1. Objective and Validation Metrics

1.1 Testing Objective

Following the outcomes of IP1 (early concept validation) and the technical failure encountered in IP2 (multi-user networked prototype), I deliberately refocused the testing objectives for IP3. Results from IP1 indicated that although the MR (host) / VR (guest) hybrid meeting concept was clearly communicated, two major concerns emerged: missing functionality and potentially unnatural communication experience. In IP2, I attempted to address these concerns by introducing shared collaboration features such as a virtual whiteboard and avatars. However, due to unexpected issues with Netcode networking and MR rendering errors, the prototype completely failed during testing, and I was unable to collect any meaningful design feedback.

As a result, the core objective of IP3 was to rigorously evaluate the interaction fidelity and naturalness of the MR collaborative tool prototype. Learning from the setbacks of IP2, I intentionally removed complex multi-user networking features and instead focused solely on the single-user interaction experience. This decision ensured that before reintroducing social communication components, the fundamental interaction layer of the system would be validated as usable, stable, and natural.

The specific testing objectives are as follows:

Usability: Evaluate whether participants can understand and operate the toolbar intuitively, including grabbing virtual tools (marker, ruler, compass) and performing actions on the whiteboard (drawing, erasing, clearing).

Naturalness of Interaction (Primary Focus): Assess whether hand-tracking interactions with virtual tools—particularly the marker—feel natural and stable, without noticeable delay, misalignment, or jitter.

Immersion and Spatial Integration: Evaluate whether virtual objects (whiteboard and tools) appear naturally integrated into the user's physical environment through Passthrough, and whether this contributes to a convincing sense of being “in a meeting setting”.

1.2 Validation Metrics

To measure whether the above objectives were achieved, the following evaluation criteria were established:

- Success Criteria: The average score for key questions in the 5-point Likert scale questionnaire must reach 4.0 or higher.
- Failure Criteria: Average scores below 3.5 on key questions, or repeated negative themes in qualitative feedback.

2. Results

This section objectively presents all data collected during the IP3 testing phase.
A total of five participants completed both the Think-Aloud usability session and a post-test questionnaire using the Meta Quest 3 headset.

This section objectively presents the key summarized findings from the IP3 testing phase. The complete raw dataset, including full participant questionnaire responses and interview transcripts, is available in **Appendix A** for detailed review.

2.1 Participant Background

All five participants reported that they often use XR devices and are familiar with using whiteboards and pens.
This indicates that potential biases caused by unfamiliarity with the technology were minimized, allowing the data to more accurately reflect the usability of the prototype itself.

3. Analysis and Insights

3.1 Insight 1: High Usability vs. Low Interaction Fidelity

The data reveals a stark disconnect between "Usability" (what to do) and "Fidelity" (how it feels). Participants found the UI and tool selection highly intuitive, awarding "Was it easy to understand how to interact?" an average score of 4.8/5.0. This is strongly supported by qualitative feedback such as "The whiteboard was easy to use" (Sean) and "Tools were easy to understand" (Leon).

However, this high usability was completely undermined by poor interaction fidelity. "How smooth was using the tools?" received a unanimously low average score of **3.0/5.0**. The qualitative data precisely identifies the cause: a "clunky" or "unnatural" feeling stemming from technical flaws.

3.2 Insight 2: "Clunkiness" as the Core Blocker to "Naturalness"

The core project goal of creating a "natural" meeting experience (Target: 4.0) was not met, achieving an average score of only 3.6/5.0. My analysis identifies that this failure is a direct consequence of the low interaction fidelity mentioned in Insight 1.
The qualitative feedback provides a detailed diagnostic of this "clunkiness":

- Misalignment: The most critical issue, cited by User 1（Jialiang He） ("stroke and pen tip were not aligned") and User 5（Peilin Li） ("line floated slightly above the board").
- Latency: User 2（Sean） noted the "pen felt laggy when drawing fast".
- Lack of Spatial Feedback: User 3（Sijian Xin） felt the ruler was "floating" and lacked a "shadow or guide" to ground it in the virtual space.

This demonstrates that in MR, "naturalness" is not just conceptual; it is a direct result of high-fidelity, low-latency spatial interaction.

3.3 Insight 3: Technical Bugs Masquerade as Design Flaws

A significant finding was the catastrophic failure of the "Did you understand the purpose of each tool?" metric, which scored 2.6/5.0. At first glance, this suggests a major design flaw in my tool icons or concepts.

However, the qualitative data reveals this is purely a technical failure. The compass tool was broken. As User 1（Jialiang He） and User 4（Leon） stated, "the compass didn't draw a circle" or "didn't work". Participants were confused about its purpose *because it was non-functional*. This teaches a critical lesson: in prototype testing, users cannot evaluate the *design* of a feature if its *implementation* is broken. The resulting data is misleading, reflecting technical debt, not design insight.

3.4 Insight 4: Core Concept Validated, Revealing a Clear Future Roadmap

Despite the interaction flaws, the core concept of an MR collaboration tool was validated. User 3（Sijian Xin） called it a "Really cool idea", and participants clearly identified its value in "Online Teaching" and "Online Meetings".

Crucially, the feedback from users naturally progressed from evaluating *current* features to requesting *future* ones. Suggestions like "laser pointer," "importing PDF slides" (User 5 Peilin Li), and "change pen thickness or color faster" (User 2 Sean) demonstrate that users were engaged enough to envision its real-world application. This provides a clear, user-driven roadmap for future iterations.

4. Reflection

4.1 Prototype Session Review (IP3)

Per the guidelines, I have selected my IP3 testing session for this review.

- **What Worked:** The session successfully validated the UI/UX design. The "ease of understanding interaction" (4.8/5.0) was the highest-rated metric, supported by qualitative feedback (User 4: "Tools were easy to understand"). This confirms my tool-selection interface is intuitive.
- **What Didn't Work:** The session failed to validate the core goal of a "natural" experience. The "smoothness" (3.0/5.0), "immersion" (3.0/5.0), and "naturalness" (3.6/5.0) metrics all failed to meet my 4.0 success criteria.
- **Critically Explain Why:** The failure was not one of *design concept* but of *technical execution*. As analyzed in Section 3, the root cause was low interaction fidelity, or "clunkiness." Specific, data-linked causes include:
  1) Misalignment between the pen tip and the drawn line (User 1, 5 Jialiang He，Peilin Li),
  2) Latency during fast drawing (User 2 Sean),
  3) Functional Bugs (the compass) which broke the experience (User 1, 4 Jialiang He，Leon).

4.2 Methodological Reflection

Looking back at the entire course, my methodology evolved significantly:

1. IP1 (Concept Validation): I used low-fidelity interviews. This was highly effective, as it correctly identified the two core challenges that would define my entire project: "function deficiency" and the potential for "unnatural communication" in my MR/VR split-concept.
2. IP2 (Technical Failure): I attempted a high-fidelity, multi-user, live demo. This methodology was completely ineffective and improper. I improperly conflated *design validation* (is the whiteboard useful?) with *technical validation* (does my Netcode work in this environment?). The technology failed, and as a result, I gathered zero design insights. The key lesson learned was the necessity of backup plans (like a video) and decoupling design from unstable technology.

4.3 Concept Evaluation

My original XR concept was a hybrid MR/VR meeting system intended to feel "natural". Based on my testing, this concept is Partially Confirmed.

- Confirmed: The core idea of an MR-based collaboration tool is valuable (User 3: "Really cool idea"). The UI/UX approach I designed is also validated (4.8/5.0 usability).
- Invalidated: My initial assumption that "providing tools equals naturalness" was invalidated. IP3 proved that *interaction fidelity* is the prerequisite for "naturalness". A "clunky" tool (3.0/5.0 smoothness) creates an *unnatural* experience, regardless of the concept's value.
- Uncertain: The most innovative part of my original concept—the split MR/VR environment—remains unvalidated. IP1 warned it might be "unnatural", and IP2's technical failure meant I never gathered data to confirm or deny this core hypothesis.

4.4 Improvements and Extensions

Based on the cumulative insights from IP1, IP2, and IP3, I propose the following changes:

- P0 Fixes: Address the fidelity blockers from IP3. This includes:
    1. Fixing the pen-tip misalignment (User 1, 5), possibly with a calibration feature
    2. Optimizing the drawing code to reduce latency (User 2).
    3. Fixing the compass bug (User 1, 4).

- P1 Affirmation: Implement the features requested by users in IP3 that affirm the concept's direction. This includes: "laser pointer" (User 4), "faster pen controls" (User 2), and "importing PDF slides" (User 4).

- P2 Concept Extensio: As required by the guidelines, my final proposal is to pivot the *core concept* based on my findings. Instead of pursuing the complex and risky MR/VR split, I would explore a unified MR (Passthrough) experience. This would have all users (remote or co-located) appear as avatars within each other's physical (or virtually scanned) spaces. This approach would simplify the technical challenge, enhance co-presence, and directly solve the "unnatural communication" risk identified in IP1.

Appendix

Appendix A

Key Validation Questions

Aspect	Key Evaluation Question
Naturalness	How smooth was the interaction when using tools such as the marker, ruler, and compass?
Immersion	Did the meeting in MR feel natural?
Spatial Integration	Did the real and virtual elements feel well integrated?
Usability	Was it easy to understand how to interact with the tools?

Appendix B

Questionnaire

What XR devices have you used before?

☐ Never ☐ Rarely ☐ Often

Are you familiar with using whiteboards or pens?

☐ Yes ☐ No

Overall Experience

Did you find holding a meeting in this mixed reality environment natural?

1 2 3 4 5 (5 is best)

Did you feel that the real and virtual elements were well integrated?

1 2 3 4 5 (5 is best)

Was it easy to understand how to interact with the tools and interface?

1 2 3 4 5 (5 is best)

Tool Interaction

How smooth did you find using the tools (pen, ruler, compass)?

1 2 3 4 5 (5 is best)

Did you understand the purpose of each tool in the inventory?

1 2 3 4 5 (5 is best)

In what situation would you use this system? (Multiple choices allowed)

☐ online meeting ☐ online teaching ☐ team working ☐ design show ☐ other: _____

Spatial & Immersion Experience

Did the virtual objects (whiteboard, pen) appear naturally placed in space?

1 2 3 4 5 (5 is best)

Did you feel a sense of "being in the meeting environment"?

1 2 3 4 5 (5 is best)

Open Feedback

What was the biggest confusion or difficulty you experienced?

What additional features would you like to see in future versions?

References (APA 7th)

Bachmann, D., Weichert, F., & Ruhl, K. (2019). *Evaluation of the Leap Motion Controller as a new contact-free pointing device for medical applications*. Sensors, 19(2), 331.

Piumsomboon, T., Lee, G., Dey, A., & Billinghurst, M. (2017). *Co-presence and shared*

Appendix C

Quantitative Data

Category	Question	Scores (5 participants)	Average	Evaluation (≥ 4.0 = Pass)
Overall Experience	Did the MR meeting feel natural?	4, 4, 4, 3, 3	3.6	Below Target
	Were real and virtual elements well integrated?	4, 4, 3, 3, 3	3.4	Below Target

	Was it easy to understand how to interact with the tools?	5, 5, 5, 5, 4	4.8	Highly Successful
Tool Interaction	How smooth was using the tools (pen, ruler, compass)?	3, 3, 3, 3, 3	3.0	Failed
	Did you understand the purpose of each tool?	3, 3, 3, 2, 2	2.6	Failed (due to compass bug)
Spatial & Immersion	Did the virtual objects appear naturally placed in space?	5, 4, 3, 3, 2	3.4	Below Target
	Did you feel a sense of "being in the meeting environment"?	3, 3, 3, 3, 3	3.0	Failed

Usage Scenario Preferences:

- Online Teaching – 2 votes
- Online Meeting – 2 votes
- Design Presentation – 1 vote

Appendix D

Qualitative Feedback

Participant	Feedback Summary	Key Observations / Issues Mentioned	Suggested Improvements
Jialiang He	The stroke drawn on the whiteboard was not aligned with the pen tip, and the line appeared twisted. The compass did not function as expected.	Pen–stroke misalignment Compass malfunction	Fix compass drawing logic - Improve stroke precision and surface alignment
Sean	The system worked well overall, but the pen felt slightly	Pen latency Lack of quick customization	Allow faster color/thickness switching

	delayed when drawing quickly. The whiteboard was easy to use.		Optimize rendering delay
Sijian Xin	The concept was interesting and useful for online classes or team meetings. However, the ruler appeared floating in space without a clear point of contact.	Floating ruler effect Lack of visual cues	Add visual guides or shadows for alignment
Leon	The tools were intuitive, but the compass failed to work. Suggested adding a laser pointer or PDF import to support more realistic meetings.	Compass bug Missing presentation features	Add laser pointer tool Enable PDF/slides import
Peilin Li	The drawing experience was enjoyable and natural, but sometimes the line appeared slightly above the whiteboard surface.	Line floating above board Minor spatial offset	Add calibration or surface adjustment option

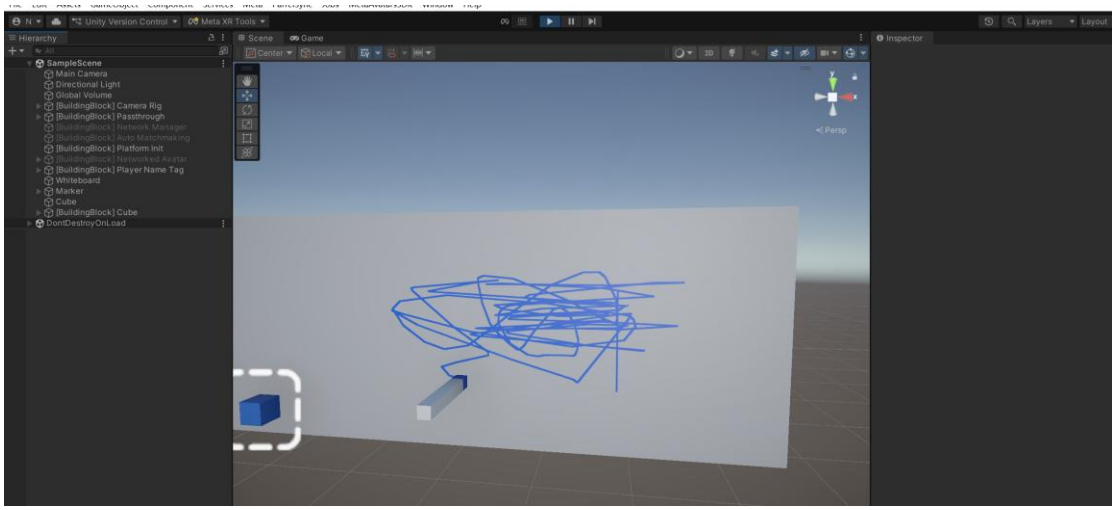
Appendix E

Screenshots of IP1, IP2, and IP3

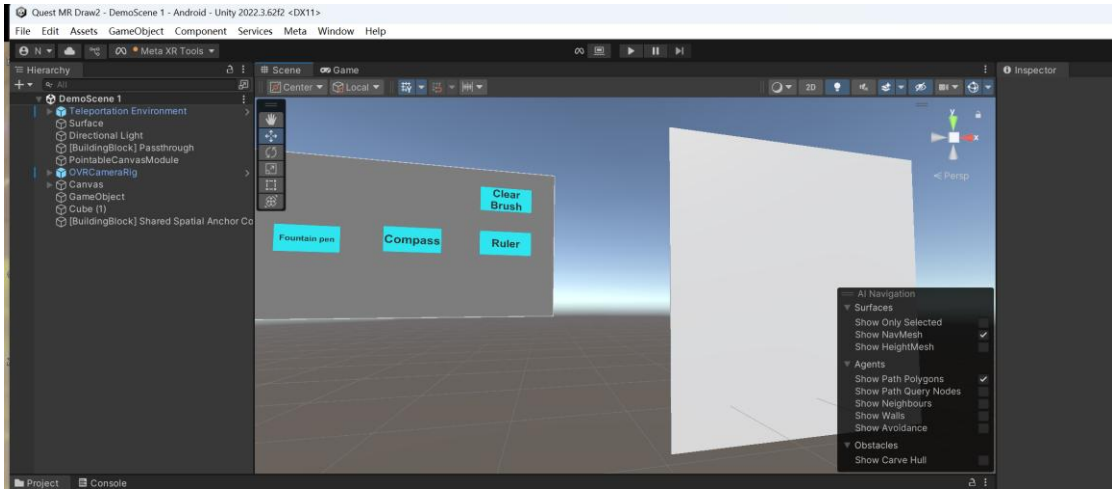
IP1:


Hierarchy
Scene
Game
Inspector
P121*
Main Camera
Directional Light
Global Volume
[BuildingBlock] Network Manager
[BuildingBlock] Auto Matchmaking
[BuildingBlock] Camera Rig
[BuildingBlock] Platform Init
[BuildingBlock] Networked Avatar
[BuildingBlock] Player Name Tag
[BuildingBlock] Passthrough
[BuildingBlock] Shared Spatial Anchor Core
[BuildingBlock] Colocation
ContentRoot
marker
grad point
[BuildingBlock] Cube
[BuildingBlock] Cube
Whiteboard
Persp
Project
Console
Favorites
All Materials
All Models
All Prefabs
Assets
Oculus
ParrelSync
Plugins
Resources
Scenes
Settings
Streaming...
TurboSqui...
TutorialInfo
Whiteboard
XR
DefaultNe...
P121
Readme

IP2 :


Hierarchy
Scene
Game
Inspector
SampleScene
Main Camera
Directional Light
Global Volume
[BuildingBlock] Camera Rig
[BuildingBlock] Passthrough
[BuildingBlock] Platform Init
[BuildingBlock] Player Name Tag
Whiteboard
Marker
Cube
[BuildingBlock] Cube
DontDestroyOnLoad
Persp

IP3:


Quest MR Draw2 - DemoScene 1 - Android - Unity 2022.3.62f2 <DX11>
File Edit Assets GameObject Component Services Meta Window Help
Hierarchy
Scene
Game
Inspector
DemoScene 1
Teleportation Environment
Surface
Directional Light
[BuildingBlock] Passthrough
PointableCanvasModule
OVRCameraRig
Canvas
GameObject
Cube (1)
[BuildingBlock] Shared Spatial Anchor Co
Clear Brush
Fountain pen
Compass
Ruler
Persp
AI Navigation
Surfaces
Show Only Selected
Show NavMesh
Show HeightMesh
Agents
Show Path Polygons
Show Path Query Nodes
Show Neighbours
Show Walls
Show Avoidance
Obstacles
Show Carve Hull
Project
Console