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
selmalaki@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's 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, 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

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

1 Introduction

Collaboration within mixed reality (MR)[1] environments is becoming increasingly prevalent across domains such as education, design, healthcare, and remote work [15]. This trend stems from MR's unique capabilities to seamlessly integrate physical and virtual realms, enabling co-located and remote users to interact through shared digital content. However, as these systems evolve to become more intricate, immersive and integral to collaborative work, researchers have explored various aspects of MR collaboration, including the effects of gaze, gesture sharing, and virtual replicas on presence and cognitive load [4, 56], as well as domain-specific applications such as radiography training [64, 9].

Despite these advances, a fundamental challenge persists: *How can we effectively understand the quality of collaboration that unfolds within these environments?* Modern MR systems generate extensive behavioral data through system logs and embedded sensors, capturing where users look [44], how they move [29], and when they speak [48], yet these signals often provide only limited insights into users' internal states [68]. The critical gap in current research is understanding whether such sensor data meaningfully reflect participants' actual collaborative experiences or whether important dimensions of collaboration remain undetected by current measurement approaches.

This paper takes a human-centered approach to this question. Rather than treating collaboration as an externally defined construct, we focus on how participants themselves perceive their group interactions. In doing so, we recognize that collaboration is not merely a behavioral pattern but a subjective [38], socially mediated experience [42]. For MR systems to effectively support teamwork, we need methods

[1]**MR** aligns with Milgram's reality-virtuality continuum [33], closely related to Augmented Reality (AR), where virtual elements seamlessly integrate with the real world.

that not only detect and interpret this experience but also bridge the gap between observable actions and users' perceptions of their collaborative interactions.

To explore this, we study a collaborative MR task using a multi-perspective approach involving 48 participants grouped into 12 teams of four. 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 [68, 65]. 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 headsets, 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 [19], and ***collaboration***, defined as the intentional, coordinated effort among members to work together toward shared goals [19], takes distinctive forms in MR contexts due to unique factors including free movement [40], embodied gesture [23], co-presence [51], and shared spatial context [39]. Research has examined MR collaboration through interface design [6], task coordination [66], communication patterns [3], and trust development [5], 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 now 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 [68] and CoCo [46] monitor nonverbal cues to track group cohesion, while wearable computing approaches assess synchrony and proximity patterns during collaborative tasks [54]. In immersive environments, headset-embedded sensors measure participation and spatial orientation, and researchers demonstrate how gaze and object manipulation reflect collaboration quality in VR tasks [65] and employ volumetric capture to assess gesture alignment in MR learning contexts [26, 22]. Despite these advances, most existing approaches emphasize either performance outcomes or structural interaction properties (such as coordination patterns and role distribution) without examining their relationship to participants' subjective experiences. 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 [62, 47], and cognitive load surveys [20], are widely used in the Human-Computer Interaction (HCI) literature to gauge how users evaluate, interpret, and internalize virtual experiences and their experience. Previous work has used these subjective measures to assess collaborative quality in various contexts, including team cohesion in distributed work [68], perceived social presence in remote collaboration [41], and mutual understanding during joint problem-solving [67]. Researchers have used self-report data in educational settings to evaluate perceived contribution equity and satisfaction with group outcomes [61]. Similarly, studies of collaborative design have incorporated subjective assessments of idea sharing and creative synergy [36].

Despite rich insights, these subjective measures are rarely systematically linked to sensor-based behavioral data. This disconnect raises questions about whether group behaviors in MR environments actually reflect users' internal collaborative experience. If MR systems can detect where users are looking, who is speaking, and how close they are, can those signals be meaningfully interpreted as indicators of subjective collaborative quality? Do behavioral patterns track with subjective perception or challenge our assumptions about collaboration? Addressing these questions requires examining the relationship between sensor data and human experience in collaborative MR contexts.

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

Collaboration is not merely the accumulation of actions performed by individuals; it is shaped by how each user interprets, feels, and reflects on those interactions. Collaboration is what group members come to understand and feel about what they did together. In MR, where social cues, task progress, and attention are distributed across a blend of physical and virtual elements, this perception is especially complex. To understand collaboration in such settings, we must look not only at visible behavior but also at the subjective interpretations users form before, during, and after interaction. These interpretations are shaped by both *individual-level experience* and *group-level awareness*, which together constitute what we refer to in this paper as *subjective perception of collaboration*. As such, we treat the subjective perception of collaboration as a layered construct, one that blends individual experiences with group-level awareness. This section unpacks that construct and explains how we operationalize it in our study.

Individual-Level Experience. Each user in a collaborative MR task brings their own perspective on the interaction. This perspective is filtered through their *sense of presence* [50], *cognitive availability* [52], and *personal contribution* [12]. These factors are not passive background states but directly influence how users perceive and make sense of social interaction [41]. At its most foundational, collaboration begins with how individuals experience being part of a group in each moment. In immersive settings, this includes:

- *Presence*: the degree to which a user feels immersed and situated ("being there") in the virtual environment is the foundational phenomenon. Presence affects how responsive or attuned a user is to the behaviors of others. Prior research has shown that presence increases sensitivity to others' gaze, gesture, and speech, and enhances the salience of social cues [49, 47]. In our context, a greater presence may facilitate awareness of others' engagement, enabling more accurate reflections on group behavior.
- *Cognitive Load*: often understood as the mental effort required to manage task complexity, modulates how much a user can track, interpret, and respond to others. A mentally taxed user may overlook subtle coordination signals, while a cognitively available user may notice shifts in attention, interruptions in conversation, or emerging group patterns [13]. In this way, we hypothesize that cognitive load shapes not just task performance [24], but the perceived quality of collaboration. Higher cognitive load may inhibit attention to group interactions or reduce perceived performance in the collaboration [55].
- *Contribution Awareness*: how much a user feels other members contributed meaningfully to the group's progress. This involves awareness of others' actions and roles within the group. This acts as a bridge between personal effort and social perception. It reflects both internal judgment and observed input from peers. In immersive collaboration, where explicit roles may not be assigned, contribution awareness can become a key proxy for how individuals interpret participation and balance.

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. While the MR system tracks each user individually, users themselves develop relational interpretations: Were we focused on the same thing? Did we support each other's efforts? Did we function as a group? These are not directly observable but accessible through post-task subjective reflection [53]. We capture these interpretations through four key constructs:

- *Shared Attention Awareness* reflects user' recognition that others are attending to the same virtual elements. This perception is not only a sign of visual convergence, but it also reflects shared intention and awareness, which are foundational for coordination and meaning-making [34].
- *Conversational Support* measures whether dialogue with others helped a user understand the task and contribute effectively. It captures the conversation and its function: Did it clarify? Did it invite participation? This reflects group cohesion and mutual scaffolding [18].
- *Proximity Impact* probes how user felt physical closeness (as mediated by MR headsets) influenced collaboration. Prior work suggests that co-located interaction, being physically near others, improves responsiveness. [17].
- *Group Collaboration* captures users' holistic sense of how well the team worked together. This reflects an integration of observed behaviors, internal judgments, and implicit comparison to collaborative expectations [16].

Though collected individually, these measures represent users' interpretations of the group. They are formed through the lens of each user's own state, engagement, and task involvement. In other words, they are **subjective readings of an inherently social phenomenon.**

3.1.2 Objective Perception of Collaboration

While subjective reports offer insight into how collaboration is experienced, immersive systems provide a complementary channel: detailed logs of where users look, how they move, and when they speak. These objective and sensor-based indicators, derived from sensor data and system instrumentation, allow us to capture collaboration not as it is reported but as 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 data are fundamentally shaped by interaction: who a user looks at, how close they are to others, and when they speak or remain silent. Each signal gains meaning only in the context of collective activity. The central challenge lies in interpreting these distributed signals in ways that meaningfully reflect group-level coordination. Our system logs three core types of interaction indicators, chosen for their theoretical grounding in studies of group behavior [65, 68] and practical availability from commodity MR headsets such as Meta Quest Pro [32], Hololens 2 [1], and Apple Vision Pro [2]:

- ***Conversation*** and speech activity has been frequently employed as a key metric for examining behavior within small groups in the fields of psychology, anthropology, and traditional observational studies [28, 10]. Even recent work in the literature utilizing ubiquitous sensors to understand group behavior continues to emphasize the significance of this metric [27]. Studies have demonstrated that the proportion of conversation among humans varies across different media, such as virtual reality and desktop environments [65]. Given that MR immersive experiences offer a unique modality for human interaction with virtual objects while remaining grounded in the physical world, capturing conversation and speech metrics is valuable for gaining insights into group behavior. To this end, speech activity is logged via audio data collected from the microphones embedded in the MR headsets to capture conversation metrics. Rather than transcribing content, we focus on conversational metrics: speaking time distribution, number of turns, and participation variance across group members. These features have been shown to correlate with dominance and conversational fluency in other settings [37].
- ***Shared attention***, a cognitive process involving the ability of individuals to focus with a social partner on a shared object of interest for either intentional or social purposes, has been extensively studied in the literature [7, 34]. Various works in the literature have consistently highlighted the critical role of shared attention in facilitating cooperation and social bonding [63]. Given its pivotal role in collaborative behavior, the significance of shared attention in the context of computer-supported collaborative work has been a subject of considerable exploration [67]. Many studies have used gaze awareness for automatic shared attention detection. Building on this, we detect shared attention episodes by analyzing overlapping gaze vectors on the same virtual object using eye-tracking data from MR headset sensors. This operationalizes shared attention [34] in a continuously measured, time-resolved, and automatically logged way. While shared attention is inherently social, it is not explicitly communicated; users may not be aware of it. We aim to assess whether sensed shared attention reflects users' perception of collaboration.
- ***Proximity*** plays a critical role in collaboration and group interactions. Prior work in organizational behavior has shown that closer physical proximity among team members enhances knowledge exchange, creativity, and team-

work quality [60]. In co-located collaborative scenarios, groups tend to perform better when members are physically near one another rather than dispersed [21, 59]. Additionally, studies have observed that individuals often move closer to those with whom they share stronger social bonds [14]. Building on these insights, we capture proximity as a scalar measure, specifically, the average pairwise distance between users, using positional data from each user's headset in a shared coordinate frame.

We deliberately abstract away from directionality in terms of proximity and conversation (e.g., who was closest to whom or who spoke to whom) and instead focus on group-level distributions. This is because our interest lies in capturing whether everyone engaged with everyone, not in identifying dominant or passive roles. Prior work has shown that proximity itself (regardless of direction) is a meaningful cue of social engagement in co-located settings [14]. Moreover, our subjective measures focus on perceived group collaboration, which is more likely influenced by general closeness than specific spatial arrangements. Prior work supports this approach, showing that evenly distributed proximity and turn-taking indicate group cohesion [68].

In addition to sensor-based metrics, we log task-relevant interactions that reflect how users engage with the collaborative activity over time. These include interaction frequency, object manipulation patterns, and changes to previously completed actions. For example, we track how often each user interacts with shared content, how frequently prior actions are revisited or overridden, and the number of unique configurations a group explores before reaching a consensus. These metrics offer a view into participation distribution, coordination dynamics, and decision negotiation, which are key elements of collaborative processes. We also record overall task completion time, which serves as a high-level indicator of successful collaboration.

3.2 Individual Reflections and Collaboration

A central question in our study is how these individual perceptions of collaboration relate to the behavioral traces we can sense and log. We do not assume that users' subjective responses can be directly averaged into a "true" group state. Instead, we examine the distribution of perceptions across the group: Are they aligned or divergent? Are certain individuals systematically disengaged or overloaded? These variations matter because they signal different interpretations of what collaboration felt like to the user. By examining the relationship between sensor-based indicators (§3.1.2) and subjective reports of internal individual perception (§3.1.1), we can start to ask: Do shared gaze events map onto users' awareness of shared attention? Does higher proximity align with perceptions of a sense of group cohesion? Does a more balanced distribution of 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. Our goal is not to classify groups as "effective" or "cohesive" based on signals but to assess where and how the system's behavioral observations reflect users' own perspectives.

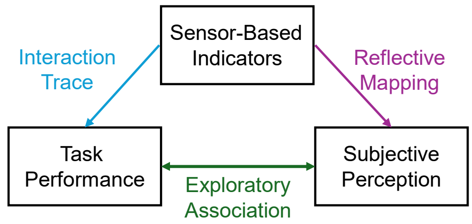

Figure 2: Sensor-to-Subjective Mapping Framework (S2S)

3.3 Sensor-to-Subjective Mapping Framework

To structure our investigation, we introduce the *Sensor-to-Subjective Mapping Framework (S2S)* in Figure 2, which conceptualizes how collaborative group behavior in MR can be interpreted across three perspectives: observable system-level interaction traces (sensor-based indicators), emergent outcomes from interaction (task performance metrics), and post-task reflective accounts from users (subjective perceptions). This framework does not assume that one perspective directly determines the other; it treats each as a complementary lens that captures different facets of collaboration.

3.3.1 From Signals to Perception

The first relationship examines whether behavioral signals captured through sensors reflect users' internal interpretations of collaboration. We refer to this relationship as reflective interaction mapping. For instance, we ask whether shared gaze events, as measured by eye-tracking, correspond to a user's sense that the group had shared attention. Similarly, do balanced patterns in voice activity align with reported conversational support in the post-hoc survey? 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. While individual states, such as presence and cognitive load, are not directly measured through sensors, they shape how users engage with others and respond to collaborative demands. For example, users who report lower cognitive load may have been more available to coordinate or contribute [13], which could result in more balanced participation and faster task completion. This mapping supports hypotheses that investigate whether the way users feel about their own experience or the group as a whole is meaningfully associated with collaborative 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 is 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.

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 [58]; 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 [65]. 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, 35]. 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.

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, 8 Female
Frequency of AR Experience	21 Never Used, 14 Rarely, 6 Occasionally, 1 Frequently, 2 Frequently in the past
Frequency of VR Experience	19 Never Used, 13 Rarely, 7 Occasionally, 4 Frequently, 1 Frequently in the past
Frequency of Gaming	4 Never Used, 8 Rarely, 14 Occasionally, 15 Frequently, 3 Frequently in the past
Familiarity to Other Members	25 No Members, 11 One Member, 6 Two Members, 2 Three Members

4.2 Materials

To capture sensor data on user interactions, we used the Meta Quest Pro [32]. 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 application is developed using Unity with Meta XR packages [31]. Unity and Meta XR API's are used in custom scripts to record the audio, location, gaze-object intersections, and user interactions with virtual objects. For precise manipulation of virtual objects, we use the controller integrated with the Meta Quest Pro [32]. The interaction recording and tracking is validated by others [43]. Participants were invited to a shared lab room with a designated 10 $\texttt{ft} \times 5$ $\texttt{ft}$ 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 cue, 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 [65]. 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 were tasked with sorting 28 carefully selected images from the Open Affective (OASIS) dataset [25]. The OASIS dataset is comprised of 900 validated images by over 822 participants for pleasantness and arousal ratings. The images were selected to represent a range of emotions while excluding potentially distressing content such as violence or graphic imagery. Participants were tasked with sorting the selected images into one of six emotion categories randomly chosen from Russell's circumplex model of affect [45]. These categories included angry, bored, relaxed, tense, pleased, and frustrated. Image sorting tasks have been shown to foster decision-making, communication, and social coordination by building shared mental models and group alignment [8]. 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 all 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 categories 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.

Figure 3: Controller gesture for image and label grabbing.

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?[11]
Joint Attention Awareness	How often did you feel you were paying attention to the same virtual object as other participants?[57]
Proximity Impact	I felt that my proximity to others affected my collaboration during the task.[17]
Conversational Support	How much did the group conversation help you understand the task and contribute effectively?[18]
Group Collaboration	The group worked together effectively to complete the task.[16]

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 the controller at an image, pressed and held the grip button, and moved the image by guiding the controller. The image followed the controller's pointer until the grip button was released, locking the image in its final position. Participants could only grab objects within reach and were instructed to release the grip button to secure an image in place, as demonstrated in Figure 3. This interaction closely mimics the physical action of picking up and placing objects. Once the image was positioned next to the label, it remained stationary when the grip button was released, providing feedback that the image was placed as intended. Only one participant could move each object at a time, but they all were allowed to hold different image objects at the same time.

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 several subjective measures through post-exposure surveys after the image sorting task, such as presence (subjective feeling of being present in a virtual environment) with the IPQ [47] and PQ [62] questionnaires, cognitive load through NASA-Task Load Index (NASA-TLX) [20][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 factors such as 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, conversational support, and overall group collaboration. Each question is informed by established research to ensure relevance to our study's context.

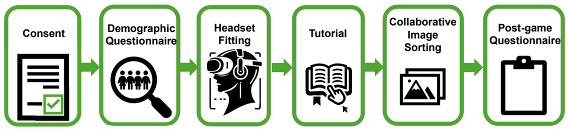

Figure 4: User study procedure: Participants provided consent and completed a demographics survey followed by headset calibration, a tutorial, the main collaborative image sorting task, and a post-exposure questionnaire.

4.5 Study Procedure

The procedure for the study involved several steps, as shown in Figure 4. Upon arrival at the laboratory, participants were given a detailed study information sheet about the study procedures, data collection, and privacy measures. The administering researcher also verbally briefed them on this information. They were also verbally instructed on the headset interactions and controller gestures needed to complete those interactions required during the tasks and how to perform them. The briefing included details about the visual stimuli used in the experiment, 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. Following the briefing, participants were asked to fill out a demographic questionnaire, with questions including their gender, age, familiarity with technology, and with other 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 comfortable with using the point-and-drag interaction from the controller. Following this, the group proceeded with the main image sorting task, where they were tasked to sort 28 images into six different categories. 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.

[2] We are aware of the criticism surrounding the use of NASA-TLX. We applied it as intended to measure "perceived" cognitive workload, rather than actual mental load [30].

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. The sub-scales are 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).

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

Upon task completion, a post-task questionnaire assessing their cognitive load, presence, and perspective of what they think about how the collaboration went with the other group members in the task condition. The time it took for each group to complete the task varied; however, in general, the total duration of the session, including consent, briefing, training, headset calibration, experiment, and questionnaires, took less than 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.

We summarize the descriptive statistics illustrating participants' subjective experiences with response variability in Table 3. Metrics such as PQ ($\mu 5.46$) and IPQ ($\mu 4.39$) indicate moderate presence levels. The PQ-REAL subscale has a mean of 5.16 ($\sigma 1.1$), suggesting moderate realism, while the ACT subscale shows a high mean of 5.8, 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 mean NASA TLX score of 2.3 points to a low perceived workload. Group cohesion and collaboration scored high ($\mu 6.55$ and 6.64), while group proximity varied ($\mu = 3.93$, higher SD), indicating differing closeness perceptions.

Next, we present two complementary sets of group-level metrics. A summary of behavioral interaction statistics captured through system instrumentation is seen at 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 (994.05s), suggesting prolonged deliberation or difficulty reaching consensus. In contrast, Group 5 completed the task most quickly (209.26s), with lower override activity, possibly reflecting more streamlined decision-making or higher initial agreement.

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

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

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 ($p = 0.2864$ at the group level; $p = 2.0125e\text{-}04$ at the individual level), determining 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 ($\rho = 0.3646$, $p = 0.0108$) and proximity duration ($\rho = 0.3027$, $p = 0.0365$) as per Spearman tests. Group analysis indicated moderate correlations with shared gaze overlap ($\rho = 0.4814$, $p = 0.1131$) and proximity duration ($\rho = 0.2262$, $p = 0.4796$), though these were insignificant. Conversational dynamics at the group level, particularly speaking proportion, were linked to perceived conversational support ($\rho = 0.6517$, $p = 0.0217$). 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 ($\rho = 0.6710$, $p = 1.80\text{e-}07$) 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 ($R^2 = 0.7411$), 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 ($R^2 = 0.6117$, $p = 0.0026$). Interaction balance metrics were strongly related to task performance. At the individual level, variance in image grabs ($\rho = 0.6713$, $p = 1.76\text{e-}07$) correlated with more effective task progression. This association was also evident at the group level, where variance in label overrides ($\rho = 0.6713$, $p = 1.76\text{e-}07$) significantly correlated with completion time ($\rho = 0.7597$, $p = 0.0041$). 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 (NASA-TLX) 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: $\rho = 0.0417$, $p = 0.7783$), proximity (mean pairwise distance: $\rho = 0.0736$, $p = 0.6190$), or interaction balance (variance in image grabs: $\rho = 0.0083$, $p = 0.9555$). 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 ($\rho = 0.3223$, $p = 0.0255$), 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 ($\rho = -0.1777$, $p = 0.2268$). These results suggest that internal experience is partially reflected in the observable sensor-based metrics of group behavior. 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 behavioral indicators from sensors and system logs, 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 identified between sensed group behavior and reported experiences. 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. 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. We affirm **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 in group-level aggregates.

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 efficiency. Notably, these patterns persisted at both individual and group scales, reinforcing the role of participation equity in enhancing fluency. 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 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.

6.1 Implications

Our study indicates 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. Therefore, MR systems should feature indicators such as shared gaze highlights or real-time attention cues to enhance awareness of group focus. Such simple interventions can foster shared attention and 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, not simply from individual activity. Systems that infer group dynamics or adapt to team behavior should avoid user-centric interpreta-

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.

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

tions 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, even in high-fidelity sensing environments. 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 be cautious about over-interpreting behavioral signals as proxies for mental state 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. 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 understand its effects better.

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

Table 8: Correlation between subjective experience and sensor-based indicators with individual and group levels test types.

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

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 number 2339266 and 2237485.

References

[1] Hololens 2. https://www.microsoft.com/en-us/hololens/hardware. Accessed Sep-2024. 3

[2] Apple Vision Pro. https://www.apple.com/apple-vision-pro/, 2023. 3

[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. 2, 5

[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 ACM Conference on Human Factors in Computing Systems*, 2020. 1

[5] 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. 2

[6] 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. 2

[7] 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. 3

[8] 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. 5

[9] M. D. Bonis, H. Nguyen, and P. Bourdot. A literature review of user studies in extended reality applications for archaeology. In *2022 IEEE International Symposium on Mixed and Augmented Reality*, pp. 92–101, 2022. 1

[10] R. Brown and S. Pehrson. *Group processes: Dynamics within and between groups*. John Wiley & Sons, 2019. 3

[11] 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. 6

[12] 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. 2

[13] 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. 3, 4

[14] 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. 4

[15] 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. 1

[16] D. R. Forsyth. Recent advances in the study of group cohesion. *Group Dynamics: Theory, Research, and Practice*, 25(3):213, 2021. 3, 6

[17] 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. 3, 6

[18] C. Goodwin and J. Heritage. Conversation analysis. *Annual review of anthropology*, 19:283–307, 1990. 3, 6

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

[20] 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. 2, 6

[21] K. Hawkey, M. Kellar, D. Reilly, T. Whalen, and K. M. Inkpen. The proximity factor: impact of distance on colocated collaboration. In *Proceedings of the 2005 ACM International Conference on Supporting Group Work*, 2005. 4

[22] 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. 2

[23] 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. 2

[24] 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. 3

[25] B. Kurdi, S. Lozano, and M. R. Banaji. Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, 49(2), Apr. 2017. 5

[26] H. Le Tarnec, O. Augereau, E. Bevacqua, and P. De Loor. Improving Collaborative Learning in Virtual Reality With Facial Expressions. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '22 Adjunct. New York, NY, USA, 2023. 2

[27] 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. 3

[28] J. M. Levine and R. L. Moreland. Progress in small group research. *Annual review of psychology*, 1990. 3

[29] 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. 1

[30] 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. 6

[31] Meta. Import Meta XR Packages | Meta Horizon OS Developers, 2024. 5

[32] Meta. Meta Quest Pro: Premium Mixed Reality | Meta Store, 2024. 3, 5

[33] 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. 1

[34] 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. 3

[35] 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. 5

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

[37] A. S. Pentland. The new science of building great teams. *Harvard business review*, 90(4):60–69, 2012. 3

[38] I. R. Pinto, J. M. Marques, J. M. Levine, and D. Abrams. Membership status and subjective group dynamics: Who triggers the black sheep effect? *Journal of personality and social psychology*, 99(1):107, 2010. 1

[39] 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. 2

[40] 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. 2

[41] 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. 2, 3

[42] T. Postmes, R. Spears, K. Sakhel, and D. De Groot. Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, 27(10), 2001. 1

[43] 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. 5

[44] 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. 1

[45] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1980. 5

[46] 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. 2

[47] T. Schubert, F. Friedmann, and H. Regenbrecht. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, 10(3), 2001. 2, 3, 6

[48] 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. 1

[49] 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. 3

[50] 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. 2

[51] 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. 2

[52] 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. 2

[53] A. Steed, M. Slater, A. Sadagic, A. Bullock, and J. Tromp. Leadership and collaboration in shared virtual environments. In *Proceedings IEEE Virtual Reality*, 1999. 3

[54] 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. 2

[55] 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. 3

[56] 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. 1

[57] M. Tomasello. Joint attention as social cognition. In *Joint attention*, pp. 103–130. Psychology Press, 2014. 6

[58] I. S. University. 8.2 Defining Small Groups and Teams. Book Title: Introduction to Public Communication Publisher: Originally published by Indiana State University. 5

[59] 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. 4

[60] G. Vissers and B. Dankbaar. Knowledge and Proximity. *European Planning Studies*, 21(5):700–721, May 2013. doi: 10.1080/09654313.2013.734459 4

[61] 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. 2

[62] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 1998. 2, 6

[63] 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 3

[64] X. Xu, A. Puggioni, D. Kilroy, and A. G. Campbell. User experience of collaborative co-located mixed reality: a user study in teaching veterinary radiation safety rules. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2023. 1

[65] 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. 2, 3, 5

[66] N. Yee and J. Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290, 2007. 2

[67] Yiqiu Zhou and Jina Kang. Characterizing joint attention dynamics during collaborative problem-solving in an immersive astronomy simulation. July 2022. 2, 3

[68] 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. 1, 2, 3, 4