FigmaXR — IP1 Evaluation

1) Objective & validation metrics

Aim:

Determine whether the IP1 prototype is usable enough to let participants judge UI visibility in context (i.e., they can learn controls quickly, receive clear feedback, make few errors, and progress with minimal prompting).

Success criteria:

- IUQ-12 total mean ≥ 4.0/5; each subscale A/B/C ≥ 3.5.
- ≥80% participants complete the core flow without heavy prompting (target: ≤1 prompt median). Measures. IUQ-12 (12 items, 5-point Likert), task completion (Y/N), prompts count, notable errors & quotes.

2) Results

2.1 Session snapshot

- Participants: n = 5 (XR exp: L=1 / M=3 / H=1)
- Completion: 100% (5/5)
- Prompts (median): 4
- IUQ-12 means (1–5): Total 3.63, A 3.96 (Learnability/Controls), B 3.13 (Feedback/Visibility), C 3.60 (Efficiency/Errors/Comfort)
- Participants at/above bar (Total ≥4.0): 2/5

2.2 Item-level means

	Item (short)	Mean
L1	Learn basics in 30s	4.2
L2	Move/look predictable	4.4
L3	Pickup/put-down intuitive	3.4
L4	Env switch easy to find/use	4.0
L5	Lighting slider easy	3.8
L6	System feedback clear	4.0
L7	Screen legible	2.4
L8	Rare hesitation (findability/prompts)	3.0
L9	Flow smooth	3.4
L10	Few errors/few prompts	3.2
L11	Comfort (no motion sickness)	4.4
L12	Confidence to use again	3.4

Optional semantic differentials (1=good end): Clarity 2.8, Contrast 3.2, Comfort 3.0.

2.3 Short-interview synthesis

Top themes (n=5):

Pickup/put-down aiming & feedback — *2/5*

"hard to aim items to choose and put down", "Take tablet"

Screen legibility tuning (contrast / size / lighting) — *3/5*

"enhance contrast", "adjust screens size", "Hard to control light"

Control discoverability / complexity — *2/5*

"change room could be more clear", "so many buttons / mouse movement"

Device placement behaviour — *1/5*

"Devices can be placed freely instead of returning to their original positions"

3) Analysis / insights

- Learnable but not self-explanatory. Subscale A ≈ 4.0 (I1/I2 high) shows low learning barrier, yet prompts median = 4 and I10=3.2 indicate self-explanation is weak: people know what to do but often need help with how/where.
- Legibility is the primary blocker. I7 = 2.4 (lowest item) and optional SD scores (clarity/contrast) point to on-device screen readability issues—likely a mix of texture resolution/filters, physical scale, and default lighting.
- Feedback & visibility cues are underpowered. B = 3.13 with I8=3.0 suggests control salience and action feedback (pickup/put-down/switch/adjust) aren't strong enough to keep users flowing independently.
- Comfort is fine. I11=4.4 indicates camera/motion tuning is acceptable; no systemic motion-sickness concerns.

4) Evaluation of aims

- Met: Completion 100% ≥ 80%.
- Not met: Total 3.63 < 4.0; B 3.13 < 3.5; prompts median = 4 (not "minimal").
- Conclusion: IP1 is not yet "usable enough to test UI" per the bar; the gating factors are screen legibility and self-explanatory guidance.

5) Concept iteration

I1 — Legibility bundle (highest priority)

- Export Figma frames at 2×/3×, disable aggressive texture compression; enable MipMaps + Trilinear + Anisotropic ≥4.
- Increase device-screen physical scale +10–15%; add optional 1.2× inspect (e.g., hold RMB/F).
- Target: I7 ≥ 3.8, Subscale B ≥ 3.5, Total ≥ 4.0.

I2 — Safer default lighting

- Start at a readable preset, clamp extremes, add "Reset to Readable" button.
- Target: +0.6 uplift on I7, smoother C.

I3 — Self-explanatory cues & control salience

- Add micro-prompts after 3–5s idle ("WASD / Click to pick up / Switch / Adjust lighting").
- Make Environment Switch unmistakable (label + icon + soft glow; position nearer center FOV).
- Target: prompts median ≤ 1, I8 ≥ 3.8, I10 ≥ 3.8.

I4 — Pickup/put-down aiming & feedback

- Enlarge pickup collider +20–30%; add reticle snap when hovering.
- On pickup/put-down, play outline + click SFX and show a brief tooltip.
- Target: I3 ≥ 3.8, observed mis-aim decreases.

I5 — Device placement option

- Provide a "Free place" toggle (don't auto-return), default can remain auto-return.
- Target: zero mentions of placement frustration in interviews.

6) Reflection & next test

Using scripted tasks and the IUQ-12, I observed 100% completion from five participants, a median of four prompts, and an IUQ-12 total mean of 3.63 with subscales 3.96 for Learnability and Controls, 3.13 for Feedback and Visibility, and 3.60 for Efficiency, Errors and Comfort. Short

interviews reinforced this: participants struggled to aim when picking up or putting down devices, on-device legibility was weak across contrast, size and lighting, controls felt unclear or busy, and one person wanted free device placement. For IP2 I will raise legibility with higher-resolution exports, improved texture filtering, a slightly larger device screen, and a readable default lighting with a quick reset; reduce hesitation with gentle micro-prompts, a clearer environment switch, stronger pickup feedback and larger colliders; consider a free-place option; and tighten my method with a standardized think-aloud reminder, time-boxed lighting adjustment, and a single immediate legibility rating plus an optional two-alternative check.

Appendix

See ***EvaluationResult1_Rawdata.pdf*** for the complete dataset.

IP1 Usability Questionnaire (IUQ-12, 5-point)

Participant ID (PID): _____ XR Experience: □ L □ M □ H Date: _____

Instructions: Please rate each statement based on your experience with the prototype.

Scale: 1 = Strongly Disagree · 2 = Disagree · 3 = Neutral · 4 = Agree · 5 = Strongly Agree

A. Learnability & Controls

1.I could learn the basic controls within **30 seconds** (WASD to move, mouse to look, click to pick up/use).

□1 □2 □3 □4 □5

2.Movement and camera look felt **predictable** and easy to control.

□1 □2 □3 □4 □5

3.**Picking up / putting down a device** was intuitive (I knew when it was picked up/put down).

□1 □2 □3 □4 □5

4.The **environment switch** was easy to find and use.

□1 □2 □3 □4 □5

5.The **lighting slider** was easy to understand and use (I knew how it affects readability).

□1 □2 □3 □4 □5

B. Feedback & Visibility

6.When I acted (pick up / switch / adjust lighting), the **system feedback** was clear.

□1 □2 □3 □4 □5

7.The **Figma interface** on the device screen was **clear and legible** (text, icons, dividers).

□1 □2 □3 □4 □5

8.I **rarely hesitated** because controls were hard to find or prompts were unclear.

□1 □2 □3 □4 □5

C. Efficiency, Errors & Comfort

9.I could **complete the tasks smoothly** (view UI → switch environment → adjust lighting) without major friction.

□1 □2 □3 □4 □5

10.I rarely mis-clicked or made errors, and rarely needed moderator prompts.

□1 □2 □3 □4 □5

11.I did not feel motion-sick or uncomfortable; overall comfort was good.

□1 □2 □3 □4 □5

12.I feel **confident using** this prototype again; overall **usability is good**.

□1 □2 □3 □4 □5

Optional: Semantic Differentials (pick 1–3 lines if desired)

Clear □1 □2 □3 □4 □5 □6 □7 Blurry

High contrast □1 □2 □3 □4 □5 □6 □7 Low contrast

Comfortable □1 □2 □3 □4 □5 □6 □7 Straining

Short Interview (1–2 sentences each)
Q1. What was the **biggest difficulty** you encountered?
Q2. Which **control or feedback** most needs improvement?
Q3. If you could **change one thing first**, what would it be?
Q4. Any other comments or suggestions?
Moderator notes (not shown to participant):
Prompts needed: __ times Task completed: □ Yes □ No Notes/Quotes: _____

PID,XR_Familiarity(L/M/H),Task_Completed(Y/N),
Prompts_Count,Item1,Item2,Item3,Item4,Item5,Item6,Item7,Item8,Item9,Item10,Item11,Item12,Optional1,Optional2,Optional3
P01,M,Y,4,4,5,2,4,3,4,2,3,2,3,3,4,4,3,3
P02,L,Y,3,2,3,2,4,4,4,1,2,2,2,4,2,6,6,4
P03,M,Y,4,5,5,4,5,3,3,4,4,5,4,5,4,1,1,3
P04,M,Y,6,5,4,5,3,5,5,3,3,3,4,5,4,2,2,3
P05,H,Y,3,5,5,4,4,4,4,2,3,5,3,5,3,1,4,2

I used AI in a limited, documented way: ChatGPT (GPT-5 Thinking) helped me draft and refine the testing plan structure and the moderator script. Claude provided small Unity C# suggestions (e.g., raycast pickup logic, binding a lighting slider, simple environment-switch state handling, and an optional audio-fade coroutine). I treated all AI outputs as drafts or references—reviewed, modified, and verified by me—and I did not rely on AI for autonomous code or asset generation.