

Evaluating the Interfaces I

Prepared by:

Dr. Fahad Ahmed

Topics of Discussion

- Personas
 - Why needed?
 - How to make them?
- Interface Evaluation
 - What is Interface Evaluation?
 - Heuristic Evaluation
- How to use the data collected?
 - Quantitative Analysis
 - Qualitative Analysis

Recap: What We Know So Far

You now have two complementary toolkits — this week they combine to evaluate a real interface.

💡 Asking People (Lecture 4)

Questionnaires & SUS

Structured scales that quantify user satisfaction

Interviews

One-to-one exploration of goals, frustrations, and mental models

Focus Groups

Group discussion surfaces shared attitudes and priorities

👁 Watching People (Lecture 5)

Think-Aloud Protocol

Participants narrate reasoning — makes invisible thought visible

Direct Observation

Naturalistic, controlled, or structured watching in context

Analytics & Log Data

Behavioural patterns at scale: completion, error, drop-off rates

💡 *Evaluation applies both toolkits to a specific target — an interface — to answer: Does it work for the people who use it?*

What is a Persona?

A research-based fictional character representing a real user type — not a stereotype, but a synthesis of observed behaviour [1, 2].

"Personas are user archetypes you can use to help guide decisions about product features, navigation, interactions, and even visual design."

— Cooper et al. (2014) [1]

✓ What a persona IS



Data-driven

Built from real user research — interviews, observations, surveys. Every detail should be traceable to a finding, not invented.



Goal-oriented

Defined by what the user is trying to achieve, not just who they are. Goals drive design decisions more than demographics.



Repeatable

A shared reference across the whole team — designer, developer, and stakeholder all mean the same user when they say 'Brenda'.



Contextual

Includes the environment, constraints, and tools the user actually works with — not an idealised context.

✗ What a persona is NOT



Not a stereotype

"Tech-savvy millennial" or "elderly technophobe" — these are assumptions, not research. Personas challenge stereotypes, they don't encode them.



Not a made-up character

Inventing a persona without research gives the team a fictional user who validates whatever you already wanted to build.



Not a marketing segment

Market segments describe buying behaviour. Personas describe usage behaviour — what people do, not what they buy.



Not permanent

Personas become outdated. If your user research changes, your personas must change too.

[1] Cooper, A., Reimann, R., Cronin, D. and Noessel, C. (2014). *About Face: The Essentials of Interaction Design*. 4th edn. Indianapolis: Wiley.

[2] Pruitt, J. and Adlin, T. (2006). *The Persona Lifecycle*. San Francisco: Morgan Kaufmann.

How to Build a Persona

Six steps from raw research data to a usable, team-facing persona document [1, 3].

1

Conduct user research

Interviews, observations, surveys — minimum 5 participants per major user group. Collect goals, frustrations, workflows, and mental models. Don't start a persona without real data.

2

Identify behavioural variables

Look for axes of difference: experience level, frequency of use, primary goal, attitude to technology, context of use. These become the dimensions that separate your user types.

3

Cluster users into types

Group participants who share similar behaviours along your variables. Aim for 2–4 distinct clusters — one per major user type. Avoid forcing everyone into one cluster.

4

Build the skeleton

Give the cluster a name, photo, age, role, and a one-line goal. Add primary and secondary goals. Include a direct quote from a real participant that captures their mindset.

5

Add context and detail

What tools do they use? What environment do they work in? What are their pain points with current systems? What would success look like for them? Every detail should earn its place.

6

Validate and circulate

Check the persona against your raw data — does it feel true? Share with the team and check for recognition. Personas that surprise your participants are probably wrong.

[1] Cooper et al. (2014). *About Face*. 4th edn. Wiley.

[3] Goodwin, K. (2009). *Designing for the Digital Age*. Indianapolis: Wiley. (Ch. 6 — Creating Personas)

Worked Example — WardWatch Persona

Built from interview data collected with staff from a job site. Each detail maps to a real research finding [4].



Company “Investigator”

Rosa Cho

Content Strategist, Freelance

Age: 34

Location: Seattle, WA

“I’m looking to join the right company that challenges me and allows me to grow and develop my skills.”

About Rosa

Rosa does not believe in settling. She won’t settle for a job with a company that isn’t as innovative and cutting edge as she believes she deserves. She wants to get the most out of every professional experience, and before moving to a new position, Rosa investigates every angle of aligning herself with a company.

Behavioral Considerations

- + Expects the site experience to reflect the business’s culture and values
- + Interested in career opportunities within the organization that fit her career goals
- + Thoroughly compares multiple companies with similar opportunities
- + Is interested in the unique benefits of working at a company, including cultural elements, mentoring programs, and continuing education policies
- + Needs to be confident the company has innovative products that will be interesting to work on
- + Needs to know company has reputable partners and customers

“I crave variety in the types of industries and goals of each content project I work on. I need to ensure I won’t get bored.”

Frustrations

- + Thinks that too many companies have career sections that just talk about open positions but not why she would actually want to work there
- + Would like to challenge herself and have a more stable job, but is comfortable as a freelancer and wouldn’t stop for just any job

Goals

- + Needs to see reasons why a company is interesting: has it won awards, had intense growth, won big contracts?
- + Wants to figure out how to get in touch with someone at the company to explore opportunities further

Tasks

- + Learn about current customers and success stories
- + Read press releases about recent big contract wins and other accolades
- + Read about culture, benefits and perks, and the people that work there
- + View job openings and apply

[4] <https://www.nngroup.com/articles/persona/>

Worked Example — WardWatch Persona

Built from interview data collected with ward staff across two NHS sites. Each detail maps to a real research finding.



Brenda Okafor

Senior Ward Nurse · 14 years NHS experience · Age 38

Tech comfort

Medium

Shift pattern

Rotating

Primary device

Desktop

Team size

12 staff

"I don't need it to be pretty — I need it to work when I'm in the middle of a handover with five things happening at once."

PRIMARY GOAL

Complete handovers accurately in under 5 minutes — zero missed tasks.

SECONDARY GOAL

Help junior staff find information without interrupting her.

 Pain Points (from interview data)

- "The search function makes me click three times to find a patient I've been looking after all week."
- Navigating between screens during a handover causes errors — she loses her place and has to start again.
- Audit log exports take 40 seconds — long enough that she gives up and writes on paper instead.

 Observed Behaviours

- Keeps a paper backup of the handover template — does not trust the system to save mid-entry.
- Uses keyboard shortcuts exclusively; never uses the mouse for navigation tasks under time pressure.
- Teaches new staff the workarounds before she teaches them the system itself.

 Design Implications for WardWatch

- Speed over aesthetics:** Every interaction must be completable in ≤ 2 clicks. Remove decorative elements that increase cognitive load.
- Auto-save, always:** Any unsaved state that can be lost to a navigation error is a safety risk, not just a UX inconvenience.

Using Personas Effectively

A persona on a shelf is useless. These practices make them a living design tool [1, 2, 5].

✓ Practical Tips



Print them and put them on the wall

Personas kept only in a document get forgotten. Physical presence in your workspace keeps them in every conversation — especially during design reviews.



Use the persona's name in meetings

Don't say "the user". Say "Would Brenda be able to do this in a handover?" Naming forces specificity and stops abstract debates about hypothetical users.



One primary persona per design

If you design for everyone you design for no one. Identify the one persona whose satisfaction is the primary success criterion. Secondary personas set constraints, not direction.



Return to the data when personas conflict

When two team members cite different personas to justify opposing decisions, go back to the research. The data resolves persona disputes — opinion does not.



Update after each research cycle

Run a new round of interviews every 6–12 months, or after a major product change. Personas that no longer match reality will lead you to design for users who no longer exist.



Link personas to user stories and tasks

"As Brenda, I need to complete a handover in under 5 minutes so that I do not make medication errors at shift change." Personas without tasks are just portraits.

✗ Common Mistakes to Avoid



The 'average user' persona

A persona built on mean values — average age, average experience, average goals — represents nobody. Cluster on behaviour, not averages.



Too many personas

More than 4–5 personas and teams stop using them. Each persona must represent a meaningfully different user type, or merge them.



Demographic-only personas

"Age 45, female, works in admin" tells designers nothing about what to build. Goals, frustrations, and behaviours are the actionable data — not the demographics.



Skipping the quote

A direct quote from a real participant anchors the persona in reality and stops the team projecting their own assumptions onto it.



Where personas fit in the Double Diamond



Built from raw research data

Used to frame 'How Might We' questions

Evaluate prototypes against persona goals

Validate that the shipped product serves them

What is Interface Evaluation?

A systematic process for assessing whether an interface meets user needs [6, 7].

ISO 9241-11 (2018) defines usability as: effectiveness (can users complete tasks?), efficiency (at what cost?), and satisfaction (how pleasant is it?). Evaluation measures all three.

WHEN do we evaluate?

- Throughout the design lifecycle — not just at the end
- Formative: during design to guide iterations
- Summative: after deployment to measure performance
- Each Double Diamond phase calls for a different evaluation approach

WHAT do we evaluate?

- Task completion — can users reach their goal?
- Error frequency and severity
- Time on task — efficiency
- Learnability — can new users get started?
- Subjective satisfaction — does it feel right?

HOW do we evaluate?

- Expert reviews (heuristic evaluation)
- User-based tests (think-aloud, observation)
- Questionnaires (SUS, custom surveys)
- Analytics (log data, session recordings)
- Often: multiple methods combined

Heuristic Evaluation

Expert inspection against established usability principles — fast, cheap, and effective before user testing [8, 9].

What is it?

- 3–5 evaluators independently inspect the interface
- Each evaluator checks the interface against Nielsen's 10 heuristics [5]
- Problems are rated by severity (0 = not a problem → 4 = usability catastrophe)
- Evaluators consolidate findings after individual reviews
- No users required — makes it ideal for early rapid cycles
- Finds ~75% of usability problems at low cost [6]

Nielsen's 10 Heuristics [5]	
H1	Visibility of system status
H2	Match between system and real world
H3	User control and freedom
H4	Consistency and standards
H5	Error prevention
H6	Recognition rather than recall
H7	Flexibility and efficiency of use
H8	Aesthetic and minimalist design
H9	Help users recognise, diagnose, and recover from errors
H10	Help and documentation

[8] Nielsen, J. (1994). Heuristic evaluation. In: Nielsen, J. and Mack, R.L. (eds.) Usability Inspection Methods. New York: Wiley.
 [9] Nielsen, J. (1994). Usability Engineering. San Francisco: Morgan Kaufmann.

Roomba Robot Redesign



1. Overview

The Roomba 680 [1] helps users make efficient use of time through automatic and scheduled cleaning. This includes cleaning while users are not in the room or home. The interface design is so minimal that it is unusable. It is not accessible for visually impaired users. Its place on the floor means the only way to access the interface is to bend down; access for people who have physical disabilities may be hard. The purpose of this project is to re-imagine the Roomba to be more accessible and usable.

2. Interface Critique



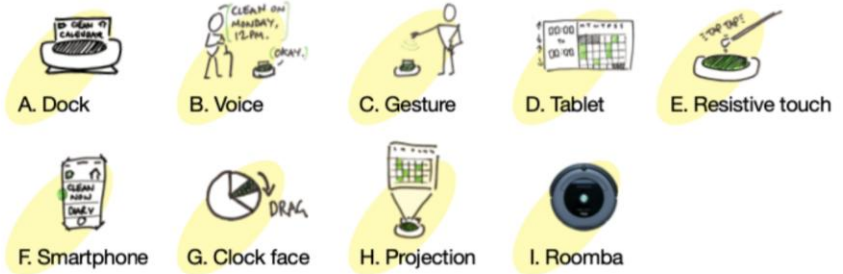
▲ Roomba 680

I identified design problems according to Nielsen's 10 Usability Heuristics [2]:

HEURISTIC	DESIGN PROBLEMS
Visibility of system status	1. Beep language is specialist and inconsistent
	2. Clock interface is unclear, areas light up inaccurately
	3. No cleaning progress indicator
User control and freedom	4. Undoing schedules requires multiple buttons and press-lengths
	5. Clock scrolling is one-way, wasting time and frustrating users
Consistency and standards	6. 'D, H, M' is not a common, recognisable standard
	7. Battery symbol is unconventional, only outline colour changes to show charge
	8. 'Clean' has inconsistent or contradictory functions, dependent on invisible system information, such as cleaning status
Recognition rather than recall	9. Users must remember complex button sequences and combinations
	10. Users must remember cleaning cycle start time to estimate finish time
Flexibility and efficiency of use	11. There is one way to perform each function
Help and documentation	12. Users must learn the interface. However, manuals do not show results of inputs, meaning users do not know if the machine is responding the way it should

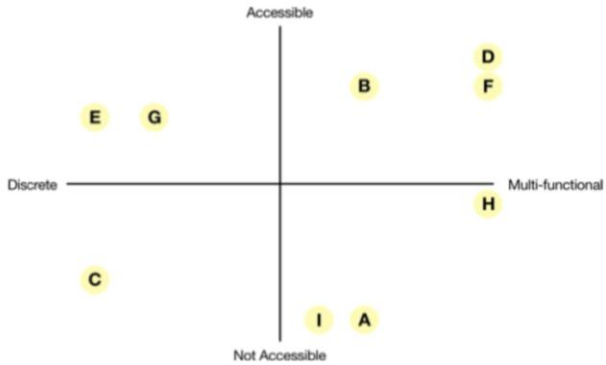
3. Interaction Ideation

Using the Crazy 8s technique [3] to quickly come up with a variety of unrestricted ideas, I sketched options to redesign interaction types.



4. Idea Evaluation

The target user groups for my design work are visually impaired people and people who might find it difficult to reach down to floor-level, such as elderly people and people with physical disabilities, as this is the biggest area for improvement. Motor control is affected in old age, so the device will be difficult to pick up. Older people may also be less digitally confident [4]. I therefore chose accessibility as the most important requirement for the device. Multi-functionality came second, as the functions of an automatic cleaning machine are many. I mapped each idea against these two criteria to understand how well they might meet accessibility and functionality requirements. Among novel interaction techniques, B. Voice and E. Resistive touch ranked best in this process.



[1] iRobot. Roomba. Retrieved from <https://www.irobot.co.uk/roomba/e-series>.
 [2] Nielsen, J. 1994. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 24 - 28, 1994, Boston, MA, 152-158. DOI: <https://doi.org/10.1145/191666.191729>.
 [3] Google. Crazy 8's. Retrieved from <https://designsprintkit.withgoogle.com/methodology/phase3-sketch/crazy-8s>.
 [4] Czaja S, Shant J. 1993. Age differences in the performance of computer-based work. *Psychology and Aging* 8, 1, 59-67. DOI: <https://doi.org/10.1037/0882-7974.8.1.59>.

Quantitative Analysis

Six methods for making sense of numbers from closed-ended questions and SUS scores — each illustrated with WardWatch / ShiftSmart data [10, 11, 12, 13].

1 Descriptive Statistics

Summarise responses: mean, median, mode, standard deviation.

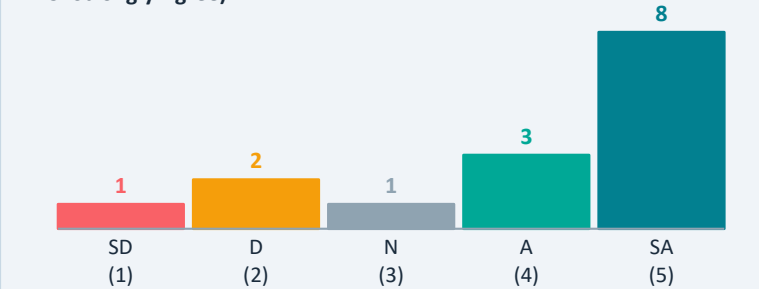
Participant	WardWatch	ShiftSmart	Diff
P1	76	68	-8
P2	82	71	-11
P3	65	80	+15
P4	78	75	-3
P5	70	83	+13
Mean	74.2	75.4	+1.2

SUS mean: WardWatch 74.2 (Good) · ShiftSmart 75.4 (Good) — near-identical means mask individual variation.

2 Frequency Distribution

Count how many chose each Likert option per item — reveals spread, not just average.

Q: "I found WardWatch easy to use" (1=Strongly Disagree → 5=Strongly Agree)



11/15 agreed or strongly agreed. Skew toward SA — positive usability signal.

3 SUS Scoring Procedure

10 items, 1–5 scale. Odd items: score – 1. Even items: 5 – score. Sum × 2.5 [7]

Item	Raw	Adjusted
Q1 (odd)	4	4 – 1 = 3
Q2 (even)	2	5 – 2 = 3
Q3 (odd)	5	5 – 1 = 4
Q4 (even)	1	5 – 1 = 4
Q5 (odd)	4	4 – 1 = 3
...Q10	–	...

Sum of adjusted scores = 29 → 29 × 2.5 = **72.5 = GOOD ✓**

4 Cross-tabulation

Compare responses across subgroups — e.g. novice vs experienced users.

Q: "I felt confident using the system" (% Agree / Strongly Agree)

Group	WardWatch	ShiftSmart
Novice users (n=8)	87.5%	50.0%
Experienced (n=7)	57.1%	85.7%
All (n=15)	73.3%	66.7%

Aggregated means hide subgroup differences — cross-tabs reveal that system preference depends on experience.

5 Non-parametric Testing

Likert data is ordinal — do not assume a normal distribution. Use Mann-Whitney U to compare two groups.

~~X~~ t-test assumes interval data and normality — Likert scales are ordinal; avoid unless justified [3].

✓ Mann-Whitney U: ranks responses, no normality needed. Use for comparing two systems or two groups.

Example: SUS scores — WardWatch vs ShiftSmart
 $U = 31$, $p = 0.48$ → No significant difference ($\alpha = 0.05$)

For correlation between two ordinal scales, use Spearman's ρ (rho), not Pearson's r .

6 Reliability — Cronbach's Alpha

If you design a custom multi-item scale, alpha measures whether all items consistently measure the same construct.



Custom 6-item learnability scale applied to WardWatch:
 Cronbach's $\alpha = 0.81$ → Good internal consistency.
 All 6 items reliably measure the same construct.

Note: SUS is already validated — you only need alpha for your own custom scales.

[10] Brooke, J. (1986). SUS: A 'quick and dirty' usability scale. In: Jordan et al. (eds.) (1996). Usability Evaluation in Industry. London: Taylor & Francis.

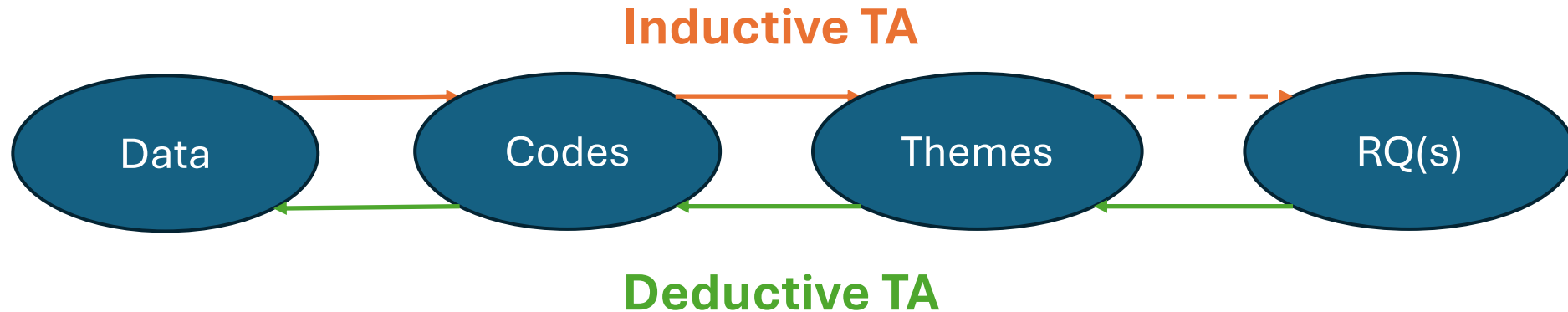
[11] Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics. 5th edn. London: Sage.

[12] Jamieson, S. (2004). Likert scales: how to (ab)use them. Medical Education, 38(12), pp.1217–1218.

[13] Norman, G. (2010). Likert scales, levels of measurement and the 'laws' of statistics. Advances in Health Sciences Education, 15(5), pp.625–632.

Qualitative Analysis: TA

- Thematic Analysis (TA) :
 - Can be inductive/deductive



Qualitative Analysis: TA

TA : Coding

Observations: *Participant 1*

- The participant mentioned that they would generally avoid digital banking due to literacy barriers. Sometimes the participant relied on family members for navigating the app.
- When interacting with the app, the participant found it difficult to switch accounts on the app with single-handed operation. Apart from this, the participant also looked at recent transactions and transfer between accounts often, but all of these are at the top the app and the participant found it hard to reach in single-handed operation.
- The participant also struggles to find loan application options, which are not present on the home page of the app but are buried behind Investment Options->Borrowing->Loans.
- When shopping online, the participant hesitated before entering bank details on the website and also double checked the transaction amount when confirmation was asked from the banking app.
- Sometimes the participant took longer than the 60 second time window to approve the transaction and so had to repeatedly try to complete a transaction but got errors on some occasions. The participant was not able to connect the errors to the cause in the first few times.

Literacy
problems

Single-handed
operation issue

Cannot find loan
option

Online Shopping
issue

Online
Shopping issue

Qualitative Analysis: TA

TA : Coding

Observations: *Participant 1*

- The participant mentioned that they would generally avoid digital banking due to literacy barriers. Sometimes the participant relied on family members for navigating the app.
- When interacting with the app, the participant found it difficult to switch accounts on the app with single-handed operation. Apart from this, the participant also looked at recent transactions and transfer between accounts often, but all of these are at the top the app and the participant found it hard to reach in single-handed operation.
- The participant also struggles to find loan application options, which are not present on the home page of the app but are buried behind Investment Options->Borrowing->Loans.
- When shopping online, the participant hesitated before entering bank details on the website and also double checked the transaction amount when confirmation was asked from the banking app.
- Sometimes the participant took longer than the 60 second time window to approve the transaction and so had to repeatedly try to complete a transaction but got errors on some occasions. The participant was not able to connect the errors to the cause in the first few times.

Literacy
problems

Navigation not
intuitive

Single-handed
operation issue

Similar features
not grouped
together

Important features
not accessible
directly

Fear of Fraud

Fear of Fraud

Error Messages are
not clear

Qualitative Analysis: TA

TA : Themes

Codes	Themes
Literacy problems	User Demographic Barriers
Fear of Fraud	
Navigation not intuitive	Usability Challenges
Single-handed operation issue	
Similar features not grouped together	
Important features not accessible directly	
Error Messages are not clear	

Thank you for your attention!

Questions?

Feedback



[Menti Link:](https://www.menti.com/alpy3vccybur)

<https://www.menti.com/alpy3vccybur>

Menti Code: **5647 6463**