What Sensors See, What People Feel: Exploring Subjective Collaboration Perception in Mixed Reality

Yasra Chandio*
ychandio@umass.edu
University of Massachusetts Amherst

Diana Romero*
dromero@uci.edu
University of California, Irvine

Salma Elmalaki
salma.elmalaki@uci.edu
University of California, Irvine

Fatima M. Anwar
fanwar@umass.edu
University of Massachusetts Amherst
*

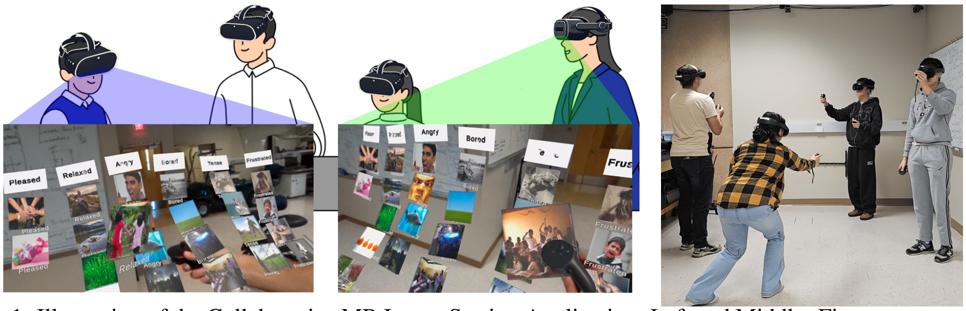

Figure 1: Illustration of the Collaborative MR Image Sorting Application. Left and Middle: First-person perspective of the application interface as experienced by two participants. Right: Third-person view showing participants interacting with the application in a shared MR environment.

Abstract

Mixed Reality (MR) enables rich, embodied collaboration, yet it is uncertain if sensor and system-logged behavioral signals capture how users experience that collaboration. This disconnect stems from a fundamental gap: behavioral signals are observable and continuous, while collaboration is interpreted subjectively, and shaped by internal states like presence, cognitive availability, and social awareness. Our core insight is that sensor signals serve as observable manifestations of subjective experiences in MR collaboration, and they can be captured through sensor data such as shared gaze, speech, spatial movement, and other system-logged performance metrics. We propose the Sensor-to-Subjective (S2S) Mapping Framework, a conceptual model that links observable interaction patterns to users' subjective perceptions of collaboration and internal cognitive states through sensor-based indicators and task performance metrics. To validate this model, we conducted a study with 48 participants across 12 MR groups engaged in a collaborative image-sorting task. Our findings show a correlation between sensed behavior and perceived collaboration, particularly through shared attention and proximity.

Index Terms: Mixed Reality, Collaboration, Perception

1 Introduction

Collaboration in mixed reality (MR)[1] environments are becoming increasingly prevalent across domains such as education, design, healthcare, and remote work [18]. MR uniquely blends physical and virtual realms, allowing co-located and remote users to share digital content seamlessly. As MR becomes more immersive and integral to collaboration, researchers have investigated the effects of gaze, gesture sharing, and virtual replicas on the presence and cognitive load [4, 62]. Despite these advances, a fundamental challenge persists: *How can we effectively understand the quality of collaboration in MR?*

Modern MR systems generate extensive behavioral data through system logs and embedded sensors, capturing where users look [47], how they move [34], and when they speak [52], yet these signals often provide only limited insights into users' internal states [73]. Prior work shows that speech, gaze, and turn-taking are tightly connected to the internal perception of collaboration [11]. However, two core limitations remain: (1) it is unclear whether sensor data meaningfully reflects how participants perceive collaboration, or if key experiential dimensions go undetected. (2) Existing studies demonstrate specific bi-modal correlation mappings such as gaze to presence [4], and physiological synchrony to co-presence [9], but stop short of offering a holistic objective-to-subjective mapping of collaboration.

We address these gaps through a systematic, multimodal Sensor-to-Subjective (S2S) mapping framework, applied in a multi-perspective study of collaborative MR involving 48 participants across 12 teams. First, *we analyze sensor-derived indicators of group interaction*, including gaze-based shared attention, conversation dynamics through audio, and spatial proximity, which can be captured through embedded sensing in MR headsets [73, 70]. Second, *we examine task performance using system logs*, measuring timing, interaction patterns, and decision changes throughout the collaboration. Finally, *we assess subjective experience through post-exposure questionnaires* that capture participants' perceptions of group dynamics as well as their individual sense of presence and cognitive effort.

Together, these perspectives allow us to ask: ***Do observable group interaction patterns, as captured by MR head-***

*Yasra Chandio and Diana Romero contributed equally to this work as co-first authors.

[1]**MR** aligns with Milgram's reality-virtuality continuum [38], closely related to Augmented Reality (AR), where virtual elements seamlessly integrate with the real world.

sets, reflect how participants experience collaboration? We formalize this inquiry through three research questions:

RQ1: Do objective sensor-based indicators of group behavior align with participants' subjective perceptions of collaboration in MR?

RQ2: How do sensor-based behavior indicators relate to task performance in a collaborative MR task?

RQ3: What insights can we derive about group behavior in MR from objective sensing and individual subjective experiences?

2 Background and Related Work

2.1 Group Behavior in Immersive Environments

Understanding collaborative group behavior has evolved from traditional settings to MR environments, where users interact in novel ways across co-located, remote, and hybrid configurations. ***Group behavior***, all forms of interaction and activity within a group [22], and ***collaboration***, defined as the intentional, coordinated effort among members to work together toward shared goals [22], takes distinctive forms in MR contexts due to unique factors including free movement [44], embodied gesture [27], co-presence [55], and shared spatial context [43]. Research has examined MR collaboration through interface design [7], task coordination [71], communication patterns [3], and trust development [6], yet it generally relies on external observations or outcomes rather than internal experiences of the participants. This creates a critical knowledge gap on how participants' internal experiences align with the observable behaviors logged by MR systems.

2.2 Sensor-Based Indicators of Group Behavior

Sensor-based methods across ubiquitous and immersive computing characterize group behavior through multiple data streams (gaze, voice, motion, location) without relying on external observation or self-reporting. Remote collaboration systems, such as TeamSense [73] and CoCo [50] monitor nonverbal cues to track group cohesion, while wearable systems assess synchrony and proximity patterns during collaborative tasks [58]. In MR, headset-embedded sensors can track gaze, object manipulation, and gesture alignment, which researchers have shown correspond to collaboration quality [70, 26]. Studies have correlated gaze and gesture sharing with co-presence [4], and linked physiological synchrony and gaze with perceived collaboration quality [9]. Other systems enhance spatial awareness by embedding gaze cues [28], while multimodal integrations of gaze and gesture have been used to improve performance and presence [5]. Additional work shows high-accuracy cognitive load prediction using multimodal sensing [25] and demonstrates that even minimal AR cues can influence team experience and [61]. Conversation-focused research further emphasizes how speech alignment, gesture, gaze, and turn-taking synchronize to create effective collaboration in co-located settings [60, 11]. These findings emphasize that collaboration is fundamentally multimodal, involving a convergence of signals across time.

Despite these insights, existing work remains limited in two ways: (1) isolated pairwise mappings rather than producing a generalizable mapping model and (2) lens-like views often aim to enhance either performance outcomes or structural interaction properties (such as coordination patterns and role distribution) without examining their relationship to participants' subjective experiences. Prior work has explicitly recognized the need for structured frameworks that integrate multiple sensing modalities and interpret internal collaboration states. Barrett et al. call for multi-method frameworks that integrate sensor modalities to capture internal collaboration states [48], while social signal processing emphasizes the need to fuse multimodal behavioral signals to capture fine-grained group dynamics [61]. Systematic reviews also note the absence of frameworks that link objective data to users' subjective experiences [17]. Despite these advances, the critical question of whether sensor-based behavioral indicators meaningfully reflect users' internal perceptions of collaboration remains largely unexplored, particularly in co-located MR tasks using commercial headsets.

2.3 Subjective Indicators of Group Behavior

Subjective self-report surveys remain central to understanding how people experience collaboration. Standardized tools, such as post-task surveys, presence questionnaires [68, 51], and cognitive load surveys [23], are widely used in the Human-Computer Interaction (HCI) to gauge how users evaluate, interpret, and internalize virtual experiences. Prior work has used these subjective measures to assess collaborative quality in various contexts, including team cohesion in distributed work [73], perceived social presence in remote collaboration [45], and mutual understanding during joint problem-solving [72]. Educational studies use them to evaluate contribution equity and group satisfaction [67], while collaborative design work leverages them to assess idea sharing and creative synergy [41].

Despite rich insights, these subjective measures are rarely systematically linked to sensor-based behavioral data. This disconnect raises key questions: can gaze, speech, and proximity signals in MR environments meaningfully reflect users' internal sense of collaboration? Do behavioral patterns align with perception, or challenge our assumptions? Answering these questions requires bridging sensor data and subjective experience in collaborative MR.

To this end, in this paper, we study how sensor-based indicators of group behavior relate to subjective perceptions in collaborative MR. Instead of treating system-logged signals as stand-alone metrics, we treat them as potential reflections of collaborative experience, positioning this alignment as a core concern for human-centered sensing in MR.

3 Approach

3.1 Theoretical Foundation

We conceptualize how group behavior can be interpreted by combining interaction traces captured through sensors and system data with subjective user reports.

3.1.1 Subjective Perception of Collaboration

We define collaboration as not just joint actions, but also how individuals interpret, feel, and reflect on their shared experience. In MR, where social cues, task progress, and attention are distributed across a blend of physical and virtual elements, shaped by both individual states and group-level awareness. We treat this as the *subjective perception of collaboration* and operationalize it through *individual-level* and *group-level* dimensions.

Individual-Level Experience. Each user's interpretation of collaboration is filtered through their *sense of presence* [54], *cognitive availability* [56], and *personal contribution* [14]. These passive background states directly influence how users perceive and make sense of social interaction [45]. At its most foundational, collaboration begins with how individuals experience being part of a group in each moment. In immersive settings, this includes:

- *Presence:* A user's sense of immersion ("being there") shapes their sensitivity to others' gaze, gesture, and speech [53, 51]. In our context, higher presence enhances awareness of others' engagement and informs group-level reflection.
- *Cognitive Load:* The mental effort to manage task demands affects users' ability to perceive coordination signals. A taxed user may miss coordination cues, while a

cognitively available one may detect attention shifts or conversational breaks [15]. Thus, we hypothesize that, cognitive load influences not only task performance [30] but also perceived collaboration quality [59].

- *Contribution Awareness:* Users' sense of how meaningfully others contributed serves as a proxy for perceived participation balance [13]. This acts as a bridge between personal effort and social perception. It links internal judgment with observed peer input and becomes especially salient in unstructured, role-free MR settings.

These individual experiences are not isolated metrics. They are influenced by how users interpret others' behaviors in real time. They form the perceptual and cognitive foundation on which group-level reflections are built.

Group-Level Reflection. MR system tracks each user individually and users form post-task judgments about shared experience [57], such as whether they attended to the same objects, communicated effectively, and collaborated cohesively. We capture this through four constructs:

- *Shared Attention Awareness* captures whether users felt others focused on the same virtual elements, indicating shared intent and coordination [39].
- *Conversational Support* reflects whether dialogue helped users understand (clarify) and contribute (invite participation), signaling cohesion and mutual support [21].
- *Proximity Impact* assesses how physical closeness, via MR, affected collaboration responsiveness. Prior work suggests that co-located interaction, being physically near others, improves responsiveness [20].
- *Group Collaboration* measures users' overall sense of team effectiveness based on observed behaviors, internal judgments, and implicit comparison to collaborative expectations [19].

These individually collected measures reflect each user's interpretation of the group, shaped by their state, engagement, and task involvement, forming **subjective readings of an inherently social phenomenon.**

3.1.2 Objective Perception of Collaboration

While subjective reports capture how collaboration feels, immersive systems log how it unfolds. Although these are individual-level measurements, they are not isolated; they reflect what *each user does in relation to others in the group*. The same user performing the task alone would not produce the same sensor traces. These sensor-based signals (derived from sensor data and system instrumentation) are shaped by interaction: who a user looks at, how close they are to others, and when they speak or remain silent. Interpreting such distributed traces in group contexts is key to understanding collaboration objectively. We focus on three core indicators supported by prior work [70, 73] and accessible on commodity MR headsets [37, 1, 2]:

- ***Conversation*** and speech activity have long been used to study group behavior across psychology, anthropology, and HCI [33, 12]. Recent work using ubiquitous sensors underscores the continued importance of conversation [32]. In MR, where interaction blends physical and virtual contexts, speech activity offers valuable insight into collaboration [70]. To this end, logged via headset microphones, we extract speaking time, turn count, and participation variance (non-transcribing content). These metrics, known to reflect dominance and fluency [42], offer insight into how evenly members contribute.
- ***Shared attention***, the ability to jointly focus on an object, underpins coordination and social connection [69]. It has been widely studied in collaborative systems [8, 39], often through gaze-based detection[72]. We build on this by using MR headset eye-tracking to detect overlapping gaze on the same virtual object, enabling continuous, time-resolved logging of shared attention [39]. Though not explicitly communicated, such alignment reflects group awareness and may shape perceived collaboration. We aim to assess whether sensed shared attention reflects users' perception of collaboration.
- ***Proximity*** influences collaboration by shaping knowledge exchange, creativity, and group cohesion [66]. In co-located collaborative scenarios, groups tend to perform better when members are physically near one another rather than dispersed [24, 65]. Studies have observed that individuals often move closer to those with whom they share stronger social bonds [16]. We capture proximity as the average pairwise distance between users using headset position data in a shared coordinate frame.

We focus on group-level distributions of proximity and conversation (overall closeness or speech participation) rather than directional relationships (who was closest to whom or who spoke to whom). Our goal is to capture collective engagement, not dominant roles. Prior work shows that proximity itself (regardless of direction) signals social engagement [16], and evenly distributed proximity and turn-taking indicate group cohesion [73]. Moreover, our subjective measures focus on perceived group collaboration, which is more likely influenced by general closeness than specific spatial arrangements.

Beyond sensor signals, we log task-relevant interactions such as how often each user interacts with content, how frequently prior actions are revised, and the number of unique configurations a group explores before reaching a consensus. These metrics reflect participation balance, coordination, and decision negotiation. We also record task completion time as a coarse measure of collaborative success.

3.2 Individual Reflections and Collaboration

A central question in our study is how individual perceptions of collaboration relate to the sensed behavioral traces. Rather than averaging subjective reports into a "single group state", we examine the distribution of perceptions across the group: Are they aligned or divergent? Do some users feel disengaged or overloaded? Such differences reflect varied interpretations of what collaboration felt like to the user. By linking sensor-based indicators (§3.1.2) with subjective internal perceptions (§3.1.1), we ask: Do shared gaze events map onto users' awareness of shared attention? Does proximity align with perceived cohesion? Do balanced interactions align with users' perception that they and their peers contributed meaningfully to the task?

Importantly, the subjective states are not outcomes of the sensor data; they are independent yet complementary accounts of collaboration. Our framework treats them as the **interpretive anchor** against which we examine the dynamics captured by the system. This allows us to understand what groups did and how collaboration was experienced in the future, yielding insight into how systems might better reflect, support, or adapt to group behavior in MR environments. Together, these indicators form a continuous stream of behavioral data interpreted as isolated features but as *observable patterns of interaction* that may or may not align with how users interpret collaboration subjectively. Rather than labeling groups as "effective" or "cohesive," our goal is to assess whether and how the system's observations reflect how collaboration was actually experienced.

3.3 Sensor-to-Subjective Mapping Framework

We introduce the *Sensor-to-Subjective Mapping Framework (S2S)* (Figure 2) to structure our investigation. It conceptualizes collaborative group behavior in MR across three perspectives: observable system-level interaction traces

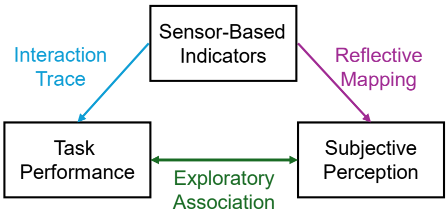

Figure 2: Sensor-to-Subjective Mapping Framework (S2S)

(sensor-based indicators), emergent outcomes from interaction (task performance metrics), and user-reported post-task reflections (subjective perceptions). Unlike prior work [4, 9, 48, 60], assuming one view determines another, our framework treats these perspectives as complementary, each capturing different facets of collaboration as described in §2.2.

3.3.1 From Signals to Perception

We first examine whether sensor-derived behavioral signals correspond to how users interpret collaboration. We call this a relationship. We call this relationship *reflective interaction mapping*. While subjective reports are post-hoc and summarized, we analyze objective indicators over task-level windows that match the granularity of users' reflections. We do not compare specific events to single survey items but assess whether aggregate behavioral patterns (gaze overlap, turn-taking) are reflected in users' overall experience. This avoids false precision and treats subjective reports as interpretive anchors rather than timestamped outcomes. For example, we ask whether shared gaze events align with a user's sense of shared attention. Do balanced speech patterns correspond to reported conversational support? This connection motivates our first set of hypotheses, which examine whether subjective perceptions of collaboration are reflected in the observable sensor-based measures.

3.3.2 Interaction Traces and Performance

The second relationship focuses on how system-logged behaviors (such as how frequently users move shared objects, override prior placements, or coordinate gaze and position) relate to overall group performance. We refer to this connection as trace-to-performance alignment. Here, we test whether groups exhibiting more balanced interaction patterns, frequent shared attention, or tighter spatial coordination complete the task more efficiently or with fewer revisions. These hypotheses assess whether sensor-based data traces align with more fluent or organized group outcomes.

3.3.3 Subjective Perception to Task Performance

Finally, we investigate whether users' internal experiences correspond to observable group-level performance metrics. We refer to this relationship as a subjective influence on performance. Though not sensor-measured, states like presence and cognitive load shape how users engage with others and respond to collaborative demands. For example, lower cognitive load may support better coordination [15], leading to more balanced participation and faster task completion. This mapping tests hypotheses that investigate whether subjective experiences (own experience or the group as a whole) align with observable group outcomes.

3.4 Hypotheses

Together, the three relationships defined by *S2S framework* guide the structure of our hypotheses. We will focus our experimental design on testing the following hypotheses:

H1: Participants' subjective perception of collaboration is reflected in sensor-based indicators (*addressing **RQ1***).

H1.1: Higher perceived collaboration is reflected in more frequent shared attention events.

H1.2: Higher perceived conversation support is reflected in more balanced conversations among group members.

H2: Collective group performance is reflected in sensor-based indicators (*addressing **RQ2***).

H2.1: Groups with more evenly shared task interactions completed the task faster.

H2.2: Shorter task completion time is reflected in more frequent shared attention events.

H3: Participants' individual experiences are reflected in sensor-based indicators (*addressing **RQ3***).

H3.1: Higher individual perceived presence and perception of contribution are reflected in more frequent shared attention events among group members.

H3.2: Higher individual presence is reflected in closer proximity among group members.

H3.3: Lower individual cognitive load is reflected in evenly distributed task interactions within the group.

Table 1: Participant Demographics. The key for frequency: never/almost never; rarely (<2 times); occasionally (a few times); frequently in the past; frequently (>2 times/month).

Demographics	Number of Participants
Gender	36 Male, 12 Female
Frequency of AR Experience	22 Never Used, 15 Rarely, 6 Occasionally, 3 Frequently, 2 Frequently in the past
Frequency of VR Experience	21 Never Used, 13 Rarely, 7 Occasionally, 6 Frequently, 1 Frequently in the past
Frequency of Gaming	4 Never Used, 10 Rarely, 15 Occasionally, 16 Frequently, 3 Frequently in the past
Familiarity to Other Members	25 No Members, 15 One Member, 6 Two Members, 2 Three Members

4 User Study

4.1 Participants

We recruited 60 participants; data from 12 participants were discarded due to technical issues, resulting in 48 participants being included in the analysis. Participants were allowed to form their own groups, or, if preferred, the research team randomly assigned them to a group. In total, 12 groups completed the study. In similar research on collaborative behavior involving small groups, the group is typically defined as having three or more members [64]; in our study, we formed a group with 4 members. Prior studies have revealed patterns of conversation, interaction, and coordination within groups of four in both desktop and virtual reality environments [70]. By adopting a group size of four participants, we aimed to create a richer collaborative environment, capturing the complexities of group behavior not as evident in smaller groups [3, 40]. Participants' ages ranged from 21 to 42, with a mean age of 24. A summary of the participant demographics can be seen in Table 1. All participants provided verbal informed consent. Each participant had normal or corrected-to-normal vision. The institution's ethics committee approved the study.

4.2 Materials

To capture sensor data on user interactions, we used the Meta Quest Pro [37]. The integrated sensors enabled audio recording, eye-tracking, six-degrees-of-freedom (6 DoF) simultaneous localization and mapping (SLAM) tracking, and MR capabilities. The collaborative app is built in Unity using Meta XR APIs to log audio, gaze, location, and object interactions [36]. For precise manipulation of virtual objects, we use the controller integrated with the Meta Quest Pro [37]. The interaction recording and tracking are validated by others [46]. Participants were invited to a shared lab room with a designated 10 ft $\times$ 5 ft space cleared of materials to minimize distractions. They were informed that they could move freely within this area during the task.

4.3 Experimental Task

This section summarizes the cues, interaction, and feedback of our collaborative image-sorting task. The study was deliberately designed around a single collaborative task under one shared condition, without varying levels of network stressors or task types. Our goal was not to compare multiple experimental conditions but to examine the richness of group behavior in a naturalistic, unconstrained collaborative setting. Prior work has shown that tightly controlled comparisons can obscure the variability and fluidity of real-time collaboration in MR environments [70]. Instead, we focused on collecting high-resolution behavioral data and subjective reflections during a consistent, shared experience across groups. Each group of four participants completed one image-sorting task using the same images and categories. Participants were instructed to work together to reach a consensus on the grouping of each image.

4.3.1 Primary Task

Participants sorted 28 images from the OASIS dataset [31], which includes 900 validated images rated by 822 participants for pleasantness and arousal. Images were selected to span diverse emotions while avoiding distressing content. Participants sorted the selected images into one of six emotion categories randomly chosen from Russell's circumplex model of affect [49]: angry, bored, relaxed, tense, pleased, or frustrated. Image sorting tasks have been shown to foster decision-making, communication, and social coordination by building shared mental models and group alignment [10]. We focus on this collaborative task marked by asynchronous, flexible participation, where participants can contribute and modify inputs independently. This repeated image-sorting task, involving open-ended discussions on the emotions elicited by each image, allows us to observe a dynamic, iterative collaborative process among the group, where participants achieve a shared goal through incremental steps and consensus. Each group of four participants completed one image-sorting task on the same 28 images and categories. Participants were instructed to work together as a group to reach a consensus on the label of each image, with no time limit for completing the task, allowing participants to engage in deliberate discussion and negotiation. The labels are not placed in a fixed position, allowing participants to organize and use the room space however they want. The task ends once they inform the researcher that they agree with the image groupings.

4.3.2 Virtual Scene and Cues

As shown in Figure 1 (left, headset's first-person perspective), all 28 virtual images are scattered around the room, and all six emotion category labels (gray virtual plates) are pasted a little higher than where the images are scattered. Participants can view these images and labels at all times via their headsets. For participants, the cue to start the interaction is not defined by the researcher but decided by each participant which image they want to discuss with the group to sort. This lack of structure in cues is by design, as our goal is to observe open and free collaborations without participants taking turns or being directed by the flow of the virtual scene designed by the researcher. The participants are assumed to take the cue for virtual interaction from the other three group members as shown in Figure 1 (right), where all four participants are physically close, probably examining the same image and discussing the final label.

4.3.3 Interaction and Feedback

To sort an image, the participants were asked to physically move the virtual image near the virtual label. Once the image is pasted close to the label, the image is recorded as sorted. Participants used a point-and-drag near-interaction motion with the grip buttons on their left or right controllers to move an image to the corresponding label. They pointed at an image, pressed and held the grip button to drag it, and released it to place it. The image followed the controller's pointer and locked into position when released. Participants could only grab objects within reach and were instructed to release the grip button to secure an image in place. Only one participant could move an image at a time, but multiple images could be held simultaneously. This interaction mimicked real-world object placement and provided immediate feedback once placed.

Table 2: Proposed group behavior perception questionnaire (superscripts refer to the conceptual basis for each item.)

Dimension	Question
Contribution Awareness	How much did you feel other group members contributed during the task?[13]
Joint Attention Awareness	How often did you feel you were paying attention to the same virtual object as other participants?[63]
Proximity Impact	I felt that my proximity to others affected my collaboration during the task.[20]
Conversational Support	How much did the group conversation help you understand the task and contribute effectively?[21]
Group Collaboration	The group worked together effectively to complete the task.[19]

4.4 Measurements

This section outlines the measures we used to capture group behavior in the image sorting task, categorized into sensor level, task-related performance metrics, and subjective measures. At the sensor level, we collected data from the headset using custom scripts. We recorded the audio signal from the microphone, x, y positions for location, and eye-tracking data to capture the data for conversation, proximity, and shared attention as described in §3.1.2.

On the task level, we recorded the various virtual object interactions, such as the number of virtual image interactions per participant in a group, to count if a participant grabs a virtual image and then releases it. Throughout the task, we capture the number of label changes per group to count the number of times a particular image changes its label. We also captured distinct groupings for each image per group to count the distinct labels for each image. For instance, if Participant A moved an image to label X, Participant B moved it to label Y, and Participant A moved it back to X, the image would have three label changes and two distinct groupings. We also collected high-level performance metrics, such as completion time. We measured completion time as the time elapsed from when the first image was grabbed to when the last image was placed, indicating the group's overall time completing the task.

Finally, we collect post-task subjective measures via surveys after the image sorting task, such as presence with the IPQ [51] and PQ [68] questionnaires, cognitive load through NASA-Task Load Index (NASA-TLX) [23][2], and perception of the group behavior through a custom-designed questionnaire. The PQ evaluates factors such as the possibility to act and examine, realism, self-evaluation, and interface quality. The IPQ measures spatial and general presence, realism, and involvement. The presence scores are derived from 33 items (14 IPQ and 19 PQ, the cognitive load score is derived from 5 items from NASA-TLX, and group behavior from 5 items from our custom-designed survey on a 7-point scale. We developed a custom questionnaire to measure participants' perspectives of how their group interacted, as shown in Table 2. The proposed group behavior characterization questionnaire assesses key aspects such as contribution awareness, shared attention, proximity impact, con-

[2]We are aware of the criticism surrounding the use of NASA-TLX. We applied it to measure "perceived" cognitive workload, rather than actual mental load [35].

Table 3: Descriptive and statistical results for PQ, IPQ, NASA TLX, and Group Behavior Questionnaire. N = 48
Metrics: Mean (μ), Standard Deviation (σ), Standard Error (SE), 95% Confidence Interval (CI), 5th/95th Percentiles (P5/P95), Min, and Max.

Measure	μ	σ	SE	95% CI	P5	P95	Min	Max
PQ Subscales								
PQ-REAL	5.16	1.10	0.17	4.82–5.49	3.49	6.86	2.86	7
ACT	5.80	0.89	0.13	5.53–6.07	4.50	7.00	3.50	7
IFQUAL	5.29	1.11	0.17	4.95–5.63	3.38	6.67	2.33	7
EXAM	5.67	0.83	0.13	5.42–5.93	4.33	6.95	4.00	7
EVAL	5.81	0.92	0.14	5.53–6.09	4.50	7.00	3.00	7
PQ	**5.46**	**0.75**	**0.11**	**5.23–5.69**	**4.27**	**6.62**	**4.11**	**6.95**
IPQ Subscales								
INV	3.68	1.53	0.23	3.22–4.15	1.54	6.60	1.00	7
SP	5.07	1.21	0.18	4.71–5.44	2.89	6.77	1.00	7
GP	5.32	1.55	0.23	4.85–5.79	2.00	7.00	1.00	7
IPQ-REAL	4.00	1.21	0.18	3.63–4.37	2.25	5.96	1.25	6.5
IPQ	**4.39**	**1.12**	**0.17**	**4.05–4.73**	**2.52**	**6.21**	**1.07**	**6.5**
PQ + IPQ	**4.92**	**0.86**	**0.13**	**4.66–5.18**	**3.73**	**6.17**	**2.69**	**6.72**
NASA TLX								
NASA TLX	2.30	0.93	0.14	2.01–2.58	1.03	3.80	1.00	4.2
Custom Group Behavior Questionnaire								
Cohesion	6.55	0.85	0.13	6.29–6.80	5.00	7.00	3.00	7
Attention	5.16	1.36	0.21	4.74–5.57	3.15	7.00	2.00	7
Proximity	3.93	1.99	0.30	3.33–4.54	1.00	7.00	1.00	7
Conversation	6.20	1.00	0.15	5.90–6.51	4.15	7.00	3.00	7
Collaboration	6.64	0.84	0.13	6.38–6.89	5.15	7.00	3.00	7

versational support, and overall group collaboration. Each question is informed by established research to ensure relevance to our study's context.

4.5 Study Procedure

Upon arrival, participants received a detailed information sheet outlining procedures, data collection, and privacy. The researcher also provided a verbal briefing, covering headset interactions, controller gestures, and visual stimuli details such as their color, shape, duration, cues, and feedback mechanism for the image-sorting task. Participants were given ample time to consider their participation in the study and were asked for their verbal consent. Participants then completed a demographic questionnaire covering age, gender, tech familiarity, and familiarity with group members.

MR headsets were then distributed to the participants, and they were instructed to calibrate the focus and fit of the headset for maximum comfort. Before the main task, participants completed a tutorial application with two images and two categories not included in the main task to prevent learning effects. This tutorial task aimed to familiarize them with the task interactions in terms of gestures and get them comfortable with using the point-and-drag interaction from the controller. Following this, the group proceeded with the main image-sorting task. They were informed that there was no time limit for the task and that the main requirement was for them to reach a consensus on the image sorting for the task to end. To encourage a more natural and unconstrained group behavior, participants were not informed that they were being timed or evaluated on their accuracy.

Upon task completion, participants filled out a post-task questionnaire on cognitive load, presence, and perceived collaboration. The time it took for each group to complete the task varied; however, the total duration of the session, including consent, briefing, training, calibration, task, and surveys, lasted under an hour.

5 Results

5.1 Perceived Group Behavior & Tasks Summary

We begin with an overview of participants' perceived experiences, task performance metrics, and group-level collaboration indicators. These results offer a joint view of how groups behaved and how they experienced the collaboration. This allows us to assess alignment and divergence later across sensor-based indicators, performance metrics, and subjective perceptions of collaboration.

Table 4: Group-Level Task Metrics Summary.

Group	Num of Images Grabbed	Total Num of Image Grabbing	Num of Image Labels Overridden	Total Images Looked At	Completion Time (seconds)	Num of Label Changes
1	50.0	232.0	22.0	109.0	415.54	56.0
2	52.0	378.0	24.0	112.0	620.59	72.0
3	71.0	497.0	43.0	111.0	676.78	88.0
4	51.0	254.0	23.0	110.0	513.09	55.0
5	54.0	306.0	26.0	110.0	**209.26**	61.0
6	60.0	320.0	32.0	112.0	622.31	71.0
7	71.0	388.0	44.0	111.0	562.77	101.0
8	58.0	**436.0**	30.0	112.0	**994.05**	59.0
9	50.0	220.0	22.0	112.0	430.89	52.0
10	70.0	458.0	42.0	112.0	652.17	84.0
11	72.0	378.0	44.0	112.0	534.00	92.0
12	60.0	318.0	84.0	112.0	573.91	65.0

We summarize the descriptive statistics illustrating participants' subjective experiences with response variability in Table 3. Metrics such as mean (μ) and standard deviation (σ) PQ ($\mu = 5.46$) and IPQ ($\mu = 4.39$) indicate moderate presence levels. We also reported sub-scales: realism (PQ-REAL, IPQ-REAL), possibility to act (ACT), interface quality (IFQUAL), possibility to examine (EXAM), self-evaluation of performance (EVAL), involvement (INV), general presence (GP), and spatial presence (SP). The PQ-REAL subscale ($\mu = 5.16, \sigma = 1.1$), suggests moderate realism, while the ACT subscale shows a high $\mu = 5.18$, indicating strong perceived action capability. INV subscale variability ($\mu = 3.68, \sigma = 1.53$) highlights differing engagement levels. SP ($\mu = 5.07$) reflects strong spatial awareness. The NASA TLX score ($\mu = 2.3$) points to a low perceived workload. Group cohesion ($\mu = 6.55$) and collaboration scored high ($\mu = 6.64$), while group proximity varied ($\mu = 3.93$), indicating differing closeness perceptions.

Next, we present two complementary sets of group-level metrics. A summary of behavioral interaction statistics captured through system instrumentation in Table 4. These include the number of images grabbed per participant, total grabs, label overrides, label changes, and task completion time. These metrics reflect the group's engagement with the task, coordination complexity, and task performance. For example, Group 8 had a notably high number of total grabs (436) and the longest task duration (994s), suggesting prolonged deliberation or difficulty reaching consensus, whereas Group 5 completed the task most quickly (209s), with lower override activity, possibly reflecting more streamlined decision-making or higher initial agreement.

Table 5 reports group-wise mean values from the post-task questionnaires. These include subjective ratings for presence (PQ+IPQ), perceived cognitive load (NASA-TLX), and five dimensions of group behavior: cohesion, attention, proximity, conversational support, and overall collaboration. The average presence score across groups was moderately high, ranging from 4.15 to 5.4, while TLX scores remained low, indicating generally low cognitive effort. Group cohesion and collaboration ratings remained consistently high (close to or at 7), whereas proximity scores were more variable across groups, aligning with previously observed differences in physical movement and spacing.

5.2 Collaboration via Sensor-based Indicators

This section discusses the participants' reflections on collaboration, alongside behavioral indicators obtained from sensors and system logs, at both individual and group scales. Findings for shared gaze overlap frequency (indicating visual focus on the same virtual image), proximity duration (measuring time spent in physical proximity during interaction), and speaking proportion and variance (representing turn-taking and conversational balance via total speaking time, frequency, and variance in participation) can be seen in Table 6. Normality checks for subjective collaboration scores were conducted using the Shapiro-Wilk test

Table 5: Mean scores per group for Presence (PQ+IPQ), NASA TLX, and group behavior metrics (cohesion, attention, proximity, conversation, collaboration).

Group	Presence	TLX	Coh.	Attn.	Prox.	Conv.	Collab.
1	4.86	2.20	5.75	4.25	5.50	6.25	6.25
2	5.04	2.55	6.75	5.25	4.75	6.25	6.75
3	4.50	**1.90**	**7.00**	4.00	2.25	**7.00**	6.75
4	**5.32**	2.55	6.00	5.00	5.00	6.25	6.25
5	5.22	**1.20**	**7.00**	4.75	4.00	5.25	**7.00**
6	5.40	1.95	**7.00**	5.00	3.50	6.50	**7.00**
7	5.07	2.30	**7.00**	**7.00**	3.50	6.75	**7.00**
8	5.06	2.50	6.50	5.75	3.00	6.50	6.75
9	4.75	2.25	6.50	5.50	3.75	6.25	5.75
10	4.81	**2.95**	6.25	5.00	4.00	5.50	6.75
11	4.15	**2.90**	6.25	5.25	4.00	5.75	6.75
12	4.50	2.70	6.00	6.00	3.75	5.50	6.00

($\rho = 0.286$ at the group level; $p = 2.01e$-04 at the individual level), determining the choice of Spearman or Pearson test based on sensor metric normality. Highlighted table rows indicate statistically significant correlations ($p < 0.05$), offering quantitative insight into how subjective assessments correlate with interaction behavior patterns.

Positive correlations between self-reported collaboration and sensor-based metrics were noted. At the individual level, collaboration scores significantly correlated with both shared gaze overlap frequency ($\rho = 0.364$, $p = 0.010$) and proximity duration ($\rho = 0.302$, $p = 0.036$) as per Spearman tests. Group analysis indicated moderate correlations with shared gaze overlap ($\rho = 0.481$, $p = 0.113$) and proximity duration ($\rho = 0.226$, $p = 0.479$), though these were insignificant. Conversational dynamics at the group level, particularly speaking proportion, were linked to perceived conversational support ($\rho = 0.651$, $p = 0.021$). No significant individual-level correlations were observed between conversational support and speech activity metrics, although directionally positive trends were present.

5.3 Performance via Sensor-based Indicators

This section investigates the association between task performance, measured by the time taken to complete a collaborative image sorting task. We evaluated whether sensor-based metrics could predict the group's performance in task completion. The normality of the metrics and completion time was assessed with the Shapiro-Wilk test. Depending on the data distribution, we employed Spearman correlation for non-normal distributions or linear regression when assumptions were satisfied. For both individual and group levels, the correlation coefficient (ρ), regression coefficient (β), p-values, and R^2 values are presented.

As illustrated in Table 7, visual coordination emerged as a significant predictor of group performance. At both individual ($\rho = 0.671$, $p = 1.80$e-07) and group levels, higher shared gaze overlap correlated with quicker task completion. A linear regression at the group level attributed 74% of the variance in completion time to shared gaze ($R^2 = 0.741$), underscoring its role in group performance. Proximity duration was also significantly related to completion time. At the group level, a linear model accounted for over 61% of the variation in task duration ($R^2 = 0.611$, $p = 0.0026$). Interaction balance metrics were strongly related to task performance. At the individual level, variance in image grabs ($\rho = 0.671$, $p = 1.76$e-07) correlated with more effective task progression. This association was also evident at the group level, where variance in label overrides ($\rho = 0.671$, $p = 1.76$e-07) significantly correlated with completion time ($\rho = 0.759$, $p = 0.004$). These findings indicate that groups with greater shared gaze frequency and balanced image interactions tend to finish tasks faster.

5.4 Experience via Sensor-based Indicators

This section explores whether participants' internal states, specifically presence and cognitive load, are associated with behavioral patterns sensed during the collaborative task. These analyses complement prior sections by focusing not on group-level constructs but on how individual experience is reflected in interaction dynamics.

We first present the correlations between presence and cognitive load scores and sensor-derived indicators of shared attention, proximity, and interaction balance in Table 8. Sensor-based metrics were drawn from the same behavioral indicators introduced earlier, capturing visual coordination, physical closeness, and contribution variability. Presence was not significantly associated with shared attention metrics (shared attention percentage: $\rho = 0.041$, $p = 0.778$), proximity (mean pairwise distance: $\rho = 0.073$, $p = 0.619$), or interaction balance (variance in image grabs: $\rho = 0.008$, $p = 0.955$). This suggests that the individual sense of presence was not directly mirrored in the observable group interaction patterns. For cognitive load, however, we found a positive association with shared gaze overlap at the individual level ($\rho = 0.322$, $p = 0.025$), indicating that users reporting higher mental effort tended to participate in more episodes of shared attention. Other indicators, including proximity and interaction balance, did not show significant relationships with cognitive load ($\rho = -0.177$, $p = 0.226$). These results suggest that internal experience is partially reflected in the observable sensor-based group metrics. In particular, shared visual focus may demand or reinforce cognitive effort. At the same time, spatial positioning and contribution balance may reflect other aspects of collaboration beyond individual mental workload or presence.

6 Discussion

This section analyzes findings in the context of our *S2S* framework by assessing objective behavioral indicators, subjective experiences, and group reflections. We evaluate how these signals match participants' reported collaboration experiences and task results, revealing insights into how collaboration is both enacted and experienced. Significant associations were found between sensed behavior and reported experiences, but only at the group level did they become truly robust. We distinguish between individual and group insights to evaluate the alignment's strength and origin. Shared gaze frequency was a consistent indicator of perceived collaboration and shared attention at both analysis levels. Group-level correlations were particularly robust, highlighting shared visual focus as both an individual and collective phenomenon, aligning with Bai et al. [4] who showed that combining gaze and gesture cues enhances copresence and collaboration effectiveness. In contrast, conversational support exhibited weaker correlations. A significant relationship between speaking balance and perceived support was only identified at the group level, implying that individual perceptions may not align with supportive speaking dynamics, but group-level participation does indicate communicative balance (supporting established findings in [60]). We accept **H1** and **H1.1**, as multiple gaze-based features correlated with perceived collaboration and shared attention. We partially confirm **H1.2**, as conversational dynamics were significant solely at the group level.

Behavioral indicators, particularly shared gaze frequency and proximity duration, demonstrated strong predictive ability for task completion time, especially at the group level. Groups that engaged in mutual observation and remained close completed tasks more rapidly, underscoring the importance of co-orientation in effective collaboration, as shown by our study's metrics. Interaction balance, such as variance in image grabs or label overrides, was also linked with effi-

Table 6: Correlation between subjective collaboration measures and sensor-based indicators with individual and group levels test types.

Sensor-based (Subjective) Metric	Target Normality (Indiv. / Group)	Metric Normality (Indiv. / Group)	Correlation (Indiv. / Group)	p-Value (Indiv. / Group)	Test Used (Indiv. / Group)
Shared Gaze Overlap (Shared Attention)	2.01e-04 / 0.286	7.40e-08 / 0.0014	0.3646 / 0.4814	0.0108 / 0.1131	Spearman / Spearman
Shared Gaze Duration (Shared Attention)	2.01e-04 / 0.286	7.32e-07 / 0.0650	0.3027 / 0.2262	0.0365 / 0.4796	Spearman / Pearson
Jointly Viewed Images (Shared Attention)	2.01e-04 / 0.286	0.2048 / 0.3740	-0.0592 / 0.0861	0.6893 / 0.7902	Spearman / Pearson
Speaking Proportion (Conversation)	2.01e-04 / 0.286	0.0021 / 0.0097	0.0063 / 0.6517	0.9659 / 0.0217	Spearman / Spearman
Speaking Variance (Conversation)	2.01e-04 / 0.286	2.71e-05 / 0.0998	-0.1019 / -0.1117	0.4907 / 0.7297	Spearman / Pearson
Proximity Duration (Collaboration)	2.01e-04 / 0.286	7.32e-07 / 0.0650	0.3027 / 0.2262	0.0365 / 0.4796	Spearman / Pearson
Mean Distance (Collaboration)	2.01e-04 / 0.286	0.5384 / 0.7793	-0.1913 / -0.0905	0.1927 / 0.7797	Spearman / Pearson
Image Grabs (Collaboration)	2.01e-04 / 0.286	3.75e-06 / 0.7734	-0.1498 / 0.1325	0.3094 / 0.6814	Spearman / Pearson
Label Overrides (Collaboration)	2.01e-04 / 0.286	0.0037 / 0.0342	0.0881 / 0.1470	0.5516 / 0.6485	Spearman / Spearman
Image Grabs Var (Collaboration)	2.01e-04 / 0.286	7.97e-11 / 7.27e-05	0.1579 / 0.4921	0.2837 / 0.1042	Spearman / Spearman
Label Overrides Var (Collaboration)	2.01e-04 / 0.286	5.52e-04 / 0.3237	0.0886 / 0.1727	0.5491 / 0.5916	Spearman / Pearson

Table 7: Correlation between task completion time and sensor-based indicators with individual and group levels test types (results show statistically significant differences).

Sensor Metric	Target Normality (Indiv. / Group)	Metric Normality (Indiv. / Group)	Correlation (Indiv. / Group)	p-Value (Indiv. / Group)	Test Used (Indiv. / Group)
Shared Gaze Overlap Frequency	3.48e-04/0.3405	7.40e-08/0.0014	0.6710/0.6923	1.80e-07/0.0126	Spearman / Spearman
Proximity Duration	3.48e-04/0.3405	7.33e-07/0.0650	0.3962/0.6117 (R^2)	1.65e-06/0.0026	Linear / Linear
Image Grabs Variance	3.48e-04/0.3405	7.97e-11/7.27e-05	0.6713/0.6713	1.76e-07/0.0168	Spearman / Spearman
Label Overrides Variance	3.48e-04/0.3405	5.53e-04/0.3237	0.6713/0.7597	1.76e-07/0.0041	Spearman / Pearson
Image Grabs	3.48e-04/0.3405	3.75e-06/0.7734	0.2037/0.4108 (R^2)	0.1649/0.0247	Spearman / Linear

ciency. Notably, these patterns persisted at both individual and group scales, reinforcing the role of participation equity in enhancing fluency [17]. We accept **H2**, **H2.1**, and **H2.2**. Group-level behavior metrics, notably shared attention and proximity, predicted performance, and interaction balance indicated collaborative fluency across both levels.

Sensory behavioral features were weakly associated with subjective states such as the presence or cognitive load. A near-significant individual-level link was observed between cognitive load and shared gaze frequency, but it lacked strength and consistency at the group level. Presence did not correspond clearly with proximity or gaze patterns, suggesting a lack of visible manifestation of internal states in this context. Likewise, no significant relationship emerged between cognitive load and participation balance. **H3** is partially accepted due to some individual-level signal, but not consistently. We reject **H3.1**, **H3.2**, and **H3.3** due to a lack of strong evidence connecting presence or cognitive load to group-level behaviors. Although disappointing at first glance, this result replicates concerns by others [9, 25] that objective signals often fail to capture subjective experience unless paired with richer data sources.

6.1 Implications

Our study indicates three key takeaways for designing collaborative MR systems and for future research that aims to bridge sensed behaviors with user experiences. The link between shared gaze and perceived collaboration implies that **visual coordination acts as a behavioral signal and a perceptual anchor for participants (1)**. Therefore, MR systems should feature indicators such as shared gaze highlights or real-time attention cues to keep users in sync without disrupting task momentum.

Across multiple hypotheses, group-level behaviors more accurately predicted collaboration and task performance than individual metrics. This suggests that **meaningful social patterns emerge through aggregation (2)**, not simply from individual activity. Systems that adapt to team behavior should avoid user-centric interpretations in isolation. Instead, they should incorporate group-level features that reflect distribution, synchronicity, and co-occurrence.

Our results show that while sensed behavior reliably maps onto collaborative perception and task efficiency, it falls short of capturing deeper subjective experiences, such as presence or cognitive availability, in its current form. This reinforces that **not all internal states are externally observable (3)**, even in high-fidelity sensing environments. This absence of mapping is not a flaw it is a vital empirical finding that delineates what current MR sensing can (and importantly cannot) reveal about subjective experience. While prior works have treated gaze or physiological synchrony as proxies for presence or co-presence [28], our results underscore their limitations without deeper context. Our findings align with recent studies showing that even rich physiological or behavioral data often fail to reliably predict internal states unless complemented with richer modalities, such as neural, situational, or self-report inputs [40, 25, 29]. To meaningfully comprehend experience from sensor data, some intermediary techniques are needed to connect high-level behavioral indicators with low-level experiential states. Designers should treat behavioral signals as interpretive anchors, not definitive measures of internal states, reinforcing the cautious approach seen in immersive cognitive sensing research, and instead consider combining sensing with self-reports or adaptive prompting. Hybrid models that integrate system-logged data with lightweight subjective inputs could enable more robust personalization and responsiveness in collaborative MR.

7 Limitations and Future Work

This study offers early insight into the relationship between sensed group behavior, subjective perception, and task performance in collaborative MR settings. At the same time, several limitations should be considered, some by design, others presenting opportunities for future work.

First, we employed a singular task in one experimental condition to distinctly observe collaboration dynamics in a regulated environment. The open-ended image sorting task was chosen due to its flexibility and inherent coordination requirements, which are ideal for observing genuine group behavior. Despite this, the controlled environment might not reflect the complete range of variability in time-sensitive or specialized tasks. Future research should investigate various task types and adaptive system conditions to generalize beyond this setup. The number of groups restricts the statistical power of analyses conducted at the group level. Despite a relatively high participant count, the nature of small-group interactions limits group-level observations. Nonetheless, the sample size was adequate to reveal consistent and interpretable patterns across both individual and group analyses.

Table 8: Correlation between subjective experience and sensor-based indicators with individual and group levels test types (no results reached statistical significance).

Sensor Metric	Target Normality (Indiv. / Group)	Metric Normality (Indiv. / Group)	Correlation) (Indiv. / Group)	p-Value (Indiv. / Group)	Test Used (Indiv. / Group)
Target: Presence (PQ + IPQ)					
Shared Attention (%)	0.8677 / 0.7494	0.3722 / 0.9363	0.0417 / 0.4802	0.7783 / 0.1141	Pearson / Pearson
Jointly Viewed Images	0.8677 / 0.7494	0.2048 / 0.3740	-0.0820 / -0.4956	0.5795 / 0.1013	Pearson / Pearson
Mean Pairwise Distance	0.8677 / 0.7494	0.5384 / 0.7793	0.0736 / -0.0282	0.6190 / 0.9307	Pearson / Pearson
Proximity Duration	0.8677 / 0.7494	7.3267e-07 / 0.0650	-0.0536 / 0.2626	0.7173 / 0.4097	Spearman / Pearson
Variance in Image Grabs	0.8677 / 0.7494	7.9738e-11 / 7.2687e-05	0.0083 / -0.1049	0.9555 / 0.7456	Spearman / Spearman
Target: Cognitive Load (NASA TLX)					
Shared Gaze Overlap	0.0225 / 0.3828	7.3990e-08 / 0.0014	0.3223 / 0.2872	0.0255 / 0.3654	Spearman / Spearman
Pairwise Distance	0.0225 / 0.3828	0.5384 / 0.7793	-0.1777 / -0.2020	0.2268 / 0.5289	Spearman / Pearson
Variance in Image Grabs	0.0225 / 0.3828	7.9738e-11 / 7.2687e-05	0.1261 / 0.2522	0.3933 / 0.4291	Spearman / Spearman

Future research could build upon this by employing larger-scale deployments or multi-session studies.

Group composition was not experimentally manipulated; participants had the autonomy to self-select or be randomly assigned to groups. This decision sought to observe genuine interaction dynamics instead of creating artificial teams. Nonetheless, pre-existing familiarity among participants might impact conversation, shared focus, and comfort. Although not measured in this study, future research could quantify prior familiarity to better understand its effects.

In addition, we focus on interaction signals such as gaze, proximity, and voice activity, which could be tracked using standard MR hardware. This choice aligned with our aim to create practical, lightweight sensing systems. Yet, using only headset sensors restricted us from accessing detailed affective, physiological, or facial data. Future research should investigate multimodal enhancements for richer sensing, such as emotion recognition or biometric monitoring. Finally, while our framework bridges sensor data and subjective reflection, it cannot fully capture internal states such as frustration, motivation, or attention lapses. These nuanced experiences often remain invisible to behavioral sensing alone. Our study highlights the gap between observable behavior and internal experience, reinforcing the need for interpretive models that integrate multiple modalities and participant feedback.

In sum, the study was purposefully scoped to explore collaborative sensing in MR using a task and instrumentation that balance ecological validity and control. The observed patterns suggest promising directions for expanding system awareness and adaptability while motivating future research to test across conditions, sensor modalities, and longer-term collaboration scenarios.

8 Conclusion

This study examines the alignment between objective sensor data and subjective perception of experiences in MR collaboration. The confirmation of **H1** and **H2** indicates that sensor-derived metrics, such as shared gaze frequency, reliably reflect perceived collaboration and task performance. Nonetheless, the partial validation of **H3** reveals limitations: complex subjective states, such as presence and cognitive load, show limited correlation with objective measures of behavioral data, suggesting that internal experiences are difficult to capture solely through external observations. Future investigations should focus on the interaction between objective and subjective metrics across varied conditions, sensor types, and collaboration contexts.

We situate this work in the growing body of research studying design-responsive, human-aware MR systems. By grounding system- and sensor-level observations in subjective experience, we move toward sensing models that prioritize users' perspectives. This paper contributes a conceptual model, a validated experimental setup, and empirical findings illuminating where sensor data and user experience align and where they diverge. Through this lens, we take a step toward MR environments that do not just support collaboration but sense it in the ways users feel it.

Acknowledgments

This work is supported by the U.S. National Science Foundation (NSF) under grant numbers 2339266 and 2237485.

References

[1] Hololens 2. https://www.microsoft.com/en-us/hololens/hardware. Accessed Sep-2024.

[2] Apple Vision Pro. https://www.apple.com/apple-vision-pro/, 2023.

[3] A. Abdullah, J. Kolkmeier, V. Lo, and M. Neff. Videoconference and Embodied VR: Communication Patterns Across Task and Medium. *Proc. ACM Humam Computer Interaction*, 5(CSCW2), 2021.

[4] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13, 2020.

[5] J. N. Bailenson, A. C. Beall, and J. Blascovich. Gaze and task performance in shared virtual environments. *The Journal of Visualization and Computer Animation*, 13(5):313–320, 2002. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/vis.297. doi: 10.1002/vis.297

[6] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments*, 13(4):428–441, 2004.

[7] J. N. Bailenson, N. Yee, J. Blascovich, A. C. Beall, N. Lundblad, and M. Jin. The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The journal of the learning sciences*, 17(1):102–141, 2008.

[8] R. Bakeman and L. B. Adamson. Coordinating Attention to People and Objects in Mother-Infant and Peer-Infant Interaction. *Wiley, Society for Research in Child Development*, 55(4):1278–1289, 1984.

[9] A. Bayro, Y. Ghasemi, and H. Jeong. Subjective and objective analyses of collaboration and co-presence in a virtual reality remote environment. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2022.

[10] M. F. Bjerre. Card Sorting as Collaborative Method for User-Driven Information Organizing on a Website: Recommendations for Running Collaborative Group Card Sorts in Practice. *Communication & Language at Work*, 4(4), 2015.

[11] D. Bohus and E. Horvitz. Facilitating multiparty dialog with gaze, gesture, and speech. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10. Associa-

tion for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1891903.1891910

[12] R. Brown and S. Pehrson. *Group processes: Dynamics within and between groups*. John Wiley & Sons, 2019.

[13] J. M. Carroll, M. B. Rosson, G. Convertino, and C. H. Ganoe. Awareness and teamwork in computer-supported collaborations. *Interacting with computers*, 18(1):21–46, 2006.

[14] Y. Chandio, V. Interrante, and F. M. Anwar. Human factors at play: Understanding the impact of conditioning on presence and reaction time in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 30(5), 2024.

[15] Y. Chandio, V. Interrante, and F. M. Anwar. Reaction time as a proxy for presence in mixed reality with distraction. *IEEE Transactions on Visualization and Computer Graphics*, 2025.

[16] M. Cristani, G. Paggetti, A. Vinciarelli, L. Bazzani, G. Menegaz, and V. Murino. Towards Computational Proxemics: Inferring Social Relations from Interpersonal Distances. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 2011.

[17] R. A. J. de Belen, H. Nguyen, D. Filonik, D. Del Favero, and T. Bednarz. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. *AIMS Electronics and Electrical Engineering*, 3(2):181–223, 2019.

[18] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131, 2019.

[19] D. R. Forsyth. Recent advances in the study of group cohesion. *Group Dynamics: Theory, Research, and Practice*, 25(3):213, 2021.

[20] M. Gonzalez-Franco, M. Hall, D. Hansen, K. Jones, P. Hannah, and P. Bermell-Garcia. Framework for remote collaborative interaction in virtual environments based on proximity. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 153–154. IEEE, 2015.

[21] C. Goodwin and J. Heritage. Conversation analysis. *Annual review of anthropology*, 19:283–307, 1990.

[22] J. R. Hackman and N. Katz. Group behavior and performance. *Handbook of social psychology*, 2:1208–1251, 2010.

[23] S. G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 2006.

[24] K. Hawkey, M. Kellar, D. Reilly, T. Whalen, and K. M. Inkpen. The proximity factor: impact of distance on co-located collaboration. In *Proceedings of the 2005 ACM International Conference on Supporting Group Work*, 2005.

[25] Y. Hou, Q. Xie, N. Zhang, and J. Lv. Cognitive load classification of mixed reality human computer interaction tasks based on multimodal sensor signals. *Scientific Reports*, 15(1):13732, 2025.

[26] A. Irlitti, M. Latifoglu, Q. Zhou, M. N. Reinoso, T. Hoang, E. Velloso, and F. Vetere. Volumetric Mixed Reality Telepresence for Real-time Cross Modality Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023.

[27] A. Jing, K. May, B. Matthews, G. Lee, and M. Billinghurst. The Impact of Sharing Gaze Behaviours in Collaborative Mixed Reality. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), Nov. 2022.

[28] A. Jing, K. W. May, M. Naeem, G. Lee, and M. Billinghurst. eyemr-vis: Using bi-directional gaze behavioural cues to improve mixed reality remote collaboration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411763.3451844

[29] G. Kim, J. Lee, D. Yeo, E. An, and S. Kim. Physiological indices to predict driver situation awareness in vr. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, pp. 40–45, 2023.

[30] M. Kocur, P. Schauhuber, V. Schwind, C. Wolff, and N. Henze. The effects of self-and external perception of avatars on cognitive task performance in virtual reality. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2020.

[31] B. Kurdi, S. Lozano, and M. R. Banaji. Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, 49(2), Apr. 2017.

[32] M.-H. Lee, Y.-K. Row, O. Son, U. Lee, J. Kim, J. Jeong, S. Maeng, and T.-J. Nam. Flower-Pop: Facilitating Casual Group Conversations With Multiple Mobile Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4), 2018.

[33] J. M. Levine and R. L. Moreland. Progress in small group research. *Annual review of psychology*, 1990.

[34] M. McGill, J. Gugenheimer, and E. Freeman. A quest for co-located mixed reality: Aligning and assessing slam tracking for same-space multi-user experiences. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*, 2020.

[35] R. D. McKendrick and E. Cherry. A deeper look at the nasa tlx and where it falls short. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 62, pp. 44–48. SAGE Publications Sage CA: Los Angeles, CA, 2018.

[36] Meta. Import Meta XR Packages | Meta Horizon OS Developers, 2024.

[37] Meta. Meta Quest Pro: Premium Mixed Reality | Meta Store, 2024.

[38] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented Reality: A Class of displays on the Reality-Virtuality Continuum. In *Telemanipulator and Telepresence Technologies*. SPIE, 1995.

[39] C. Moore and P. J. Dunham, eds. *Joint attention: Its origins and role in development*. Joint attention: Its origins and role in development. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1995. Pages: vii, 286.

[40] N. Numan and A. Steed. Exploring User Behaviour in Asymmetric Collaborative Mixed Reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, VRST '22. New York, NY, USA, 2022.

[41] P. Paulus. Groups, teams, and creativity: The creative potential of idea-generating groups. *Applied psychology*, 49(2):237–262, 2000.

[42] A. S. Pentland. The new science of building great teams. *Harvard business review*, 90(4):60–69, 2012.

[43] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 6:5, 2019.

[44] I. Podkosova and H. Kaufmann. Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 2018.

[45] S. Poeschl and N. Doering. Measuring co-presence and social presence in virtual environments–psychometric construction of a german scale for a fear of public speaking scenario. *Annual Review of Cybertherapy and Telemedicine 2015*, pp. 58–63, 2015.

[46] N. S. published. Are Quest Pro controllers worth buying for Quest 2? https://www.androidcentral.com/gaming/virtual-reality/are-quest-pro-controllers-worth-buying-for-quest-2, Oct. 2022.

[47] Z. Qu, R. Byrne, and M. Gorlatova. "looking" into attention patterns in extended reality: An eye tracking-based study. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2024.

[48] M. A. Rosen, A. S. Dietz, T. Yang, C. E. Priebe, and P. J. Pronovost. An integrative framework for sensor-based measurement of teamwork in healthcare. *Journal of the American Medical Informatics Association*, 22(1):11–18, 2015.

[49] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, Dec. 1980.

doi: 10.1037/h0077714

[50] S. Samrose, R. Zhao, J. White, V. Li, L. Nova, Y. Lu, M. R. Ali, and M. E. Hoque. CoCo: Collaboration Coach for Understanding Team Dynamics during Video Conferencing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4), 2018.

[51] T. Schubert, F. Friedmann, and H. Regenbrecht. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, 10(3), 2001.

[52] C. Shi, X. Xu, T. Zhang, P. Walker, Y. Wu, J. Liu, N. Saxena, Y. Chen, and J. Yu. Face-mic: inferring live speech and speaker identity via subtle facial dynamics captured by ar/vr motion sensors. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021.

[53] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[54] M. Slater, V. Linakis, M. Usoh, and R. Kooper. Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 1996.

[55] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small-group behavior in a virtual and real environment: A comparative study. *Presence*, 9(1):37–51, 2000.

[56] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *IEEE Virtual Reality*, 2016.

[57] A. Steed, M. Slater, A. Sadagic, A. Bullock, and J. Tromp. Leadership and collaboration in shared virtual environments. In *Proceedings IEEE Virtual Reality*, 1999.

[58] Y. Sun, D. A. Greaves, G. Orgs, A. F. de C. Hamilton, S. Day, and J. A. Ward. Using Wearable Sensors to Measure Interpersonal Synchrony in Actors and Audience Members During a Live Theatre Performance. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(1), 2023.

[59] Q. Tang, Y. Wang, H. Liu, Q. Liu, and S. Jiang. Experiencing an art education program through immersive virtual reality or ipad: Examining the mediating effects of sense of presence and extraneous cognitive load on enjoyment, attention, and retention. *Frontiers in Psychology*, 13:957037, 2022.

[60] J. Thomas, S. W. Lee, A. Giovannelli, L. Lane, and D. Bowman. A communication-focused framework for understanding immersive collaboration experiences. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 301–304. IEEE, 2023.

[61] L. Thomaschewski, N. Feld, B. Weyers, and A. Kluge. I sense that there is someone else: an exploratory study on the influence of the media richness of augmented reality-based assistance systems on team experience and performance. *Frontiers in Virtual Reality*, 4, 2023.

[62] H. Tian, G. A. Lee, H. Bai, and M. Billinghurst. Using virtual replicas to improve mixed reality remote collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 29(5), 2023.

[63] M. Tomasello. Joint attention as social cognition. In *Joint attention*, pp. 103–130. Psychology Press, 2014.

[64] I. S. University. 8.2 Defining Small Groups and Teams. Book Title: Introduction to Public Communication Publisher: Originally published by Indiana State University.

[65] J. S. Valacich, J. F. George, J. Nunamaker Jr, and D. R. Vogel. Physical proximity effects on computer-mediated group idea generation. *Small Group Research*, 25(1):83–104, 1994.

[66] G. Vissers and B. Dankbaar. Knowledge and Proximity. *European Planning Studies*, 21(5):700–721, May 2013. doi: 10.1080/09654313.2013.734459

[67] S. Volet, M. Vauras, A.-E. Salo, and D. Khosa. Individual contributions in student-led collaborative learning: Insights from two analytical approaches to explain the quality of group outcome. *Learning and Individual Differences*, 53:79–92, Jan. 2017.

[68] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 1998.

[69] W. Wolf, J. Launay, and R. I. M. Dunbar. Joint attention, shared goals, and social bonding. *British Journal of Psychology*, 107(2):322–337, 2016. doi: 10.1111/bjop.12144

[70] Y. Yang, T. Dwyer, M. Wybrow, B. Lee, M. Cordeil, M. Billinghurst, and B. H. Thomas. Towards Immersive Collaborative Sensemaking. *Proceedings of the ACM on Human-Computer Interaction*, 6:722–746, Nov. 2022.

[71] N. Yee and J. Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290, 2007.

[72] Yiqiu Zhou and Jina Kang. Characterizing joint attention dynamics during collaborative problem-solving in an immersive astronomy simulation. July 2022.

[73] Y. Zhang, J. Olenick, C.-H. Chang, S. W. J. Kozlowski, and H. Hung. TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3), 2018.