeable lithium-ion battery  
*Up to 18 hours<sup>8</sup>*
- USB-C magnetic fast charging cable



# Technical Evaluation

- Lifecycle
  - Usually needs a mechanised/automated setup



<https://giphy.com/>

# Technical Evaluation

- Safety
  - Electrical safety
  - Material safety
  - Radiation / Electromagnetic waves
  - Information / data safety

# Evaluating with Users

- You as the designer (and your colleagues) may think that an interface is usable and attractive
- But others may disagree



**"It's the latest innovation in office safety.  
When your computer crashes, an air bag is activated  
so you won't bang your head in frustration."**

funnybizblog.com

# Evaluating with Users

- Evaluation **enables designers to check that their design is appropriate and acceptable** for the **target user** population

# Evaluating with Users

- **Focuses on both usability and user experience**

- The usability of the system

- How easy and helpful it is to learn and to use?

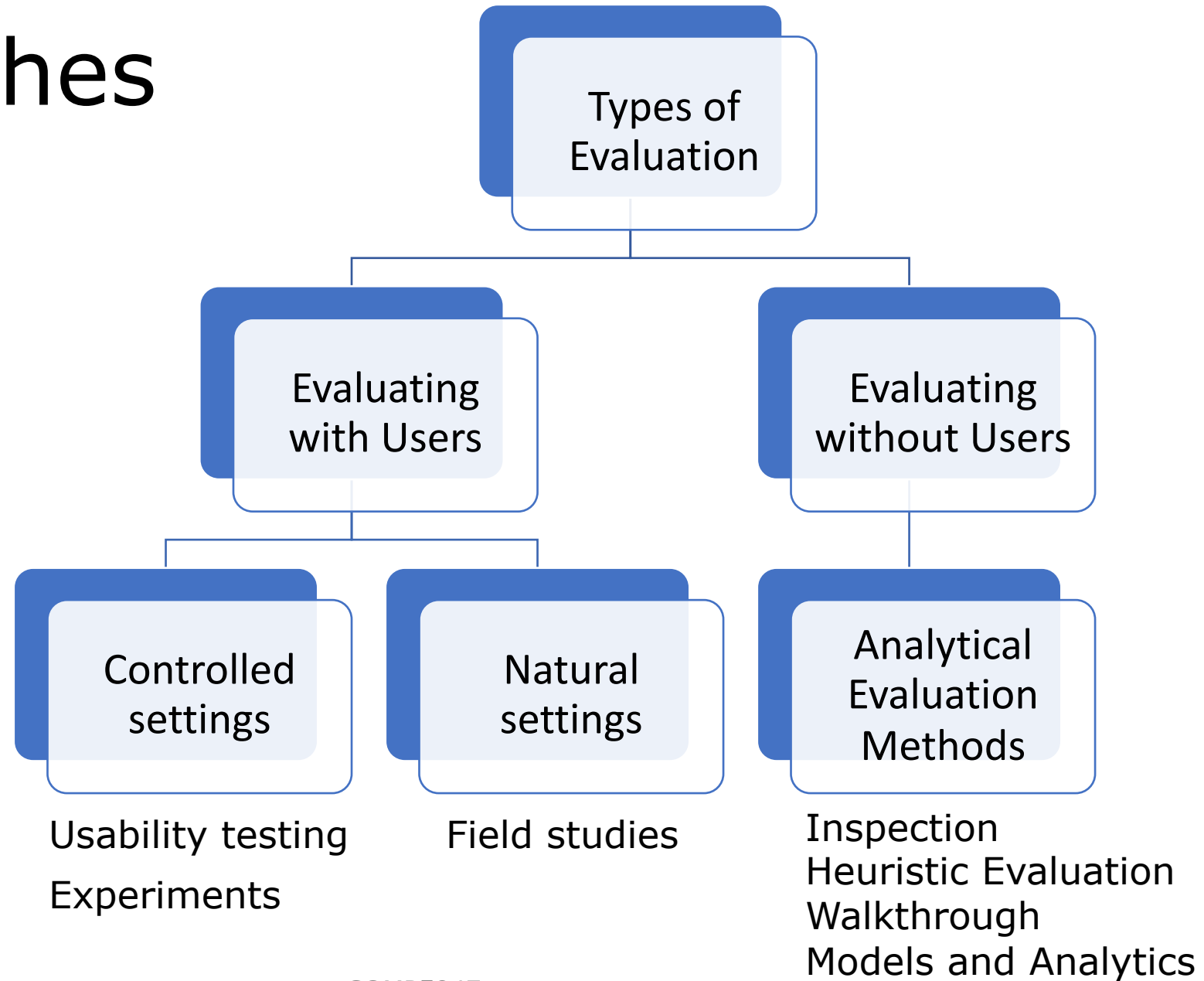


- The users' experiences

- How satisfying, enjoyable, or motivating



# Approaches





# Think-alouds

*"Thinking aloud may be the single most valuable usability engineering method."* I wrote this in my 1993 book, [\*Usability Engineering\*](#), and I stand by this assessment today. The fact that the same method has remained #1 for 19 years is a good indication of the **longevity of usability methods**.



Jacob Nielsen  
nngroup.com

<https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>



## Method Manual: Think Aloud Protocol

[https://www.youtube.com/watch?v=pxsJkAk\\_eo0](https://www.youtube.com/watch?v=pxsJkAk_eo0)

# What do you need for a Think-Aloud

- System
  - Ready to evaluate
- Tasks
  - You prescribe the task (high level)
- Recruit People
  - to do the task; must be representative users
- Observe
  - And record data

# Think-Aloud

- Helps to understand users ***mental model***
  - How the user thinks
- Many tools available
  - Even to do remotely (e.g. you can do over zoom)
- If a GUI interface, participants even can use their own devices

# Think-Aloud - Process

- **Recruit Participants** – Must be representative
  - Demographics: Age, gender, etc.
  - Background: Expertise, experience
  - Interest/Motivation
- How many users?
  - For your projects, try to do at least 4
  - Over 12 is good in general
- Think about the effect on actual user population?

# Think-Aloud - Process

- **Introduction**

- Welcome the participant and explain the purpose
- System, not the participant is tested
- Upfront about what is being recorded
- Opt out at anytime

- **Demo the process**

- Explain
- Invite questions

# Think-Aloud - Process

- **The Test**

- User walks through the task
  - Record
  - Ensure user is supported
  - Show positive feedback for issues identified
  - Help user if they are stuck

- **Task**

- Example task for a wearable physical activity monitor?
  - Find abstract tasks
  - And concrete tasks under each abstract task

# Think-Aloud - Process

- **Questionnaire**

- Open and closing questionnaires for additional data



# Think-Aloud - Process

- **Data**

- Notes
- Video/Audio/Screen recordings
- Software/hardware logs

- **Questionnaire**

- Open questions
  - Best thing about <this system> ?
  - If you could change one thing bout <this system>, what would it be ?
- Other tools
  - UMUX/UMUX-Lite
  - <https://measuringu.com/umux-lite/>

# Think-Aloud - Process

- **Debrief**

- Thank the participant
- Remind the usefulness of results
- Check everything is properly recorded
- Confirm the details with the participant

# Think-Aloud - Pitfalls

- **Defining the task**
  - Test all key aspects
  - Multiple tasks for same aspect
- **Instructions to the user**
  - Do not lead the user
  - Do not use same terms as it appear in the interface
- **Record details you observed**
  - Success/failure/success with help
  - User comments / problems / hesitations / errors

# Think-Aloud - Reporting

- **Summary**

- A table with columns for users and rows for tasks
- Each cell
  - Complete easily (green tick)
  - Complete with difficulty or help (grey tick) – add comments on the problem
  - Could not complete (red-cross) – list the identified issue
- Identify main issues and possible solutions

# Pilot Studies are Critical

- Always run a pilot study
  - Use several representative users
  - Do ASAP
  - Expect major changes
  - plan accordingly



[economictimes.indiatimes.com](http://economictimes.indiatimes.com)

- Discover problems in experimental planning – save a lot of time later

# Ethics

- Evaluation with users involve humans

Ethics approval is an important  
requirement

# University Human Ethics Policy

- Make sure you are familiar with the university policies
  - <https://intranet.sydney.edu.au/research-support/ethics-integrity/human-ethics.html>
- Other relevant documents
  - <https://www.nhmrc.gov.au/research-policy/ethics/national-statement-ethical-conduct-human-research>
  - <https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018>

# Summary

- Understand different types of evaluation methods
- Learn how to do technical and user aspect evaluation
- Plans and conduct a think-aloud
- How to report findings
- Ethical considerations

Image <https://www.dnsstuff.com/>



# Practical Work

- Write the main purpose of the system you develop for your group work (probably already done)
- Abstract Tasks: Briefly describe what a person should be able to do?
- Concrete Tasks: For each abstract task, write two or more concrete tasks (a user should understand what they need to do)
- Run a pilot study within your group (do this in the tutorial)