

Evaluating the Interfaces II

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 Covered in Lecture 6

Three tools now in your evaluation toolkit, this session adds a few more and shows you how to act on what you find.



Personas

- Research-based fictional user types
- Goal-oriented, data-driven, contextual
- Brenda Okafor: Senior Ward Nurse, handover under time pressure
- Used across the Double Diamond to test decisions against real user goals



Heuristic Evaluation

- Expert inspection against Nielsen's 10 heuristics
- 3–5 evaluators, working independently
- Severity rated 0–4 (cosmetic → catastrophe)
- Finds ~75% of usability problems cheaply, before any user testing



Analysing Data

- Quantitative: SUS scoring, descriptive stats, Mann-Whitney U
- Qualitative: Thematic Analysis
- TA example: literacy problems, navigation issues, error messages
- Both perspectives needed for a complete picture

Cognitive Walkthrough — Worked Example

Task: Brenda needs to flag a patient as high-risk in WardWatch. She has never done this before [1].

Step	Action	Q1: Know?	Q2: Notice?	Q3: Connect?	Q4: Feedback?	Verdict
1	Locate patient 'Edwards' on ward list	✓	✓	✓	✓	✓ Pass
2	Open patient record	✓	✓	?	✓	⚠ Q3
3	Find the 'Risk Flag' option	✗	✗	—	—	✗ Fail
4	Click the flag icon	—	✗	—	—	✗ Fail
5	Read confirmation message	—	—	—	✗	✗ Fail

 Key finding: The risk flag icon is invisible to a first-time user — it is neither labelled nor logically grouped with patient data. Steps 3–5 all fail. This maps to Nielsen H1 (visibility) and H6 (recognition over recall).

[1] Wharton, C. et al. (1994). The cognitive walkthrough method. In: Usability Inspection Methods. New York: Wiley. Task scenario and data are illustrative; method is authentic.

Heuristic Evaluation vs Cognitive Walkthrough

Two expert methods, these are complementary, not competing [1, 2, 3].

Dimension	Heuristic Evaluation	Cognitive Walkthrough
Question asked	Does this violate a usability principle?	Would a new user succeed at this step?
Origin	User's experience of the whole interface	Task-based, step-by-step simulation
Input needed	Interface + heuristics list	Interface + task scenario + user profile
Output	Prioritised problem list (severity 0–4)	Pass/fail for each action + problem descriptions
Effort	~2 hours per evaluator (whole interface)	~30 min per task (thorough but focused)
Best for	Finding broad range of issues quickly	New users, learnability, specific critical tasks
Misses	Task-specific new-user errors	Global inconsistencies, aesthetic issues
Combine?	✓ Run both — HE first (broad), then CW (task-specific) for key user flows	

[1] Wharton et al. (1994). Cognitive walkthrough. In: Usability Inspection Methods. Wiley. [2] Nielsen, J. (1994). Heuristic evaluation. In: Usability Inspection Methods. Wiley. [3] Jeffries, R. et al. (1991). User interface evaluation in the real world: a comparison of four techniques. CHI '91. pp.119–124.

WCAG 2.1 — Web Content Accessibility Guidelines

The international standard for web accessibility, published by W3C. Organised around four core principles [4].

 P Perceivable	 O Operable	 U Understandable	 R Robust
<i>Information and UI components must be presentable in ways users can perceive — they cannot be invisible to all senses.</i> • Text alternatives for images (alt text) • Captions for video and audio content • Sufficient colour contrast (minimum 4.5:1 for normal text) • Content not conveyed by colour alone <i>WardWatch fail: The risk flag uses colour only — red/green — to show patient status. A colour-blind nurse cannot distinguish them.</i>	<i>UI components and navigation must be operable — users cannot be required to perform interactions they cannot perform.</i> • All functionality available from keyboard (no mouse required) • Users can control timing of content • Navigation aids: skip links, clear page titles, logical focus order • No content that flashes more than 3 times per second <i>WardWatch fail: Brenda uses keyboard shortcuts — but risk-flagging requires a mouse hover to reveal the icon. Keyboard-only users cannot access it.</i>	<i>Information and UI operation must be understandable — content or operation cannot be beyond users' comprehension.</i> • Language of page identified in code • Unusual words or abbreviations explained • Error messages identify the problem and suggest correction • Consistent navigation and labelling across pages <i>WardWatch fail: Error message on failed handover sign-off reads 'ERR_402'. No plain-language explanation. Violates this principle directly.</i>	<i>Content must be robust enough to be interpreted by a wide variety of user agents, including assistive technologies.</i> • Valid HTML — assistive technologies can parse it • ARIA landmarks to identify regions of the page • Name, role, and value provided for all UI components • Compatibility tested with JAWS, NVDA, VoiceOver <i>WardWatch fail: Dynamic patient updates are injected via JavaScript but not announced to screen readers — ARIA live regions are absent.</i>

[4] W3C (2018). Web Content Accessibility Guidelines (WCAG) 2.1. World Wide Web Consortium. Available at: <https://www.w3.org/TR/WCAG21/> [5] Henry, S.L. (2006). Understanding Web Accessibility. In: Thatcher, J. et al. Web Accessibility: Web Standards and Regulatory Compliance. Berkeley: Apress.

Evaluating Accessibility in Practice

Combine automated scanning with manual checks, neither alone is sufficient [5, 6].



Automated Tools

Run first — catch the easy wins

WAVE (WebAIM)

Browser extension — visual overlay showing errors, alerts, and structure

Axe DevTools

Chrome/Firefox DevTools panel provides zero false positives; integrates with CI/CD

Google Lighthouse

Built into Chrome DevTools that does accessibility audit as part of overall quality score

Colour Contrast Analyser

Desktop tool that checks foreground/background contrast ratios against WCAG thresholds

⚠ *Automated tools catch ~30–40% of WCAG failures (Deque, 2020). The rest require human judgement.*



Manual Checks

Always required — automation cannot replace these

Keyboard navigation

Tab through every interactive element — can you reach and operate everything without a mouse?

Screen reader test

Use NVDA (free, Windows) or VoiceOver (built-in, Mac/iOS) — does the interface make sense without sight?

Zoom to 200%

Resize to 200% — does content reflow without horizontal scrolling or content loss?

Cognitive load review

Plain language check, consistent navigation, error identification — requires human judgement

⚠ *Manual testing requires expertise — pair with an accessibility specialist or use a standardised WCAG audit checklist.*

Severity Rating — Consolidating Findings

After individual HE reviews, evaluators consolidate by the combining list that drives the prioritised problem backlog [8].

0 — Not a problem

Cosmetic issue only — fix only if time permits

1 — Cosmetic

Minor usability problem — fix with low priority

2 — Minor

Usability problem — fixing should be given medium priority

3 — Major

Important to fix — high priority before release

4 — Catastrophe

Imperative to fix — prevents task completion

WardWatch — Consolidated issue list (extract)

#	Issue	Heuristic	Found by	Severity	Frequency	Priority
1	Risk flag invisible — no label, icon only	H1, H6	3/3 eval	4	Every session	Critical
2	Audit export takes 40s — no progress indicator	H1	2/3 eval	3	Daily	High
3	'Sign off' button absent on mobile viewport	H7	3/3 eval	3	Every shift	High
4	Error code 'ERR_402' — no plain language description	H9	2/3 eval	2	Occasional	Medium
5	Colour-only risk status (WCAG 1.4.1 fail)	H4, WCAG	1/3 eval	2	Always	Medium

[8] Nielsen, J. (1994). Severity ratings for usability problems. Available at: nngroup.com Priority = Severity × Frequency. Source: Nielsen, J. and Mack, R.L. (eds.) (1994). Usability Inspection Methods. New York: Wiley.

The Impact / Effort Matrix

Turns a prioritised problem list into a realistic design roadmap, balancing user value against development cost [9].

HIGH IMPACT

QUICK WINS

Do now



BIG BETS

Plan carefully

Quick Wins

High user impact, low dev cost. Do these first. Often: copy changes, CSS tweaks, missing labels. Issues #1 (risk flag CSS) and #4 (ERR_402 copy) land here.

Big Bets

High user impact but significant effort. Plan for the next sprint or release. Issue #2 (export progress bar — backend + UI) and #3 (mobile sign-off — layout rework) land here.

Fill-Ins

Low impact but quick. Add to a quiet sprint. Issue #5 (icon alongside colour flag) lands here i.e., low severity, but WCAG required.

Rethink

Low impact and high effort. Challenge whether the feature is needed. If yes, redesign scope before committing.

FILL-INS

Do if time allows

RETHINK

Consider dropping

LOW IMPACT

LOW EFFORT

HIGH EFFORT

Extending the Qualitative Toolkit

Thematic Analysis is not the only qualitative method — each below asks a fundamentally different question [10, 11].



Thematic Analysis

Research question: What themes emerge across this dataset?

Unit: Dataset

Output: Named themes with supporting quotes

Lecture 6 — already covered



Affinity Diagramming

Research question: How can we organise raw data points into design-relevant clusters?

Unit: Data points

Output: Physical/visual cluster map

Collaborative team-based sense-making



IPA

Research question: How does this person make sense of and experience this technology?

Unit: Individual cases

Output: Experiential themes per participant

Emotional/identity dimensions of UX

[10] Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), pp.77–101.

[11] Spencer, D. (2009). *Card Sorting: Designing Usable Categories*. New York: Rosenfeld Media.

What is Affinity Diagramming?

A bottom-up method for organising hundreds of raw data points into a hierarchical map of design insights [12, 13].

Origin: Developed by Kawakita Jiro (the 'KJ method') in the 1960s for organising field data in anthropology. Adopted into software design by Beyer & Holtzblatt (1998) as part of Contextual Design — a practice that is now standard across UX teams at Google, Microsoft, and IDEO.



How it relates to Thematic Analysis

TA codes textual data and derives themes through close reading. Affinity Diagramming works with individual data notes (one insight per sticky note), physically clusters them, and builds a hierarchy from the bottom up. TA is analytical and interpretive; affinity is spatial and collaborative.

TA = analyst reads → codes → names themes
Affinity = team writes notes → clusters → names groups → names clusters



What it produces

A wall (physical or digital) of notes organised into groups and clusters with labels at each level. The top-level cluster labels become the high-level design themes. It produces a shared, visible, team-owned model of the user's world — not a document that lives in one analyst's head.

Typical scale: 200–1,000 notes · 20–100 groups · 5–15 top-level clusters



When to use it

Best used after fieldwork (interviews, observations, diary studies) when you have large volumes of raw qualitative data and need to make sense of them collaboratively as a team. Ideal in the Define phase of the Double Diamond — converting research data into a shared problem framing.

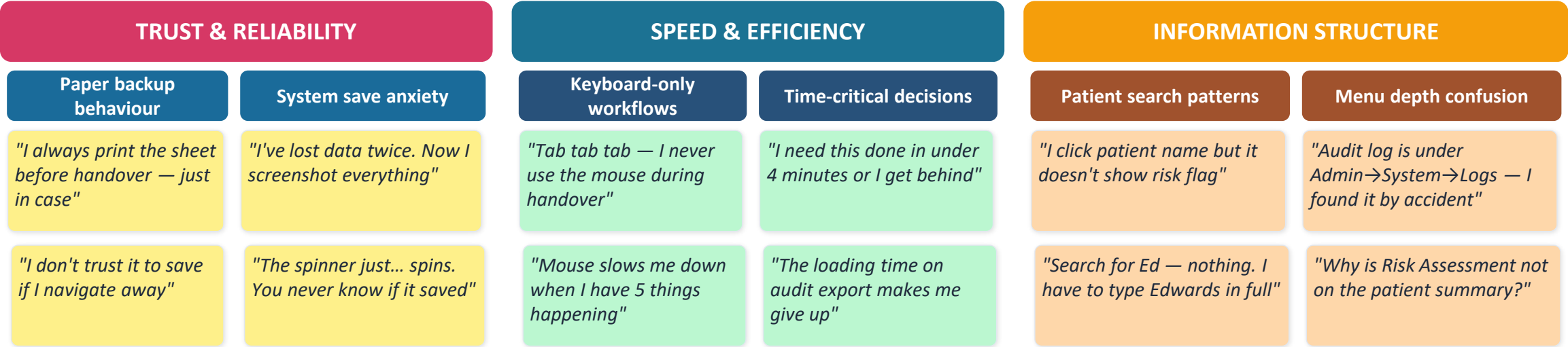
Tools: physical Post-it notes on wall · Miro · FigJam · Mural (digital alternatives)

[12] Beyer, H. and Holtzblatt, K. (1998). *Contextual Design: Defining Customer-Centred Systems*. San Francisco: Morgan Kaufmann.

[13] Holtzblatt, K., Wendell, J.B. and Wood, S. (2005). *Rapid Contextual Design*. San Francisco: Morgan Kaufmann.

Worked Example — WardWatch Affinity Diagram

Built from 5 nurse interviews and 2 observation sessions. Each note = one verbatim observation or quote [12].



What the affinity diagram tells the design team:

TRUST & RELIABILITY	Auto-save must be visible and real-time. Every save event needs a confirmation indicator. Loss of data = loss of patient safety.
SPEED & EFFICIENCY	The entire handover flow must be keyboard-navigable. Mouse interactions are a design failure for this user group under time pressure.
INFORMATION STRUCTURE	Risk flag and patient summary must be co-located. Search must support partial matching. Audit logs belong in a dedicated 'Reports' section.

[12] Beyer & Holtzblatt (1998). Contextual Design. Morgan Kaufmann. Example data is illustrative — method and structure are authentic. Affinity hierarchy: pink = top-level clusters, blue = groups, yellow = data notes.

Interpretative Phenomenological Analysis (IPA)

A qualitative method for understanding how individuals make sense of significant personal experiences [14, 15].

Origin: Developed by Jonathan Smith (1996) in health psychology. Rooted in phenomenology (Husserl, Heidegger), hermeneutics (Gadamer), and idiography (Allport). Applied to HCI and UX to understand the experiential and emotional dimensions of living with technology — dimensions that usability metrics alone cannot capture.



Phenomenological

IPA is concerned with experience as it is lived — the texture of everyday life with a system, not just task performance. It asks: what is it like to be this person using this technology in this context? The focus is on meaning, not frequency.

vs TA: Unlike TA (what themes emerge across participants?), IPA asks: how does this individual make sense of their experience?



Interpretative

The 'double hermeneutic' — the researcher interprets the participant's interpretations. The analyst is not a neutral recorder but an active sense-maker. This interpretive layer is acknowledged, not hidden. The analyst's positioning must be stated.

vs TA: Unlike content analysis (counting occurrences), IPA makes claims about meaning from small samples. One participant's account can constitute a valid finding.



Idiographic

IPA analyses individual cases in depth before looking across cases. Each participant's account is treated as a world of its own. Cross-case patterns emerge only after individual analysis — not imposed top-down from the start.

vs TA: Unlike TA (any N), IPA uses 3–8 participants. A study of one participant is methodologically legitimate. Depth over breadth.

[14] Smith, J.A., Flowers, P. and Larkin, M. (2009). *Interpretative Phenomenological Analysis: Theory, Method and Research*. London: Sage.

[15] Smith, J.A. (1996). *Beyond the divide between cognition and discourse: Using interpretative phenomenological analysis in health psychology*. *Psychology and Health*, 11(2), pp.261–271.

IPA — The Six-Step Process

Applied case by case — only after each individual analysis is complete do you look across cases [16].

1 Read and re-read
Immerse in the transcript. Read multiple times without coding. Note initial impressions, emotions, surprises. The goal is familiarity — not yet analysis.

At this stage: annotate freely in the margins. No coding categories. Just responses.

2 Initial noting
Annotate the transcript freely into 3 groups:

- descriptive (what is happening?)
- linguistic (how is it being said?)
- conceptual (what does this imply?)

Three levels of comment: descriptive, linguistic ('notice the hesitation here'), conceptual ('this suggests loss of professional identity').

3 Develop emergent themes
Working from your annotations, create concise theme statements. Each theme captures the psychological essence of a cluster of notes. Themes are phrases, not codes: not 'trust' but 'eroding confidence in a system that should be trustworthy'.

Themes are distinct from TA themes — they are experiential, not topical.

4 Search for connections across themes
Look for relationships between themes: which themes cluster? Which contradict? Which are superordinate? Draw a map if helpful. This produces a hierarchical structure of experiential themes for this participant.

A theme may absorb another. Some may be minor. Prioritise depth over exhaustiveness.

5 Move to the next case
Repeat steps 1–4 for each participant independently. Resist the urge to let the first case dominate. Each case must be allowed to speak for itself before cross-case patterns are imposed.

Each participant gets their own theme structure. Only then do you look across them.

6 Look for patterns across cases
Identify themes that recur, vary, or are unique across participants. Produce a master theme table. Note convergence (shared themes) and divergence (different experiences of the same situation).

Report both convergence and divergence — IPA is not looking for consensus but for the range of experience.

IPA — Worked Example

Participant: Brenda, Senior Ward Nurse, 14 years NHS experience. Interview topic: her daily experience of WardWatch [8].

Transcript extract with initial noting

“I always end up going back to my paper sheet.
I know I shouldn't, but... it's mine. I trust it.”

Descriptive: paper backup behaviour
Linguistic: 'mine', ownership language
Conceptual: digital system excludes her from ownership of her own work

“When the system loses data I feel like I've let the patient down. That's not the system's fault. That's on me.”

Descriptive: self-blame for system failure
Linguistic: transfers system agency to self
Conceptual: professional identity threatened when system fails

“I taught myself the shortcuts because otherwise I can't do my job properly. No one showed me.”

Descriptive: self-taught coping strategy
Linguistic: 'my job' — professional pride
Conceptual: competence constructed despite the system, not through it

From initial notes → emergent themes

Emergent theme 1:

Paper as a trusted extension of professional self — the system is external, the paper is internal

Emergent theme 2:

Absorbing system failure as personal moral failure — technology's errors become her professional inadequacy

Emergent theme 3:

Competence constructed through self-taught workarounds — expertise in spite of, not because of, the design

Master theme (this participant):

Master theme: WardWatch undermines professional identity — it positions Brenda as a user of someone else's system rather than an expert practitioner in control of her own work.



IPA finding: The usability problem (invisible risk flag) is not just a navigation issue, it is experienced as a threat to professional identity. A redesign that returns agency to the nurse (e.g. customisable dashboard, visible confirmation of actions) may be more effective than a purely functional fix.

[8] Smith, J.A., Flowers, P. and Larkin, M. (2009). *Interpretative Phenomenological Analysis*. Sage. Data is illustrative; method and analytical framework are authentic. Three-level annotation: descriptive, linguistic, conceptual (after Smith et al., 2009, Ch.4).

Thank you for your attention!

Questions?

Feedback



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

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

Menti Code: **5647 6463**