

# Evaluating Interfaces

COMP2044: Human-Computer Interaction (2024-2025)

---

Matthew Pike

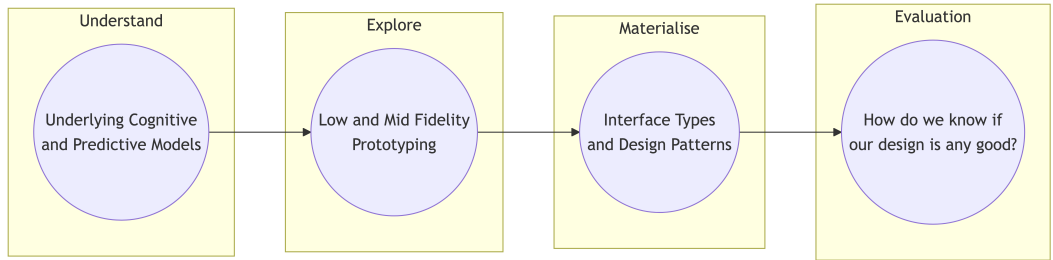
## Overview

---

## Objectives for today

- Reflect on our progress so far, and what's left to do.
- Understand the different types of evaluation methods.
- Considerations when designing and conducting a questionnaire.
- Introduce Cognitive Walkthroughs as an evaluation method.

# Reflecting on What We've Learned So Far



## Evaluation

---

# Evaluation Methods

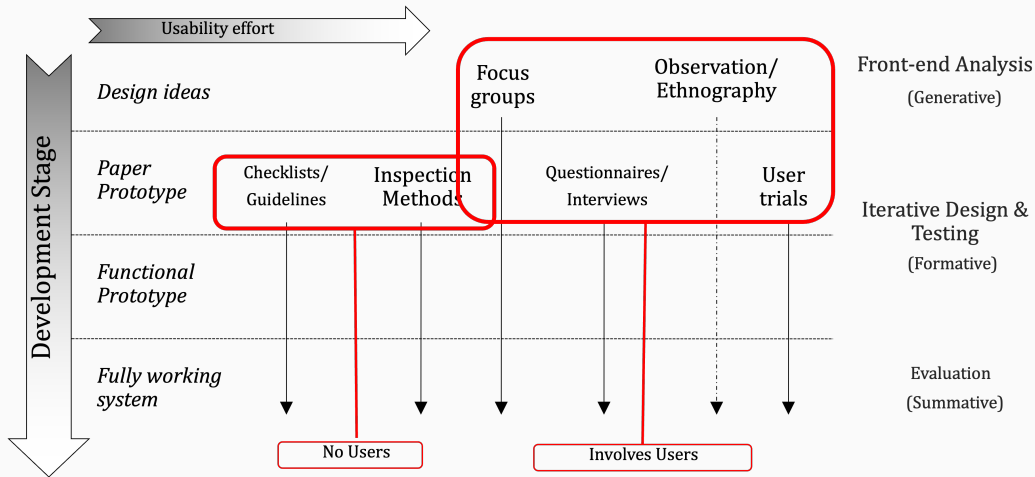
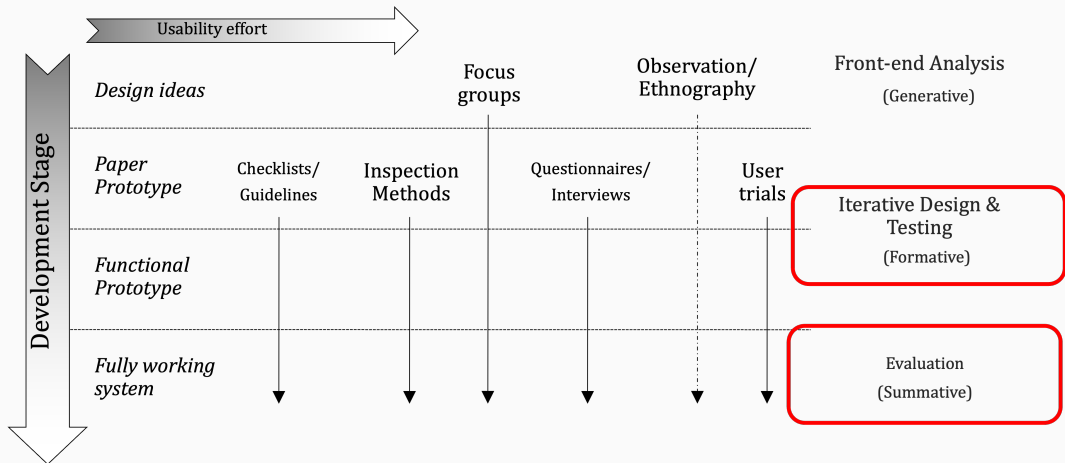


Figure 1: The evaluation methods that are available at various stages of the development process.

## Formative vs. Summative



**Figure 2:** Formative evaluations are conducted during the design and development process, while summative evaluations are conducted after the system has been developed.

## Conducting an Evaluation (1/2)

- Who:
  - Identify the target group. Ensure they are representative of real users.
- Task(s):
  - What tasks will the users perform during the evaluation? Important to be specific and to test before the evaluation.
- Recruit Participants:
  - How many participants? What are the demographics? Is it a representative sample?
  - How to recruit? (e.g. incentives). This can often be the most difficult part of the evaluation.
- Prepare the Evaluation:
  - Protocols, Task, Timings, etc.
  - Instructions: What does the user need to be told to do? Equipment: including test materials.
  - Consent Forms.



## Conducting an Evaluation (2/2)

- Run the Evaluation:
  - Make sure you have signed (informed) consent for taking part (before beginning the evaluation).
  - Encourage users to be critical and reflective. Reinforce - it is the system being tested not them.
  - Collect data (think about how you are recording).
- Generate Results:
  - Executive summary, Aims, Study Design, Methods (Participants, Procedure, Equipment), Results, Discussion, Conclusions.
  - What structure will be best to represent your data?
  - What evidence is appropriate?

- Before conducting any evaluation, it is important to consider the ethical implications of your work. This is often required by your institution/employer.
- Often codes of conduct will include the following:
  - Emphasise system test not user test.
  - Clarify what is expected of the participant and inform that they are free to leave at any point.
  - Explain the purpose of the test (try not to deceive the participants).
  - Make sure the participant is comfortable (or at least normally positioned).
  - Explain that results are confidential and anonymous and how they will be used.
  - Get the participant to agree in writing to take part in the study.
  - Never do anything to embarrass, hurt or otherwise distress the participant.

# Example Consent Form

{STUDY TITLE}

"I state that I am {specific requirements} and wish to participate in a study being conducted by {name/s of researchers/ evaluators} at the {organisation name}.

The purpose of the study is to {general study aims}.

The procedures involve {generally what will happen}.

I understand that I will be asked to {specific tasks being given}.

I understand that all information collected in the study is confidential, and that my name will not be identified at any time.

I understand that I am free to withdraw from participation at any time without penalty"

Signature of participant and date

---

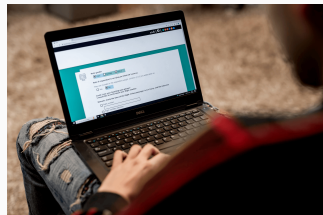
**Figure 3:** The consent form must be completed before data collection with a participant can begin. This is often accompanied by an information sheet to ensure that the participant is providing informed consent.

## Questionnaires

---

# Questionnaires

- A tool for gathering information from users.
- They are cheap and easy to administer. Can be done remotely.
- Very easy to design a bad/misleading questionnaire.
  - E.g. leading questions, double-barrelled questions, ambiguous questions, etc.
- Questionnaires need to be carefully designed to ensure that they are valid and reliable.
  - **Validity:** Does the questionnaire measure what it is supposed to measure?
  - **Reliability:** Does it produce consistent results over time when the same (types of) people are tested?



**Figure 4:** Questionnaires can be completed remotely and scale well to large numbers of participants.

## Types of Questions

- **Factual:** Observable information.
  - E.g. “What is your age?”
- **Opinion:** Attitudes or beliefs (outward facing).
  - E.g. “How satisfied are you with the system?”
- **Attitude:** Attitudes or beliefs (inward facing).
  - E.g. “I find the system easy to use.”

## Question Styles

- **Closed:** Fixed set of responses.
  - Quick and easy to complete and analyse.
  - E.g. “How satisfied are you with the system?” (Very satisfied, Satisfied, Neutral, Unsatisfied, Very unsatisfied)
- **Open:** User gives free-form answers.
  - Captures more detail but harder to analyse.
  - E.g. “What do you like about the system?”

# Example Questionnaire: System Usability Scale (SUS) (Brooke et al., 1996)

- Simple, ten-item attitude Likert scale which provides a broad, subjective assessments of usability.
- Cheap, quick and simple to administer.
- Can be used to compare usability across different systems.

## System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree								Strongly agree
1. I think that I would like to use this system frequently	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
2. I found the system unnecessarily complex	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
3. I thought the system was easy to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
4. I think that I would need the support of a technical person to be able to use this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
5. I found the various functions in this system were well integrated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
6. I thought there was too much inconsistency in this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
7. I would imagine that most people would learn to use this system very quickly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
8. I found the system very cumbersome to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
9. I felt very confident using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				
10. I needed to learn a lot of things before I could get going with this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				
	1	2	3	4	5				

Figure 5: The System Usability Scale (SUS) questionnaire.

## Standard SUS Questions (Brooke et al., 1996)

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.



# Calculating SUS Scores

- The SUS score is calculated as follows:
  - For odd-numbered questions, subtract 1 from the response value.
  - For even-numbered questions, subtract the response value from 5.
  - Total the scores for each question.
  - Multiply the total by 2.5 to get the SUS score.
- The SUS score ranges from 0 to 100, but is not a percentage.
- The average SUS score is around 68.

## System Usability Scale

© Digital Equipment Corporation, 1986.

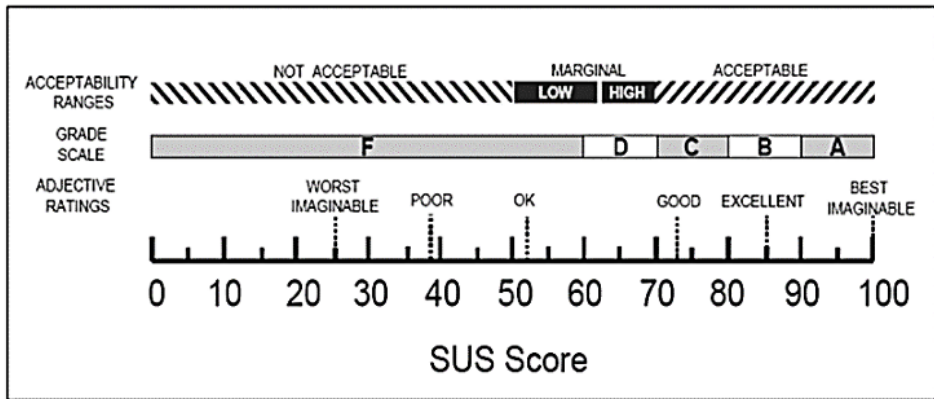
	Strongly disagree						Strongly agree	
1. I think that I would like to use this system frequently							✓	4
2. I found the system unnecessarily complex					✓			1
3. I thought the system was easy to use			✓					1
4. I think that I would need the support of a technical person to be able to use this system	✓							4
5. I found the various functions in this system were well integrated			✓					1
6. I thought there was too much inconsistency in this system					✓			2
7. I would imagine that most people would learn to use this system very quickly			✓					1
8. I found the system very cumbersome to use					✓			1
9. I felt very confident using the system							✓	4
10. I needed to learn a lot of things before I could get going with this system			✓					3

Total score = 22

SUS Score =  $22 \times 2.5 = 55$

**Figure 6:** Example SUS calculation (Brooke et al., 1996).

## SUS Results Interpretation



**Figure 7:** The interpretation of SUS scores. A score of 68 is considered average. This provides a convenient (if imperfect) way to compare the usability of different systems (Brooke, 2013).

## Cognitive Walkthrough

---

# What is a Cognitive Walkthrough?

*“The cognitive walkthrough is a theoretically structured evaluation process that takes the form of a list of questions (Figure 8). The questions focus the designer’s attention on individual aspects of the interface that the CE+ theory claims are important in facilitating the problem-solving and learning processes.”*

(Lewis et al., 1990)

- Many variations of the cognitive walkthroughs (CW) exist. A nice summary of the different approaches was produced by Mahatody et al. (2010).

CE+ Design Walkthrough	Evaluator _____	Date _____
Interface _____	Task _____	Step # _____
<i>Actions/choices should be ranked according to what percentage of potential users are expected to have problems: 0 = none; 1 = some; 2 = more than half; 3 = most.</i>		
1. Description of user's immediate goal:		
2. (First/next) atomic action user should take:		
2a. Obvious that action is available ? Why/Why not?		
2b. Obvious that action is appropriate to goal ? Why/Why not?		
3. How will user access description of action?		
3a. Problem accessing? Why/Why not?		
4. How will user associate description with action?		
4a. Problem associating? Why/why not?		
5. All other available actions less appropriate ? For each, why/why not?		
6. How will user execute the action?		
6a. Problems? Why/why not?		
7. If timeouts, time for user to decide before timeout? Why/why not?		
8. Execute the action. Describe system response:		
8a. Obvious progress has been made toward goal? Why/why not?		
8b. User can access needed info. in system response? Why/why not?		
9. Describe appropriate modified goal, if any:		
9a. Obvious that goal should change? Why/why not?		
9b. If task completed, is it obvious? Why/why not?		

**Figure 8:** The questions posed by Lewis et al., when conducting a CW (Lewis et al., 1990).

## Conducting a Cognitive Walkthrough (Lewis et al., 1990)

1. First, the researcher specifies a series of tasks on which one will evaluate the design.
2. Next, the sequence of user actions that will successfully perform a given task is specified by the researcher.
3. Finally, the main part of the walkthrough involves using the questionnaire to evaluate the ease of learning for the proposed design on a particular task.
  - If an evaluator expects no problems at a given step, that judgment has to be defended.
  - If problems are expected, they should be documented.
4. At each stage, the evaluator asks themselves a series of questions (Figure 8, for example) about the interface.
  - There are many variations of the questions that can be asked. Another example is given by the **NNGroup**.

# NNGroup Cognitive Walkthrough Questions

Task				Action success		Action failure	
Action step							
Will the user try to achieve the right result?	yes <input type="checkbox"/>	from experience	the system tells them to	no <input type="checkbox"/>			
Will the user notice that the correct action is available?	yes <input type="checkbox"/>	from experience	they would see a call-to-action	no <input type="checkbox"/>			
Will the user associate the correct action with the effect they're trying to achieve?	yes <input type="checkbox"/>	from experience	a prompt/label matches action	no <input type="checkbox"/>			
After the action is performed, will the user see that progress is being made toward the goal?	yes <input type="checkbox"/>	from experience	there's a connection between the system response and user goal	no <input type="checkbox"/>			

Figure 9: Available via NNGroup -

[https://media.nngroup.com/media/articles/attachments/Blank\\_Cognitive-Walkthrough\\_Template.pdf](https://media.nngroup.com/media/articles/attachments/Blank_Cognitive-Walkthrough_Template.pdf)

*“A persona is a fictional, yet realistic, description of a typical or target user of the product. It is used to promote empathy, increase awareness and memorability of target users, prioritize features, and inform design decisions.”*

NNGroup

- Popularised by Alan Cooper in his book “The Inmates are Running the Asylum” (Cooper, 1999).
- Personas (Chang et al., 2008) allow:
  - Designers to focus on the primary user, his/her behavior patterns and needs.
  - Researchers to categorise feedback from real users onto a small number of personas - making it easier to understand and act upon.
- Personas can be used in conjunction with CW to better model the approaches that different users might take to complete a task.

- System Usability Scale (SUS) Score Calculator
  - <https://stuart-cunningham.github.io/sus/>
- SUStified? Little-Known System Usability Scale Facts
  - <http://uxpamagazine.org/sustified/>
- Evaluate Interface Learnability with Cognitive Walkthroughs
  - <https://www.nngroup.com/articles/cognitive-walkthroughs/>
- How to Conduct a Cognitive Walkthrough
  - <https://www.interaction-design.org/literature/article/how-to-conduct-a-cognitive-walkthrough>
- Personas Make Users Memorable for Product Team Members
  - <https://www.nngroup.com/articles/persona/>



## References

- Brooke, J. et al. (1996). SUS-a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7.
- Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, 8(2), 29–40.
- Chang, Y., Lim, Y., & Stolterman, E. (2008). Personas: From theory to practices. *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges*, 439–442.
- Cooper, A. (1999). *The inmates are running the asylum*. Springer.
- Lewis, C., Polson, P. G., Wharton, C., & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 235–242.
- Mahatody, T., Sagar, M., & Kolski, C. (2010). State of the art on the cognitive walkthrough method, its variants and evolutions. *Intl. Journal of Human–Computer Interaction*, 26(8), 741–785.