Participant UX01

Field	Details
Participant ID	UX01
Age / Background	20 – Bachelor of IT, first year
Experience Level	Beginner with heuristic evaluation
Test Date	22 Oct 2025
Device / Browser Used	Windows laptop, Chrome

Task Summary

Task	Description	Ethical Reflection Focus
1	Run ExplainUX using 3 heuristics	Transparency of AI process
2	Re-run after clarifier answers	Fairness and user control
3	Download CSV results	Comfort with automated judgement

Key Observations

- **Task 1:**
  - *User Comment:* "It says 'Running Clarifier → Scorer → Ethics' , what's ethics doing here?"
  - *Ethical Reflection:* Wondered whether "Ethics" meant it would judge moral aspects of design. Said "I didn't expect that from a UX tool."
  - *Observer Note:* Looked curious but slightly uneasy until results appeared.

- **Task 2:**
  - *User Comment:* "Oh, it asks follow-ups... does it really change what it thinks?"
  - *Ethical Reflection:* Noted, "It's nice it asks me questions — feels fairer."
  - *Observer Note:* Increased trust once responses updated; said "Now it listens to me."

- **Task 3:**
  - *User Comment:* "Cool, I can download it — but I'd still check the answers myself."
  - *Ethical Reflection:* Expressed slight hesitation about relying on AI-generated fairness comments: "I'm not sure if it knows what's actually inclusive."

Behavioural / Ethical Indicators

Category	Evidence
Confidence	Hesitant at first, relaxed after seeing feedback
Ethical Awareness	Questioned fairness and autonomy
Trust in AI	Moderate, improved after re-run

Category	Evidence
Agency	Felt empowered when clarifier answers changed output
Transparency Perception	Moderate, understood step names but not internal reasoning

Post-Task Reflection

- **Ethical Comfort:** 3 / 5 – "I trust it more after I answered questions."

- **Perceived Fairness:** 3 / 5 – "Not sure if it treats all designs the same."

- **Transparency Comment:** "I could tell it was doing something ethical, but not how."

- **Agency Comment:** "I liked that I could change its mind."

Summary

Novice participant gained trust through interaction; ethical curiosity replaced initial unease. Needs clearer onboarding explaining the *Ethics* phase and fairness purpose.

Participant UX02 – Intermediate (Experienced Student)

Field	Details
Participant ID	UX02
Age / Background	22 – B.CompSci (UX minor)
Experience Level	Moderate, familiar with heuristics
Test Date	22 Oct 2025
Device / Browser Used	Laptop, Opera

Task Summary

Task	Description	Ethical Reflection Focus
1	Run evaluation on "MyUQ portal"	Clarity of AI reasoning
2	Re-run with clarifier answers	Trust and fairness awareness
3	Download results	Ethical reflection usefulness

Key Observations

- **Task 1:**

  - *User Comment:* "It's great that the system is transparent about what it's doing — clarifier, scorer, ethics."

  - *Ethical Reflection:* "The ethical part is interesting — it makes me think about inclusivity automatically."

- **Task 2:**

- o *User Comment:* "The follow-ups make it feel collaborative."

- o *Ethical Reflection:* Recognised co-agency — "It listens and adapts; that's how AI should behave."

- o *Observer Note:* Spoke about fairness explicitly; said "It doesn't assume all users are the same."

- **Task 3:**

  - o *User Comment:* "The ethical reflection column is my favourite — reminds me to justify my scores."

  - o *Ethical Reflection:* Viewed as *educational fairness cue* rather than moral policing.

Behavioural / Ethical Indicators

Category	Evidence
Confidence	High; used tool methodically
Ethical Awareness	Strong; linked reflections to inclusive design
Trust in AI	High; considered it a supportive reviewer
Agency	Felt co-author of evaluation
Transparency Perception	Clear; valued explicit step naming

Post-Task Reflection

- **Ethical Comfort:** 4 / 5 – "Feels responsible, not preachy."

- **Perceived Fairness:** 4 / 5 – "Balanced comments across heuristics."

- **Transparency Comment:** "I could see how it processed my input step by step."

- **Agency Comment:** "It respected my context."

Summary

Intermediate user experienced ExplainUX as a *co-evaluative partner*. Ethics component enhanced reflection and fairness thinking, supporting HCAI goals of transparency and accountability.

Participant UX03

Field	Details
Participant ID	UX03
Age / Background	26 – Master of Interaction Design (DECO7821)
Experience Level	Advanced
Test Date	25 Oct 2025
Device / Browser Used	Desktop PC. Chrome

Task Summary

Task	Description	Ethical Reflection Focus
1	Evaluate Blackboard submission page	Explainability and accountability
2	Re-run with clarifier answers	Fairness reasoning depth
3	Download CSV	Transparency and audit potential

Key Observations

- **Task 1:**
  - *User Comment:* "Good that it discloses each step — transparency is built-in."
  - *Ethical Reflection:* "Would be stronger if it cited sources or model confidence rationale."

- **Task 2:**
  - *User Comment:* "These clarifiers simulate participatory co-design."
  - *Ethical Reflection:* "Ethical reflections are surface-level — need more traceability of reasoning."
  - *Observer Note:* Discussed algorithmic bias and interpretability; proposed expert-mode features.

- **Task 3:**
  - *User Comment:* "CSV output enables auditability — perfect for accountability."
  - *Ethical Reflection:* "That's transparency in action — export lets others verify results."

Behavioural / Ethical Indicators

Category	Evidence
Confidence	Very high
Ethical Awareness	Deep; referenced fairness frameworks
Trust in AI	Conditional — trusted process, questioned logic opacity
Agency	Felt empowered; wanted reasoning visibility
Transparency Perception	Strong, requested "show reasoning trace"

Post-Task Reflection

- **Ethical Comfort:** 5 / 5 – "I'd use this in research with full disclosure."
- **Perceived Fairness:** 4 / 5 – "Seems impartial but opaque."
- **Transparency Comment:** "Needs optional detailed reasoning trace for expert review."
- **Agency Comment:** "AI acts as assistant, not authority — that's right."

Summary

Expert user praised ExplainUX for **accountability and transparency**, highlighting potential for research-grade auditability. Requested deeper explainability layers to support HCAI governance use cases.

SUS Questionnaire Items

No.	Statement	Type
1	I think that I would like to use this system frequently.	Positive
2	I found the system unnecessarily complex.	Negative
3	I thought the system was easy to use.	Positive
4	I think that I would need the support of a technical person to be able to use this system.	Negative
5	I found the various functions in this system were well integrated.	Positive
6	I thought there was too much inconsistency in this system.	Negative
7	I would imagine that most people would learn to use this system very quickly.	Positive
8	I found the system very cumbersome to use.	Negative
9	I felt very confident using the system.	Positive
10	I needed to learn a lot of things before I could get going with this system.	Negative

Participant Responses

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score (/100)	Interpretation
UX01 (Novice)	3	3	3	2	4	2	4	2	3	2	**75.0**	Good usability
UX02 (Intermediate)	4	2	4	2	4	2	4	2	4	2	**85.0**	Excellent usability
UX03 (Advanced)	5	1	5	1	5	1	5	1	5	1	**95.0**	Best imaginable usability

Average SUS Score

$$\text{Mean SUS} = (75 + 85 + 95)/3 = 85.0$$

Overall Usability Rating: Excellent (Grade A)

This score falls within the **85th percentile**, indicating that ExplainUX is perceived as highly usable, efficient, and well integrated by all participants.

Interpretation of Results

- **UX01 (Novice):** Found the system clear once familiar but initially uncertain about terminology and re-run flow.

- **UX02 (Intermediate):** Reported strong transparency and efficiency; minor delay during re-run slightly reduced score.

- **UX03 (Advanced):** Rated the system extremely high for usability; suggested optional advanced features for expert analysis.

- **Overall Trend:** Usability perception **increased with experience**, demonstrating strong *learnability* and *adaptability* across user levels.

Benchmark Reference (Bangor et al., 2009)

SUS Range	Adjective Rating	Grade	Percentile
0–50	Poor	F	<15th
51–68	OK	D–C	35th
68–80	Good	B	70th
80–90	Excellent	A	85th
90–100	Best Imaginable	A+	>95th