

**Attitudinal, Social, and Behavioral Outcomes in Group Collaboration with Passthrough
Mixed Reality**

Monique Santoso¹, Portia Wang¹, Eugy Han², Jeremy Bailenson¹

¹Department of Communication, Stanford University, Stanford, CA 94305, USA

²Department of Media Production, Management, and Technology, College of Journalism and Communication, University of Florida, Gainesville, FL 32611, USA

Corresponding Author: Monique Santoso

Address: 450 Jane Stanford Way, Stanford University, Stanford, CA 94305, USA

Email: mtsantoso@stanford.edu

Author Contributions: M.S. and J.B. designed research, M.S., P.W., E.H., performed research, M.S. analyzed data, M.S., P.W., E.H., J.B. wrote the paper.

Competing Interest Statement: The authors declare no competing interests.

ORCID ID:

Monique Santoso: <https://orcid.org/0000-0002-7513-3196>

Portia Wang: <https://orcid.org/0000-0003-1704-5718>

Eugy Han: <http://orcid.org/0000-0002-1066-6186>

Jeremy Bailenson: <https://orcid.org/0000-0003-2813-3297>

Abstract

With the rise of full-color, low-latency video passthrough in mixed reality (MR) headsets, this study explores how this medium's technological characteristics influence group communication. Using a mixed-factorial design, 159 university students across 24 groups completed a spatial assembly task and a group discussion. Results indicate that video passthrough increased mental and physical demand during spatial tasks. Behavioral and linguistic analysis revealed that video passthrough mediation led to longer task completion times and lower levels of positive-emotion language compared to face-to-face settings. These findings suggest that the cognitive and social costs of video passthrough are contingent upon specific demand dimensions.

Keywords: Mixed Reality, Group Communication, Social Information Processing

1. Introduction

Mixed reality (MR) with full-color, low latency video passthrough has been marketed as a transformative medium for everyday social interactions, fundamentally altering how we perceive and engage with both physical and digital environments (Doughty, 2022; Rolland et al., 1995). Recent technological advances in MR headsets with full-color video passthrough such as the Meta Quest 3, Apple Vision Pro and the HTC Vive XR Elite have positioned video passthrough as not only a tool for specialized tasks but increasingly as a medium for group communication and collaboration such as music education, theatrical performances, medical training, architectural visualizations and exergames (Banquero et al., 2024; Salter et al., 2023; Zeng et al., 2024; Schier et al., 2023; Guo et al., 2024).

Group interactions introduce unique complexities that are not present in dyadic interactions, including multi-party gaze coordination and speaking order (Harrigan & Steffan, 1983; Kalma, 1992; Lesko & Schneider, 1978), peripheral awareness (Kantharaju & Pelachaud, 2021), social facilitation effects (Buck et al., 1992), and complex turn-taking behaviors (Sacks, Schegloff, & Jefferson, 1974; McKinlay et al., 1994), all of which may be fundamentally altered by the perceptual constraints of video passthrough (Bailenson et al., 2024).

While MR devices are increasingly integrated into collaborative workflows (Holt, 2023; Yoon et al., 2025), empirical research has largely overlooked group-level dynamics, often failing to utilize unmediated face-to-face (FtF) benchmarks. This study addresses this gap by comparing passthrough MR directly to FtF interaction to isolate the medium specific cognitive load and nonverbal coordination. By establishing this benchmark, we move beyond describing MR as a novel tool and instead quantify the specific behavioral and attitudinal shifts that occur when collaborative environments are mediated by current hardware constraints.

Mixed Reality Groups

Two theoretical frameworks that together shed light on group communication in video passthrough environments are Social Information Processing Theory (SIPT) and Cognitive Load Theory (CLT) (Walther, 1992; Hart & Steveland, 1988; Bobis et al., 1993). We integrate these perspectives into an integrative account: technological constraints in video passthrough are associated with increased perceptual and attentional demands (CLT), which may limit the cognitive resources available for encoding, transmitting, and interpreting social cues (SIPT), potentially influencing behavioral coordination and perceived social presence in groups.

SIPT suggests that mediated communication results in less social information transfer and requires more time for processing due to reduced nonverbal bandwidth. CLT offers a complementary perspective, suggesting that the additional processing requirements of video passthrough, such as depth perception adjustments and limited field of view, consume an individual's processing capacity (Walther, 1992). In our framework, cognitive load functions as a plausible intermediary through which technological affordances influence social information processing. Rather than an omnibus effect, these theories suggest that MR mediation taxes specific subdimensions of mental and physical demand, which may in turn restrict the resources available for social coordination.

When applied to video passthrough together, SIPT and cognitive load theory suggest that performance differences in tasks will differ due to the attentional demands of video passthrough compared to face-to-face, however, with enough interaction between participants and familiarity between individuals in groups, groups will eventually achieve comparable communication effectiveness across conditions. Additionally, Social Presence Theory functions as the downstream outcome layer of our model. Social presence provides scaffolding for understanding how changes in social information processing manifest in perceived connection and belonging

(Biocca et al., 2001). In our integrative account, technological constraints lead to increased cognitive load, which causes reduced social information exchange and altered behavioral coordination, creating potential shifts in perceived social presence. This is particularly relevant when considering outcomes such as social presence and loneliness in MR collaboration resulting from technological affordances such as limited field-of-view (Biocca et al., 2001).

2. Background and Previous Work

2.1 Cognitive Load Theory in Video Passthrough Mediated Environments

Cognitive load refers to the mental resources used during task performance (Sweller, 1988). Hart and Steveland's (1988) construct consists of six subdimensions: (1) mental demand, (2) physical demand, (3) temporal demand, (4) performance, (5) effort, and (6) frustration. Measuring how these specific dimensions are influenced by video passthrough is important as it may not necessarily increase overall load in all contexts; rather, it taxes the mental and physical resources required to process a video-recorded version of reality. In video passthrough, measuring cognitive load is particularly relevant as users must process both physical and digital information simultaneously through a video recorded version of reality (Kockord & Bodensiek, 2021; Michalski et al., 2023).

Video passthrough in MR headsets introduces unique sources of cognitive load that are not present in face-to-face interactions. First, the reduced field of view leads to users needing to maintain spatial awareness through increased head movements, which may be especially true in group settings where participants are distributed all around the individual (Bailenson et al., 2024). Second, the stereoscopic rendering of the physical environment through front-facing cameras introduces individuals to subtle but significant alterations to depth perception (Rolland et al., 1995). Third, latency, wherein there exists a difference between the physical and digital

Mixed Reality Groups

signal, results in a perceptual mismatch between proprioceptive feedback and visual information. Finally, the physical presence of the headset itself, introduces a physical discomfort that requires attentional resources to manage due to its weight over one's head - 18.20 oz (Santoso & Bailenson, 2024).

These technological constraints have been shown to impact performance in individual tasks. Baumeister et al. (2017) demonstrated that video passthrough elicits higher mental effort and longer task completion times compared to optical see-through AR or desktop displays. Even simple tasks like brewing coffee required greater cognitive resources when performed using AR headsets (He et al., 2019).

While these cognitive load demands have been studied for individuals using AR, there has been limited research examining these effects in collaborative group contexts. In group settings, users must simultaneously manage technology-related cognitive burdens while attending to complex social dynamics. This situation parallels the general "dual-task interference" effect studied in cognitive psychology research (Pashler, 1994), where attempting to perform two attention-demanding tasks concurrently can lead to performance decrements in either or both domains.

2.2 Social Information Processing Theory

Social information processing theory (SIPT) provides a framework for understanding how communication adapts to the constraints of mediated environments. Walther (1992) proposed that computer-mediated communication (CMC) transfers social information at a slower rate than face-to-face interaction. However, given sufficient prior interactions, communicators can achieve equivalent relational development. Information transfer is slower in CMC primarily because nonverbal cues that typically accompany face-to-face communication, such as facial

Mixed Reality Groups

expressions, body language, and paralinguistic cues, are either absent, reduced, or transformed when filtered through technology (Walther, 1992). This reduced bandwidth necessitates more explicit verbal encoding of social and affective information that might otherwise be communicated nonverbally.

When applied to group contexts, research has illustrated that participants in computer-mediated conferencing used two to three times as many communication units than face-to-face groups during the same elapsed time and were less likely to reach agreement and more recent research illustrated that individuals have greater facial expressivity in face-to-face compared to computer-mediated conferencing (Hiltz et al., 1986). In video passthrough specifically, while many visual cues remain accessible, they are degraded by technical limitations mentioned above, for instance, reduced field of view constraints peripheral awareness, subtle facial expressions may be less discernible, and spatial perception alterations may impact proxemic communication (Santoso & Bailenson, 2024).

2.3 Group Dynamics in Mediated Communication

Unlike dyadic interactions, group interactions rely heavily on peripheral awareness, the ability to monitor multiple participants simultaneously, including those not engaged in the current exchange (Spring & Vathanophas, 2003). The interactivity and fidelity of a medium and how accurately it reflects real-time social signals, is crucial in group communication. Group contexts also introduce complex mutual gaze and attention patterns that are fundamentally different from dyadic exchanges. For instance, being part of a dyad led to more social attention compared to being in a group situation as measured through gaze behavior (Maran et al., 2020). This is because in groups, attention becomes distributed, and the ability to perceive when other

Mixed Reality Groups

group members are attending to each other through mutual gaze provides crucial information about subgroup formation and conversation flow (Jokinen et al., 2013; Vertegaal et al., 2001).

Word count has been used as a proxy for such investment (Huang & Hancock, 2021). Additionally, research in group conversational flow has illustrated that there is a correlation between perceived high- or low-influence of group members among one another by the number of words spoken and conversational flow. Group members with more spoken words had higher perceived social influence and groups with higher conversational flow had more entitativity compared to those with lower conversational flow (Koudenberg et al., 2013; Brooke & Ng, 1986).

Moreover, cooperative behavior is a result of theory of mind activation (Neufeld et al., 2025). Theory of mind posits that social interactions become egocentric when individuals are unable to predict others' behaviors and thoughts, whereby they then predict others' behaviors and thoughts to be similar to their own (Epley et al., 2004). Direct eye contact is an essential factor for theory of mind activation (Calder et al., 2002) and research has shown that eye contact can influence cognitive and affective processes (Neufeld et al., 2025). Given that individuals are unable to see eye-to-eye in many MR headsets, this could lead to a lack of full theory of mind activation leading to increased egocentrism in making social predictions and a lack of cooperation in groups.

2.4 Social Presence and Absence

Social presence refers to the sense of being with another person and experiencing authentic interpersonal interaction, even when that interaction is mediated by technology (Biocca et al., 2001). Social absence, on the other hand, occurs when individuals in a shared environment are perceived as less real or less present (Bailenson et al., 2024). While video passthrough

Mixed Reality Groups

preserves many visual and auditory cues present in face-to-face interaction, subtle technological mediations, such as the slight delay in video processing, distortions in spatial perception, and reduced field of view, may nonetheless undermine the perception of others as fully present (Bailenson et al., 2024). Moreover, participants positioned at the periphery of one's visual field may experience greater social absence effects than those in direct focus (Miller & Bailenson, 2021), potentially creating uneven participation patterns and fracturing the group.

Connected to social presence is the subjective experience of loneliness, where belonging needs are being insufficiently met (Baumeister & Leary, 1995). While seemingly paradoxical in physically co-located groups, loneliness can emerge even in shared physical spaces if the quality of social connection is perceived as insufficient in quality (Wheeler et al., 1983). Prior research in social VR has linked social presence with perceived social support (van Brakel et al., 2023), suggesting that diminished social presence in mediated environments might contribute to feelings of isolation despite physical proximity. In group MR contexts, if participants experience reduced social presence from their peers due to technological mediation, they may paradoxically feel lonelier than in unmediated interaction, despite being physically co-present.

Beyond social presence, how individuals are socially perceived in MR headsets has also been studied. Due (2015) found that Google Glass wearers exhibited poorer rapport development in conversations compared to unmediated dyads, while Koelle et al. (2015) documented more negative social perceptions of AR users in non-work contexts. Miller et al. (2019) investigated dyadic interactions where one participant wore a Microsoft HoloLens, finding that AR users felt less connected to non-headset-wearing partners than to fellow headset users. These findings connect directly to the presence theories discussed earlier, demonstrating how technological interfaces can fragment the social experience even in physically shared spaces.

2.5 Theoretical Integration

Integrating CLT and SIPT provides a conceptual framework for interpreting how video passthrough may influence group outcomes. Our central argument is that technological affordances (e.g., restricted field of view, latency, depth distortions) are associated with additional perceptual and attentional demands (CLT). A plausible account is that these demands limit the cognitive resources available for encoding and decoding social cues (SIPT), particularly in multi-party contexts that require distributed attention. This theoretical lens suggests that the resulting behavioral coordination and perceived social presence in groups are shaped by these underlying resource constraints.

While this study focuses on video passthrough in MR headsets, it is important to situate these findings within broader dimensions that characterize communication technologies. Scholars such as Poole and Walther emphasized that technologies vary along theoretically meaningful continua, such as media richness (Carlson & Zmud, 1999), interactivity (Burgoon et al., 2002), and social bandwidth, that predict social and cognitive outcomes. Video passthrough in MR headsets can be considered a highly rich medium due to its multimodal visual and auditory cues yet constrained in interactivity and nonverbal transparency due to field-of-view and latency limitations. Conceptualizing MR along these dimensions allows our findings to generalize beyond current hardware specifications to future systems, providing a future-proof framework for understanding how increasing sensory fidelity and responsiveness influence group processes and outcomes.

3. Current Study

The current study addresses several critical gaps in the existing literature on group communication in mixed reality. First, while previous research has examined either cognitive or

Mixed Reality Groups

social dimensions of MR interaction, few studies have integrated these perspectives to understand how cognitive load might influence social dynamics in group settings. Second, the specific effects of video passthrough MR on co-located group communication remain largely unexplored, with most existing group studies focusing on either fully virtual environments or optical see-through AR. Third, methodologically, our use of both subjective measures and objective behavioral coding provides a more comprehensive assessment of group dynamics than previous studies relying primarily on self-report. Finally, our inclusion of both spatial assembly and discussion tasks enables examination of how task characteristics influence technological mediation on group processes.

We hypothesize several outcomes for groups in video passthrough compared to face-to-face conditions: (1) increased cognitive load, particularly mental and physical demand; (2) longer task completion times, especially for spatially complex tasks; (3) reduced quantity and emotional richness of verbal communication; (4) diminished awareness of group members, particularly those positioned peripherally; and (5) potential behavioral-perceptual dissociation, defined as a gap where objective behavioral markers of connection decline while subjective perceptions of presence remain stable.

To systematically examine these complex group dynamics in MR environments, we developed a comprehensive codebook based on established educational observation frameworks. This methodological approach is necessary because standard questionnaires alone cannot capture the nuanced behavioral patterns and social interactions that emerge in group MR settings. Our codebook, adapted from the Behavioral Observation of Students in Schools (BOSS) and Classroom Assessment Scoring System (CLASS), provides a structured way to evaluate critical dimensions of group behavior including emotional climate, engagement levels, and off-task

Mixed Reality Groups

behaviors. This observational codebook is particularly valuable for examining how the cognitive and perceptual constraints of MR technology, as predicted by our theoretical framework, manifest in observable group behaviors. By coding dimensions such as positive and negative climate, sensitivity to peers, and active versus passive engagement, we can better understand how the technological mediation of MR affects not only task performance but also the quality of social connection and collaborative engagement beyond self-report. To evaluate the potential for behavioral-perceptual dissociation, we mapped objective behavioral indicators to their corresponding subjective self-report counterparts. Social presence was paired with behavioral coding of group sensitivity scores; closeness and loneliness were paired with total word count and behavioral coding of group awareness; and perceived cognitive load was paired with task completion time. This mapping allows us to identify where technological mediation impairs the objective mechanics of interaction without necessarily degrading the subjective experience of the social bond.

Drawing on our integrated framework combining Cognitive Load Theory (Sweller, 1988; Hart & Steveland, 1988) and Social Information Processing Theory (Walther, 1992), and supported by Social Presence Theory (Biocca et al., 2001), we derive hypotheses that directly test the proposed framework:

H1: Groups interacting in video passthrough will experience higher overall cognitive load than face-to-face groups, particularly in mental and physical demand.

H2: Video passthrough groups will take longer to complete tasks due to increased perceptual and attentional demands than groups without MR.

H3: Video passthrough groups will exchange fewer words and less emotionally rich language than face-to-face groups due to reduced nonverbal bandwidth and higher cognitive load.

H4: Behavioral indicators of connection (e.g., verbal engagement, sensitivity to peers) will diverge from self-reported presence in video passthrough groups, indicating a behavioral-perceptual dissociation between self-reported and behavioral social presence.

These hypotheses provide a link between our theoretical frameworks and the measurable outcomes in this study, clarifying the expected directions of effects across cognitive, behavioral, and affective dimensions of group communication.

4. Methods

4.1 Experimental Design

This study employed a 2 (Condition: Video Passthrough vs. Face-to-Face) x 2 (Activity Type: Spatial Assembly vs. Group Discussion) mixed-factorial design. Condition was a between-groups factor, while Activity Type was a within-groups factor. The unit of assignment was the group (N = 24 groups). The unit of analysis was determined by the level at which the dependent variable was measured. For variables with individual variance (e.g., NASA-TLX subscales, Social Presence, and Closeness), the individual was the unit of analysis. To account for the social interdependence and non-independence of observations, we employed linear mixed models (LMMs) with a random intercept for Group ID. For variables where the metric was inherently shared by all members (e.g., task completion time, total word count), the group served as the unit of analysis.

During the study, participants in the Video Passthrough condition remained in-headset for both the spatial assembly and the group discussion tasks.

4.2 Procedure

Participants met in groups of 6 to 10. Groups were randomly assigned to one of the two between-subjects levels of the Passthrough Function: (1) video passthrough or (2) face-to-face.

Mixed Reality Groups

Within these groups, a within-subjects manipulation of Activity Type was applied; every group completed both the spatial assembly task (assembling an HTC Vive tracking system) and the seated group discussion. To prevent order effects, the sequence of activities was counterbalanced across groups. Specific instructions for each activity are presented in **Table 1**. This study was approved by the Institutional Review Board at [institution blinded] (protocol blinded).

Table 1. Instructions provided to participants for the study tasks.

Task Type	Task Instructions
Spatial	<p>Assemble an HTC Vive headset, controller, and tracking space.</p> <ol style="list-style-type: none">1) First, you want to set up two tripods on opposite sides of the room to outline your tracking space.2) Set up the base stations on top of the tripods.3) Connect the base stations with their cables to an outlet and wait till you see a green light that means your base stations are communicating with each other.4) Setup your headset by connecting its cables to the computer and to an outlet.5) When you have connected these, check the computer and run Steam and SteamVR, two applications on the desktop.6) Look at the SteamVR application. You should see the 2 base stations and headset light up.7) Next, connect your controllers by clicking the menu in SteamVR and click on Controllers.8) Follow the instructors in the pop-up window.9) When they have all lit green in your steam VR, you are now ready to set up your tracking space.10) Follow the instructions on the PC to set up your tracking space.
Discussion	<p>Discuss the following questions:</p> <ol style="list-style-type: none">1) How does ENGAGE, the social VR application you used for discussion last week, enhance or diminish the experiences of individuals with disabilities, such as mobility or visual impairments?2) In what ways can VR provide inclusive educational opportunities for students with different learning needs?3) How can VR platforms facilitate social interactions for individuals who may find in-person interactions challenging?4) How does ENGAGE's ability to customize VR environments improve accessibility for diverse user groups?5) How does tracking play a role for accessibility?

A researcher was present in the room with the groups to help participants with headset logistics, such as putting on headsets, answering questions on the task, and troubleshooting hardware-related issues. This researcher guided participant groups through the tasks. Following the completion of each task, participants completed a questionnaire that inquired about cognitive load, closeness, social presence and loneliness.

Participants were video and audio recorded throughout the session. Video recordings were captured using two cameras that taped the entirety of the activity and researchers used a timer to assess task completion time. Similarly, audio recordings were captured by an external microphone by researchers for the activity in the form of .mp4 files that were then uploaded to Otter.ai for transcription. Given the 85-90% accuracy of Otter.ai, research team members manually reviewed and edited the automated transcription. Edited transcripts were then inputted to the Linguistic Inquiry and Word Count (LIWC) software, which uses dictionary-based natural language processing to extract emotions and sentiments (Pennebaker et al., 2015).

4.3 Apparatus and Setup

Participants used the Meta Quest 3 headset (515 g) with 2064 x 2208 pixels per eye, 110 x 96-degree field-of-view, 39 milliseconds latency, 120 Hz refresh rate, and full-color video passthrough. Hand controllers were not required for use during this study, however, participants were able to hold them. By default, participants' hands were tracked in the video passthrough system and participants were able to see their hands.

***Image 1.** Images taken from inside video passthrough in the Meta Quest 3 and on an Apple iPhone held at eye level of the same speaking position to illustrate field of view limitations. The*

Mixed Reality Groups

headset (left) has a 110° field of view while the Apple iPhone has a 130° field of view (right). Both are considerably less than the human field of view at 200°.



4.4 Participants

Participants were drawn from a sample of university students enrolled in a 10-week course about VR. At the start of the course, students were invited to participate in a study of how VR learning influences individual and group behavior. All students who were part of the course took part in all the activities, however, we only included the data of those who consented to participate in this study after signing a consent form authorized by the university's IRB and an additional oversight organization for students. To eliminate any potential perceptions of coercion, the procedure ensured that researchers and teaching team members were unaware of the identities of participating students until after the course finished by using a third-party arbiter to allocate individuals into groups. Of the 198 student participants who participated in the course, 182 were present for the study, and 159 consented to participate in the study. All participants had the opportunity to interact with others in the group 5 times for group activities prior to the study. These participants were divided into 24 groups, 12 which used video passthrough in MR headsets and 12 which did not (**Supplemental Information A** illustrates the distribution of participants per group).

Mixed Reality Groups

The 159 participants who participated in this study ($M = 78$, $F = 76$, Nonbinary = 2, Decline to Respond = 3), were between 18 and 32 years of age ($M = 20.76$, $SD = 1.74$) and identified as Asian-American or Asian ($n = 65$), Indigenous/Native American, Alaska Native, First Nations ($n = 1$), White ($n = 39$), Middle Eastern ($n = 4$), Native Hawaiian or other Pacific Island ($n = 1$), African, African-American or Black ($n = 14$), Hispanic or Latinx ($n = 14$), more than one race ($n = 17$), other not listed ($n = 2$) and decline to answer ($n = 2$). Participants reported varying levels of experience with VR ($M = 2.66$, $SD = 1.29$).

A post-hoc power analysis for a factorial mixed effects model with fixed and interaction effects was run on G*Power shows that this sample size with an $N = 159$, an effect size of 0.30, error probability of 0.05, for 4 groups reveals a power of 0.89, indicating that there is a relatively high likelihood of detecting a true effect.

4.5 Codebook Development

In addition to the use of validated questionnaire items, we conducted top-down evaluations of classroom group behavior in MR through video observation, as research has illustrated that observational research increases the sensitivity with which one can capture interpersonal relationships and affect (Jones et al., 2021). Drawing from the categories of the Behavioral Observation of Students in Schools (BOSS) and Classroom Assessment Scoring System (CLASS) (Allen et al., 2013), we formulated a coding scheme that fits the nature of classroom tasks in MR. Through an iterative process with two raters, the first author and the two raters discussed and modified the previously published coding schemes.

Following the development of the scheme, the two raters independently labeled each video. Each dimension was rated on a 7-point Likert scale, with 1 = low evidence of behavior reflected in the classroom and 7 = high evidence of behavior reflected in the classroom during

Mixed Reality Groups

the observation cycle. Each observation cycle was the length of the full video, where coders labelled how many times each behavior was displayed for each dimension. All variables in the codebook were coded at the group-level. Disagreements in coding were discussed synchronously after the completion of all videos to ensure an interrater reliability score of 0.90. Prior to the discussion, kappa scores ranged between 0-0.86 (none to near perfect agreement) across all groups. Following the discussion of discrepancies, coders had perfect agreement ($\kappa = 1$) for all dimensions (Cohen, 1960). The group-level measures collected from our codebook were:

Positive Climate. Positive climate refers to the emotional connection between peers and the amount of warmth, respect, and enjoyment communicated through classroom interactions. Examples of positive climate include helping peers with tasks, attending to what peers say in a discussion.

Negative Climate. Negative climate refers to the negativity expressed in the classroom. Examples of negative climate include being distracted from the task and pulling peers' attention away from the task.

Sensitivity. Sensitivity is defined as the participants' awareness of and responsiveness to their peers' academic and emotional needs, high levels of sensitivity for their peers to be able to actively explore the task and the levels of comfort, reassurance and encouragement the group displayed. Examples include being attentive to peer needs, helping out when a peer needed a hand with the spatial task.

Passive Engagement. Passive engagement was defined as times when the participant is passively attending to the assigned work, on task, but passively participating. Examples include head turning when someone asks for help but does not go over to help with a task, looking around instead of at the speaker when they are speaking.

Mixed Reality Groups

Active Engagement. Active engagement was defined as times when the participant is actively attending to the assigned work and is on task and actively participating. Examples include reading the instructions on tasks together, actively participating in assembly tasks or discussion tasks.

Off-Task Technology Use. Off-task technology was defined as using technology for other purposes rather than to stimulus or activity that is assigned or using technology present not in the way that was directed by the instructor during a lesson. Examples include using controllers when they did not need to, opening social media sites during activity, texting.

Irrelevant Discussion. Irrelevant discussion was defined as discussion that was not relevant to the task at hand. Example: discussing social plans during group discussion, making out-of-task related discussion with peers.

Image 2. (1) Participants in the face-to-face condition participated in a group discussion (top-left); (2) Participants in the video passthrough condition participated in a group discussion (top-right); (3) Participants without headsets engaged in assembling an HTC Vive tracking space (bottom-left); (4) Participants in video passthrough engaged in assembling an HTC Vive tracking space (bottom-right).



4.6 Behavioral Group-Level Measures

Task Completion Time. Task completion time was measured by the amount of time taken by a group to complete the spatial task. A researcher timed the duration of each task during the task performance by groups (Miller et al., 2019).

Number of Messages Exchanged. The number of messages exchanged between group members is coded by the number of words that are exchanged by participants during the interaction (Walther, 1992).

Positive Emotion Language. Positive-emotion language was calculated using the text analysis program, LIWC, which provides the percentage of total positive and negative emotion words within each corpus that fall under psychologically relevant language categories (Pennebaker et al., 2015).

4.7 Self-Reported Measures

Social Presence. Social presence was measured using the Networked Minds Social Presence Inventory (Biocca et al., 2001) using a 7-point Likert scale (1=Not at all, 7 = A very high degree) (McDonald's Omega = 0.86). This scale was adapted to ensure that participants were aware that the question queried on their social presence among the peers in their same room, as opposed to avatars or agents in virtual environments (see Appendix for scale).

Closeness. Closeness was measured using the one-item Inclusion of the Other in the Self scale where participants reported their levels of self-other overlap with their group members by selecting one of seven drawings with different degrees of overlapping circles (1=Not overlapping, 7=Almost fully overlapping) (Aron et al., 1992).

Loneliness. Loneliness was measured using the adapted UCLA Loneliness Scale (Russell, 1996) using a 4-point Likert scale (1 = Never, 4 = Always) (McDonald's Omega = 0.80).

Cognitive Load. Cognitive load was measured using the NASA TLX Cognitive Load Scale (Hart & Steveland, 1988) using a 7-point Likert scale (1 = Very Low and 7 = Very High) (McDonald's Omega = 0.72).

5. Results

To compute our results, the unit of analysis varied depending on the nature of the dependent variable. For individual-level outcomes (e.g., NASA-TLX subscales, Social Presence, and Closeness), we employed Linear Mixed Models (LMMs) with a random intercept for Group ID to account for the nested structure of individuals within groups. This analytical approach was chosen to account for the non-independence of individual observations nested within groups (the unit of assignment). Prior to analysis, we assessed the normality of the data using Shapiro-Wilk tests. Due to non-normal distributions in several dependent variables, we applied the Aligned Rank Transform (ART) to the data before conducting the LMMs. All outliers were winsorized. Pairwise comparisons were performed using Bonferroni adjustments and post-hoc Tukey's HSD tests. As noted in the procedure, no order effects were observed for any outcome variables. In these models, Passthrough Function and Activity Type were treated as fixed effects, while Group ID was included as a random effect to control for group-level variance. To verify the necessity of the multilevel modeling approach, we computed Intraclass Correlation Coefficients (ICC) for each variable across the 24 groups. ICC values ranged from .00 to .14, confirming that up to 14% of the total variance was attributable to group membership, necessitating the nested analysis structure (**Supplemental Material B**). For group-level outcomes where the metric was shared by all members and not measured at the individual level (e.g., task completion time, total word count, and the observational codes for emotional climate and engagement), the group served as the primary unit of analysis (N = 24 groups), we employed Wilcoxon tests and t-tests.

Mixed Reality Groups

Table 2. *Means and Standard Deviations of Self-Reported Measured Variables.*

Mixed Reality Groups

Variables	In Video Passthrough – Means (SD)	Face-to-Face – Means (SD)
Social Presence		
Spatial Activity	4.62 (0.62)	4.56 (0.76)
Discussion Activity	4.79 (0.45)	4.61 (0.70)
Closeness		
Spatial Activity	4.47 (1.39)	4.19 (1.57)
Discussion Activity	4.42 (1.67)	4.13 (1.51)
Loneliness		
Spatial Activity	1.92 (0.43)	1.98 (0.53)
Discussion Activity	1.90 (0.48)	2.04 (0.60)
Cognitive Load		
Spatial Activity	3.41 (0.89)	3.07 (0.87)
Discussion Activity	3.04 (0.84)	3.04 (0.75)
Mental Demand		
Spatial Activity	3.20 (1.44)	2.58 (1.26)
Discussion Activity	3.08 (1.38)	2.90 (1.30)
Physical Demand		
Spatial Activity	2.80 (1.51)	2.17 (1.23)
Discussion Activity	2.08 (1.37)	1.70 (1.19)
Temporal Demand		
Spatial Activity	2.59 (1.22)	3.06 (1.51)
Discussion Activity	2.18 (1.16)	2.89 (1.53)
Performance		

Mixed Reality Groups

Spatial Activity	5.79 (1.36)	5.55 (1.80)
Discussion Activity	5.02 (1.68)	5.34 (1.44)
Frustration		
Spatial Activity	2.53 (1.49)	2.02 (1.22)
Discussion Activity	2.42 (1.53)	2.23 (1.27)
Effort		
Spatial Activity	3.53 (1.57)	3.10 (1.51)
Discussion Activity	3.41 (1.61)	3.12 (1.39)

Table 3. Means and Standard Deviations of Behavioral and Linguistic Variables.

Variables	In Video Passthrough – Means (SD)	Face-to-Face – Means (SD)
Task Completion (seconds)		
Spatial Activity	558.83 (27.34)	378.42 (41.40)
Discussion Activity	608.75 (72.82)	574.33 (22.48)
Linguistic Indicators		
Word Count		
Spatial Activity	338.00 (103.09)	488.83 (132.24)
Discussion Activity	688.21 (291.19)	1228.50 (297.47)
Negative Emotion		
Spatial Activity	0.05 (0.11)	0.15 (0.23)
Discussion Activity	0.25 (0.22)	0.31 (0.19)
Positive Emotion Language through LIWC		
Spatial Activity	0.40 (0.46)	0.74 (0.39)
Discussion Activity	0.39 (0.26)	0.42 (0.24)

Table 4. Means and Standard Deviations of Behavioral Variables with Coding Scheme.

Mixed Reality Groups

Variables	In Video Passthrough – Means (SD)	Face-to-Face – Means (SD)
Passive Response		
Spatial Activity	5.67 (1.56)	6.50 (1.00)
Discussion Activity	5.00 (1.65)	6.50 (0.80)
Active Response		
Spatial Activity	5.94 (1.13)	4.92 (0.72)
Discussion Activity	5.75 (1.04)	5.57 (0.53)
Off-Task Use of Technology		
Spatial Activity	2.67 (1.56)	1.25 (0.45)
Discussion Activity	2.75 (1.42)	1.25 (0.45)
Positive Climate		
Spatial Activity	6.18 (0.83)	6.60 (0.50)
Discussion Activity	6.02 (0.74)	6.67 (0.49)
Negative Climate		
Spatial Activity	1.03 (0.08)	1.00 (0.00)
Discussion Activity	1.05 (0.09)	1.00 (0.00)
Sensitivity		
Spatial Activity	4.75 (1.48)	4.92 (1.38)
Discussion Activity	5.91 (1.31)	6.75 (0.45)
Irrelevant Discussion		
Spatial Activity	1.18 (0.27)	1.24 (0.29)
Discussion Activity	1.00 (0.00)	1.00 (0.00)
Productivity		
Spatial Activity	6.20 (0.83)	6.59 (0.50)
Discussion Activity	6.35 (0.77)	6.83 (0.39)

Mixed Reality Groups

Bolded findings illustrate significant differences with significant levels detailed under each category in Results.

All descriptives and correlations between dependent variables tested in the study are present in

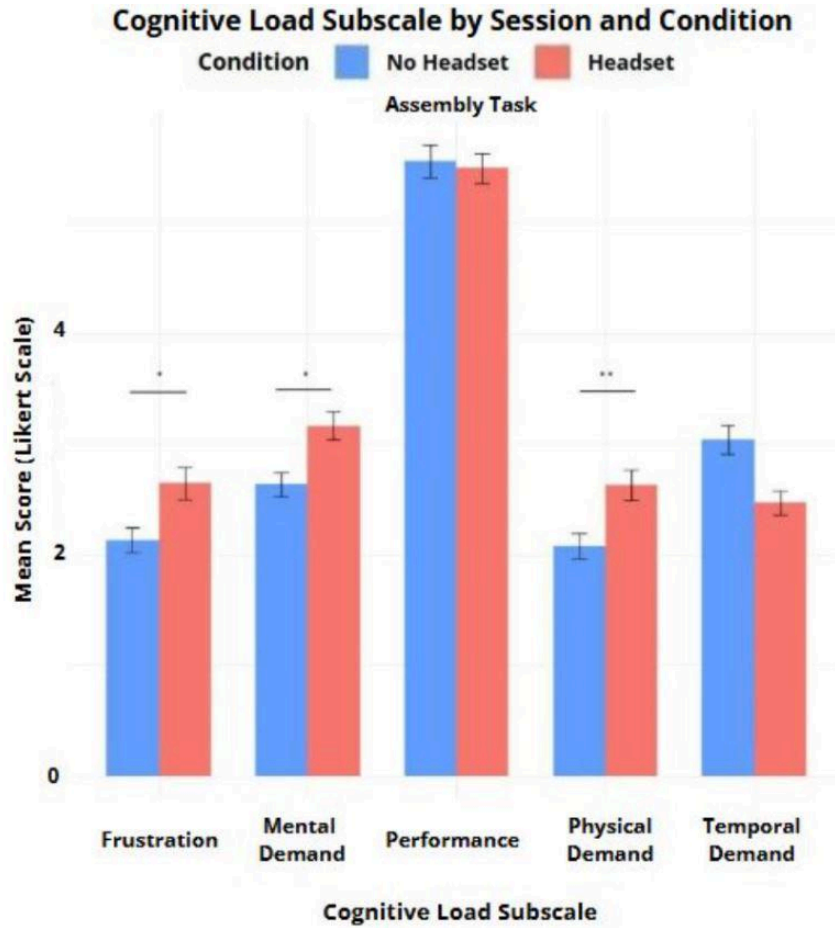
Supplemental Material C, D, E and F.

Cognitive Load

While results illustrate no significant findings in cognitive load as a composite factor, assessing the emotional and cognitive processing variables associated with cognitive load, we found differing effects of cognitive load on video passthrough use and tasks, namely, (1) mental demand, (2) physical demand, (3) temporal demand, (4) frustration, and (5) performance. Results revealed a significant main effect of condition (Headset vs. FtF), $F(1, 33.12) = 7.16, p = .012$. There was also a significant interaction between condition and activity type, $F(1, 292.42) = 10.27, p = .002$ (Table 2). A similar pattern was observed for frustration. The LMM indicated a significant main effect of Condition, $F(1, 29.51) = 4.71, p = .038$, with headset users reporting higher frustration levels than face-to-face participants. However, unlike the other demand dimensions, frustration did not vary significantly by activity type ($p = 0.81$), nor was there a significant interaction ($p = .34$).

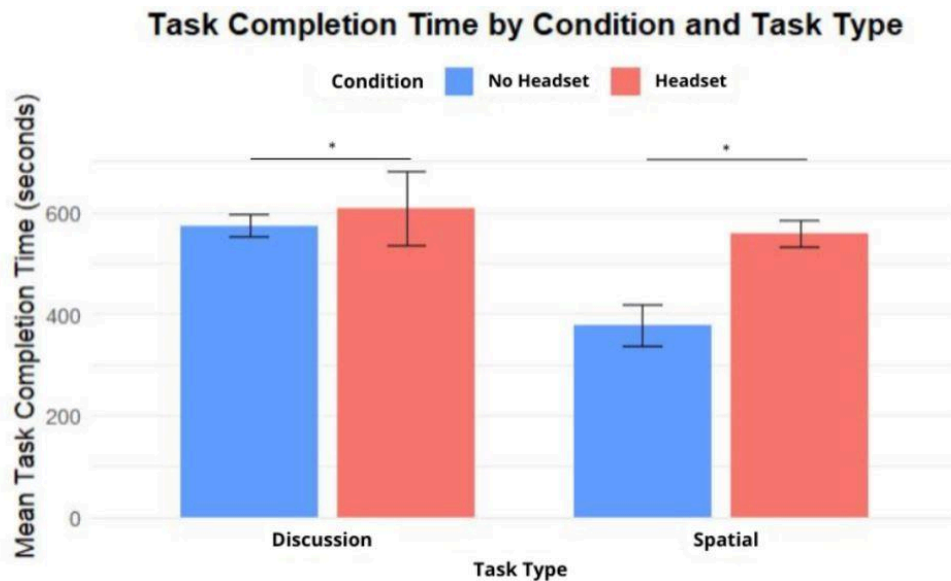
Beyond mental demand, linear mixed-effects models revealed significant main effects of Condition for both physical demand, $F(1, 37.84) = 9.90, p = 0.003$, and temporal demand, $F(1, 40.80) = 4.68, p = 0.036$. In both instances, participants in the video passthrough condition reported higher levels of demand than those in the face-to-face condition. Additionally, a main effect of Activity Type was found for both physical ($p < 0.001$) and temporal ($p < 0.001$) dimensions, indicating that the spatial assembly task was inherently more taxing in these areas than the group discussion (**Figure 1, Table 2**).

Figure 1. *Cognitive Load Subscale Values by Activity and Passthrough Condition.*



A mixed-factorial ANOVA revealed a significant main effect of Condition, $F(1, 22) = 5.16$, $p = 0.033$, indicating that participants were generally faster without a headset than with one. There was also a significant main effect of Activity Type, $F(1, 22) = 13.01$, $p = 0.002$, as groups completed the discussion task more quickly than the assembly task (**Figure 2**).

Figure 2. Task Completion Time by Activity and Video Passthrough Condition.



Social Presence and Emotional Language

The behavioral measure of the number of messages exchanged also differed significantly, with participants in the video passthrough condition having exchanged less words in the spatial activity than in the discussion activity ($W = 416, p = 0.01$) (**Table 3**).

Moreover, on analyzing results from our linguistic behavioral measure through LIWC for positive emotion language, our results found that participants had more positive emotional words in the face-to-face condition compared to the video passthrough condition ($W = 392, p = 0.03$) (**Table 3**).

Passive and Active Engagement

Results from our linear mixed effects model on passive engagement illustrated that compared to the video passthrough condition, the face-to-face condition had more passive engagement across activities ($t(24) = 3.14, p = 0.004$) (**Table 4**). Number of group members significantly decreased passive engagement where more group members resulted in higher passive engagement ($b = -0.49, SE = 0.19, df(24) = -2.56, p = 0.02$).

Off-Task Technology Use

Mixed Reality Groups

Results from our linear mixed effects model on off-task technology use illustrated that compared to the video passthrough condition, the face-to-face condition had more off-task technology use ($t(24) = -6.43, p = 0.001$) across conditions (**Table 4**). Number of group members significantly increased off-task technology use where more group members resulted in higher off-task technology use ($b = -0.59, SE = 0.12, df(24) = 5.11, p < 0.001$).

Sensitivity

Results on sensitivity illustrated that compared to the video passthrough condition, the face-to-face condition had greater awareness of their peers in the spatial activity and the discussion ($t(24) = 4.21, p = 0.001$) (**Table 4**).

Positive and Negative Climate

Results on positive climate illustrated that compared to the video passthrough condition, the face-to-face condition had no significant difference on positive climate. In terms of negative climate, our results illustrated that participants in the video passthrough condition had no significant differences in negative climate than in the face-to-face condition. A higher number of group members resulted in significantly lower positive climate ($b = -0.29, SE = 0.13, df(24) = -2.14, p = 0.04$) and significantly increased negative climate ($b = 0.19, SE = 0.08, df(24) = 2.11, p = 0.045$).

Irrelevant Discussion and Productivity

No significant differences in the video passthrough condition were found for both irrelevant discussion and productivity. However, significant differences were found in activity type for both irrelevant discussion and productivity wherein participants had higher irrelevant discussion in the assembly task ($F(1, 29) = 1.06, p = 0.02$) and higher levels of productivity in the discussion ($F(1, 23) = -0.92, p = 0.005$) (**Table 4**). The number of group members

contributed to the amount of irrelevant discussion with higher number of members significantly increasing irrelevant discussion ($b = 0.45$, $SE = 0.21$, $df(17.31) = 2.20$, $p = 0.04$).

6. Discussion

6.1 Impact of Passthrough and Everyday Tasks

This study extends prior work on social VR group behavior in MR by evaluating the effects of video passthrough on group communication through the lens of fundamental technological dimensions (Walther, 1992; Guo et al., 2019). The current results illustrate that in the video passthrough, groups of student participants experience higher cognitive load in video passthrough compared to face-to-face interactions, particularly in assembly tasks, supporting our theoretical framework integrating cognitive load theory and social information processing theory. However, the types of load they experience in video passthrough differ in the video passthrough compared to the face-to-face condition. Participants in video passthrough felt higher mental demand compared to those having to complete the task face-to-face. Participants in video passthrough were also seen adjusting their headsets frequently, suggesting that discomfort from the headset and wearing one resulted in higher cognitive load as found in prior work, which could be due to accommodation and binocular convergence (Bailenson et al., 2024; Santoso & Bailenson et al., 2024; Chan et al., 2022).

Although the assembly task may appear distinct from decision-making discussions, including both task types allowed us to demonstrate that media effects are not limited to verbal deliberation but also extend to embodied, spatial coordination, an increasingly common component of mixed reality collaboration (Brument et al., 2025).

Task completion patterns further support our theoretical integration. Spatial assembly tasks were completed more slowly in the MR condition. In addition, discussion durations were

Mixed Reality Groups

longer in the headset condition, both of which can also be explained by cognitive load theory, which posits that perceptual and attentional demands can interfere with processing efficiency and group coordination (Kim et al., 2025). Moreover, the technological affordances of video passthrough, including latency, depth misperception and distortions may interfere with natural spatial processing, resulting in mismatched physical and visual feedback, causing users to take a longer time orienting themselves to manipulate objects (Westermeier et al., 2025). Nevertheless, with the use of well-designed, spatially integrated MR overlays, better head and hand tracking, as well as the use of MR guides over real human instruction, task performance may be increased (Wiethuchter et al., 2024; Zaldivar-Colado et al., 2017; Cheng et al., 2024).

Speech behavior also varied across conditions. Participants in video passthrough engaged in shorter conversations with fewer positive emotion language words as measured through LIWC, consistent with our hypothesis that increased cognitive burden may reduce conversational richness and emotional expression. This pattern suggests a shift toward more utilitarian communication in passthrough contexts, where limited cognitive resources are prioritized for task completion rather than social engagement. These findings are consistent with the plausible account that when perceptual and attentional resources are taxed, fewer resources remain available for social engagement.

The observed effects of increased cognitive load and reduced behavioral coordination are likely intertwined with the physical and optical constraints of current-generation video passthrough headsets. Despite advancements in visual fidelity and low latency, inherent limitations in the passthrough system, such as a restricted field of view and residual system latency, impact the fidelity of social cues (Jerald, 2015). A reduced field of view diminishes peripheral awareness, forcing participants to exert extra mental demand to monitor the group

Mixed Reality Groups

environment, which in turn explains the reduced indicators of group awareness and the displacement of attention toward off-task behavior (Billinghurst et al., 2015). This suggests that the current findings reflect the limitations of the medium's implementation rather than a fundamental flaw in the concept of MR, pointing to future hardware improvements as a crucial factor in mitigating these friction points and enhancing collaborative efficacy.

Finally, while prior work has illustrated social presence in video passthrough and found lower social presence in video passthrough than face-to-face (Bailenson et al., 2024; Santoso & Bailenson, 2024), our work illustrates non-significant results as it relates to social presence, and closeness, perhaps due to a ceiling effect. Similarly, we found floor effects for loneliness. This aligns with prior work on social interaction in augmented reality, wherein dyads who did not know one another did not exhibit less closeness when interacting with an individual with a headset than without one (Miller et al., 2019). Moreover, we recognize that prior work has found social presence between a researcher and a subject that did not know one another (Santoso & Bailenson, 2024), where the concept of civil inattention may come into play (Goffman, 1971), the groups of participants in this study have interacted with each other at least once before and anticipate having to interact with one another again, different social-psychological relationships come into play. Additionally, Walther's social information processing theory (SIPT) further suggests that relational tone in mediated communication converges with face-to-face interaction over time, provided sufficient message exchange (Walther, 1992). Thus, while MR impairs behavioral indicators of connection, subjective measures of social presence may remain stable in familiar groups as seen by the dissociation between sensitivity scores of participants. However, it can also be posited that while behavioral indicators showed reduced word count and lower awareness of group members in MR, self-reported measures revealed no condition differences in

loneliness or closeness. However, we see that for cognitive load self-reported measures and task completion time, both behavioral and self-reported measures align. This behavioral-perceptual dissociation highlights an important insight: while the MR medium may support a stable internal sense of subjective closeness in familiar groups, the cognitive costs of the technology manifest as measurable declines in verbal engagement and group awareness.

6.2 Limitations and Future Work

This study is limited by a lack of precise control over participants' prior social familiarity and interaction history. While participants had previous group contact, the degree of existing relationship strength was not systematically measured. Future work should examine the role of interpersonal history and anticipated future interaction in moderating MR communication effects, as this could influence social presence as cited through impression management theory (Goffman, 1971; Sanaria, 2016; Baumeister, 1987). Another avenue of future research could also look into modeling when participants begin to exhibit similar relational tone in CMC as they would in face-to-face interactions.

Another limitation concerns the sampling method. Participants were university students enrolled in a VR course, who likely possessed above-average motivation to engage with immersive technology and greater familiarity with mixed reality systems compared to the general population. This may have introduced a selection bias that limits the external validity of the findings, as individuals less experienced or less enthusiastic about technology might respond differently to passthrough MR environments. Future research should aim to replicate these findings in more diverse populations and professional contexts to assess whether similar cognitive and social effects emerge among users with varying levels of technological experience and motivation.

Mixed Reality Groups

Additionally, although self-reported measures showed no condition differences in linguistic and behavioral indicators revealed reduced word count and lower awareness of group members in MR, echoing prior work (Due et al., 2015; Koelle et al., 2015). These findings suggest that behavioral metrics may be more sensitive to subtle shifts in social investment and engagement. Future studies could leverage and build on our behavioral coding scheme of MR behaviors in everyday tasks and also explore whether frustration mediates affective expression in mediated communication environments and assess the validity of linguistic proxies in interpreting group rapport.

Another avenue for future work involves examining the temporal dynamics of relational development in MR, particularly as it relates to interpersonal variables such as trust in teams, which have been shown to affect face-to-face interactions (Webber, 2008). Longitudinal designs may reveal whether extended exposure mitigates the initial cognitive and affective costs observed in video passthrough environments. Moreover, we acknowledge that we are generalizing about MR based on experiences with the Meta Quest 3. Device-specific features such as field of view, resolution, and passthrough quality may influence user behavior and perception, and findings may not generalize to other MR systems. However, by focusing on the dimensions of richness and interactivity, we provide a roadmap for future hardware evaluations for groups. Future work should investigate whether increasing sensory richness (e.g., expanding FoV to 180°+) or interactivity (e.g., reducing latency to <10ms) changes communication dynamics in groups.

Moreover, while our analysis drew primarily from Social Information Processing Theory and Cognitive Load Theory, future research could examine how groups adapt to MR over longer time horizons. Such extensions would build on, rather than replace, the theoretical relationship

articulated here. As mixed reality tools become more integrated into ongoing organizational and educational settings, groups with shared histories may develop stable patterns of technological adaptation, potentially mitigating some of the cognitive and social costs observed here.

Examining how group history and culture interact with MR affordances would be a valuable direction for future research.

7. Conclusion

This work provides empirical evidence that video passthrough alters group behavior across cognitive and affective dimensions. As organizations and educational institutions increasingly explore MR for collaborative work and learning, these insights become crucial for developing systems and protocols that support, rather than hinder, effective group interaction. These findings underscore the need for further research into MR systems that are not only socially attuned to support subjective presence but also cognitively supportive to maintain high-fidelity behavioral coordination.

8. Author Bios

Monique Santoso is a PhD candidate in Communication at Stanford University.

Portia Wang is a PhD candidate in Communication at Stanford University.

Eugy Han is an Assistant Professor in the Department of Media Production, Management, and Technology (in the College of Journalism & Communication) at the University of Florida.

Jeremy Bailenson is founding director of Stanford University's Virtual Human Interaction Lab, Thomas More Storke Professor in the Department of Communication, Professor (by courtesy) of Education, a Senior Fellow at the Woods Institute for the Environment, and a Faculty Leader at Stanford's Center for Longevity.

References

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of Effective Teacher-Student Interactions in Secondary School Classrooms: Predicting Student Achievement With the Classroom Assessment Scoring System-Secondary. *School psychology review, 42*(1), 76–98.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596–612. <https://doi.org/10.1037/0022-3514.63.4.596>

Bailenson, J. N., Beams, B., Brown, J., DeVeaux, C., Han, E., Queiroz, A. C. M., Ratan, R., Santoso, M., Srirangarajan, T., Tao, Y., & Wang, P. (2024). Seeing the World Through Digital Prisms: Psychological Implications of Passthrough Video Usage in Mixed Reality. *Technology, Mind, and Behavior, 5*(2: Summer 2024). <https://doi.org/10.1037/tmb0000129>

Banquero, M., Valdeolivas, G., Ramón, D., & Juan, M. C. (2024). A color Passthrough mixed reality application for learning piano. *Virtual Reality, 28*(2), 67. <https://doi.org/10.1007/s10055-024-00953-w>

Baumeister, J., Ssin, S. Y., ElSayed, N. A., Dorrian, J., Webb, D. P., Walsh, J. A., ... & Thomas, B. H. (2017). Cognitive cost of using augmented reality displays. *IEEE transactions on visualization and computer graphics, 23*(11), 2378-2388. <https://doi.org/10.1109/tvcg.2017.2735098>

Baumeister, R. F., & Hutton, D. G. (1987). Self-presentation theory: Self-construction and audience pleasing. In *Theories of group behavior* (pp. 71-87). New York, NY: Springer New York.

Mixed Reality Groups

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529.

<https://doi.org/10.1037/0033-2909.117.3.497>

Billingham, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. *Foundations and Trends® in Human-Computer Interaction*, *8*(2-3), 73-272. <https://doi.org/10.1561/11000000049>

Biocca, F., Harms, C., & Gregg, J. (2001, May). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA* (pp. 1-9).

Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning and Instruction*, *3*(1), 1-21. [https://doi.org/10.1016/s0959-4752\(09\)80002-9](https://doi.org/10.1016/s0959-4752(09)80002-9)

Brooke, M. E., & Ng, S. H. (1986). Language and social influence in small conversational groups. *Journal of Language and Social Psychology*, *5*(3), 201-210.

Buck, R., Losow, J. I., Murphy, M. M., & Costanzo, P. (1992). Social facilitation and inhibition of emotional expression and communication. *Journal of Personality and Social Psychology*, *63*(6), 962.

Burgoon, J. K., Bonito, J. A., Ramirez Jr, A., Dunbar, N. E., Kam, K., & Fischer, J. (2002). Testing the interactivity principle: Effects of mediation, propinquity, and verbal and nonverbal modalities in interpersonal interaction. *Journal of communication*, *52*(3), 657-677.

Brument, H., De Pace, F., & Podkosova, I. (2025). Does mixed reality influence joint action? Impact of the mixed reality setup on users' behavior and spatial interaction. *Journal on Multimodal User Interfaces*, *19*(2), 119-138.

Mixed Reality Groups

Carlson, J. R., & Zmud, R. W. (1999). Channel expansion theory and the experiential nature of media richness perceptions. *Academy of management journal*, 42(2), 153-170.

Chan, T. T., Wang, Y., So, R. H. Y., & Jia, J. (2022). Predicting subjective discomfort associated with lens distortion in VR headsets during vestibulo-ocular response to VR scenes. *IEEE Transactions on Visualization and Computer Graphics*, 29(8), 3656-3669.

Cheng, L., Schachtler, D., Schreiner, M., & Kunz, A. (2024, October). Task Guidance in Laser Alignment Experiments using Video See-Through Head-Mounted Display. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 431-434). IEEE.

Doughty, M., Ghugre, N. R., & Wright, G. A. (2022). Augmenting performance: A systematic review of optical see-through head-mounted displays in surgery. *Journal of Imaging*, 8(7), 203. <https://doi.org/10.3390/jimaging8070203>

Due, B. L. (2015). The social construction of a Glasshole: Google Glass and multiactivity in social interaction. *PsychNology Journal*, 13(2).

Goffman, E. (2017). *Relations in public: Microstudies of the public order*. Routledge.

Guo, Z., Wang, H., Deng, H., Xu, W., Baghaei, N., Lo, C. H., & Liang, H. N. (2024). Breaking the Isolation: Exploring the Impact of Passthrough in Shared Spaces on Player Performance and Experience in VR Exergames. *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/tvcg.2024.3372114>

Mixed Reality Groups

Hart SG, & Staveland LE (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183).

Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)

Harrigan, J. A., & Steffen, J. J. (1983). Gaze as a turn-exchange signal in group conversations. *British Journal of Social Psychology*.

Hiltz, S. R., Johnson, K., & Turoff, M. (1986). Experiments in group decision making communication process and outcome in face-to-face versus computerized conferences. *Human communication research*, 13(2), 225-252. <https://doi.org/10.1111/j.1468-2958.1986.tb00104.x>

He, S., Jia, Y., Sun, Z., Yu, C., Yi, X., Shi, Y., & Xu, Y. (2019). AR assistive system in domestic environment using HMDs: Comparing visual and aural instructions. In *Virtual, Augmented and Mixed Reality. Multimodal Interaction: 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21* (pp. 71-83). Springer International Publishing.

Holt, S. (2023). Virtual reality, augmented reality and mixed reality: For astronaut mental health; and space tourism, education and outreach. *Acta Astronautica*, 203, 436-446.

Huang, S. A., & Hancock, J. T. (2022). Will you go on a date with me? Predicting first dates from linguistic traces in online dating messages. *Journal of Language and Social Psychology*, 41(4), 371-395. <https://doi.org/10.1177/0261927x211066612>

Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

Mixed Reality Groups

Jokinen, K., Furukawa, H., Nishida, M., & Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems (TiIS)*, 3(2), 1-30.

Jones, C., Volet, S., & Pino-Pasternak, D. (2021). Observational research in face-to-face small groupwork: Capturing affect as socio-dynamic interpersonal phenomena. *Small Group Research*, 52(3), 341-376.

Kalma, A. (1992). Gazing in triads: A powerful signal in floor apportionment. *British Journal of Social Psychology*, 31(1), 21-39.

Kantharaju, R. B., & Pelachaud, C. (2021, September). Social signals of cohesion in multi-party interactions. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (pp. 9-16).

Kim, D., Min, Y., Jeong, J. Y., Han, S., Hwang, J., & Jeong, J. W. (2025, April). "Through the Looking Glass, and What We Found There": A Comprehensive Study of User Experiences with Pass-Through Devices in Everyday Activities. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-30).

Kockord, R., & Bodensiek, O. (2021). Cognitive load during first contact with mixed reality learning environments. In *Proceedings of mensch und computer 2021* (pp. 260-264).

Koelle, M., Kranz, M., & Möller, A. (2015, August). Don't look at me that way! Understanding user attitudes towards data glasses usage. In *Proceedings of the 17th international conference on human-computer interaction with mobile devices and services* (pp. 362-372).

<https://doi.org/10.1145/2785830.2785842>

Mixed Reality Groups

Koudenburg, N., Postmes, T., & Gordijn, E. H. (2014). Conversational flow and entitativity: the role of status. *The British journal of social psychology, 53*(2), 350–366.

<https://doi.org/10.1111/bjso.12027>

Lesko, W. A., & Schneider, F. W. (1978). Effects of speaking order and speaker gaze level on interpersonal gaze in a triad. *The Journal of Social Psychology, 104*(2), 185-195.

Maran, T., Furtner, M., Liegl, S., Ravet-Brown, T., Haraped, L., & Sachse, P. (2021). Visual attention in real-world conversation: Gaze patterns are modulated by communication and group size. *Applied Psychology, 70*(4), 1602-1627.

McKinlay, A., Procter, R., Masting, O., Woodburn, R., & Arnott, J. (1994). Studies of turn-taking in computer mediated communications. *Interacting with computers, 6*(2), 151-171.

Michalski, C., Cowan, M., Bohinsky, J., Dickerson, R., & Plochocki, J. H. (2023). Evaluation of cognitive load for a mixed reality anatomy application. *Translational Research in Anatomy, 31*, 100247.

Miller, M. R., & Bailenson, J. N. (2021). Social presence outside the augmented reality field of view. *Frontiers in Virtual Reality, 2*, 656473. <https://doi.org/10.3389/frvir.2021.656473>

Miller, M. R., Jun, H., Herrera, F., Yu Villa, J., Welch, G., & Bailenson, J. N. (2019). Social interaction in augmented reality. *PloS one, 14*(5), e0216290.

<https://doi.org/10.1371/journal.pone.0216290>

Neufeld, D. J., Roghanizad, M. M., & White, R. E. (2025). The Impact of Video-Mediated Communication on Social Predictions and Theory of Mind Activation. *International Journal of Human-Computer Interaction, 1-14*.

Mixed Reality Groups

Pashler H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, 116(2), 220–244. <https://doi.org/10.1037/0033-2909.116.2.220>

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015.

Rolland, J. P., Holloway, R. L., & Fuchs, H. (1995, December). Comparison of optical and video see-through, head-mounted displays. In *Telem manipulator and Telepresence Technologies* (Vol. 2351, pp. 293-307). SPIE.

Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of personality assessment*, 66(1), 20-40.

https://doi.org/10.1207/s15327752jpa6601_2

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *language*, 50(4), 696-735.

Salter, C., Thomasson, T., & Uro, P. (2023). Animate: A Theatrical Exploration of Climate Transformation through the Medium of Extended Reality (XR). ISEA2023.

Sanaria, A. D. (2016). A conceptual framework for understanding the impression management strategies used by women in Indian organizations. *South Asian Journal of Human Resources Management*, 3(1), 25-39.

Santoso, M., & Bailenson, J. (2024). How Video Passthrough Headsets Influence Perception of Self and Others. *Cyberpsychology, behavior and social networking*, 10.1089/cyber.2024.0398.

Advance online publication. <https://doi.org/10.1089/cyber.2024.0398>

Mixed Reality Groups

Schier, F., Zeidler, D., Chandran, K., Yu, Z., & McGinity, M. (2023). ViewR: Architectural-Scale Multi-User Mixed Reality with Mobile Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/tvcg.2023.3299781>

Spring, M. B., & Vathanophas, V. (2003). Peripheral social awareness information in collaborative work. *Journal of the American Society for Information Science and Technology*, 54(11), 1006-1013.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.

van Brakel, V., Barreda-Ángeles, M., & Hartmann, T. (2023). Feelings of presence and perceived social support in social virtual reality platforms. *Computers in Human Behavior*, 139, 107523. <https://doi.org/10.1016/j.chb.2022.107523>

Vertegaal, R., Slagter, R., Van der Veer, G., & Nijholt, A. (2001, March). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 301-308).

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication research*, 19(1), 52-90. <https://doi.org/10.1177/009365092019001003>

Walther, J. B. (1992). Relational communication in computer-mediated interaction. *Human Communication Research*. <https://doi.org/10.1111/j.1468-2958.1992.tb00295.x>

Westermeier, F., Murmu, C., Kohm, K., Pagano, C., Wienrich, C., Babu, S. V., & Latoschik, M. E. (2025, March). Interpupillary to Inter-Camera Distance of Video See-Through AR and its

Mixed Reality Groups

Impact on Depth Perception. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)* (pp. 537-547). IEEE.

Webber, S. S. (2008). Development of cognitive and affective trust in teams: A longitudinal study. *Small group research*, *39*(6), 746-769.

Wheeler, L., Reis, H., & Nezlek, J. B. (1983). Loneliness, social interaction, and sex roles. *Journal of Personality and Social Psychology*, *45*(4), 943–953.

<https://doi.org/10.1037/0022-3514.45.4.943>

Wiethüchter, R., Kalloori, S., & Lindlbauer, D. (2024, March). Side-By-Side Might Win: Occlusion Negatively Affects The Performance Of Augmented Reality Task Instructions. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 913-914). IEEE.

Yoon, S. H., Bianchi, A., Seifi, H., Kim, J. R., Vatavu, R. D., Lee, J., & Lee, G. (2025, April). Beyond Glasses: Future Directions for XR Interactions within the Physical World. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

Zaldivar-Colado, U., Garbaya, S., Tamayo-Serrano, P., Zaldivar-Colado, X., & Blazevic, P. (2017, August). A mixed reality for virtual assembly. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 739-744). IEEE.

Zeng, W., Ilo, C., Bowman, D., & Thompson, J. (2024). Virtual Loupes: A Pilot Study on the Use of Video Passthrough Augmented Reality in Plastic Surgery. *Surgical Innovation*, 15533506241265544.

Supplemental Material A. Number of Participants Per Group.

Section Number	Number of Participants	Type of Group
1	6	Mixed Reality Passthrough Condition
2	7	Mixed Reality Passthrough Condition
3	10	Mixed Reality Passthrough Condition
4	8	Mixed Reality Passthrough Condition
5	9	Mixed Reality Passthrough Condition
6	7	Mixed Reality Passthrough Condition
7	7	Mixed Reality Passthrough Condition
8	6	Mixed Reality Passthrough Condition
9	7	Mixed Reality Passthrough Condition
10	9	Mixed Reality Passthrough Condition
11	7	Mixed Reality Passthrough Condition
12	6	Mixed Reality Passthrough Condition
13	7	No Mixed Reality Passthrough Condition
14	7	No Mixed Reality Passthrough Condition
15	8	No Mixed Reality Passthrough Condition
16	9	No Mixed Reality Passthrough Condition
17	8	No Mixed Reality Passthrough Condition
18	8	No Mixed Reality Passthrough Condition
19	8	No Mixed Reality Passthrough Condition
20	7	No Mixed Reality Passthrough Condition
21	9	No Mixed Reality Passthrough Condition
22	9	No Mixed Reality Passthrough Condition
23	7	No Mixed Reality Passthrough Condition
24	6	No Mixed Reality Passthrough Condition

Supplemental Material B. Interclass Correlations for Self-Reported Measures.

Measure	ICC	CI [95% Lower Bound]	CI [95% Upper Bound]	Between Group Variance	Within Group Variance	Number of Groups	Number of Observations
Discussion - Social Presence	0.14	0.02	0.32	0.12	0.73	24	160
Spatial Activity - Social Presence	0.09	0.00	0.27	0.09	0.85	24	161
Discussion - Loneliness	0	0.00	0.13	0	0.12	24	160
Spatial Activity - Loneliness	0	0.00	0.13	0	0.16	24	161
Discussion - Mental Demand	0	0.00	0.13	0	1.73	24	160
Spatial Activity - Mental Demand	0.09	0.00	0.27	0.18	1.8	24	160
Discussion - Physical Demand	0.04	0.00	0.20	0.05	1.04	24	157
Spatial Activity - Physical Demand	0.06	0.00	0.23	0.14	2.02	24	161
Discussion - Temporal Demand	0.19	0.06	0.39	0.32	1.37	24	159
Spatial Activity - Temporal Demand	0.09	0.00	0.26	0.19	2.01	24	160
Discussion - Performance	0.04	0.00	0.20	0.1	2.22	24	160
Spatial Activity - Performance	0.22	0.09	0.42	0.58	2.1	24	160
Discussion - Effort	0.004	0.00	0.14	0.01	2.19	24	159
Spatial Activity - Effort	0.007	0.00	0.14	0.02	2.44	24	160
Discussion - Frustration	0	0.00	0.13	0	1.78	24	160
Spatial Activity - Frustration	0	0.00	0.13	0	2.05	24	161
Discussion - Closeness	0.01	0.00	0.15	0.03	2.47	24	160
Activity - Closeness	0.06	0.00	0.22	0.13	2.1	24	161

Mixed Reality Groups

Supplemental Material C. Descriptives for Individual-Level Variables.

Variable	Mean	SD	Min	Max	Median
Discussion - Social Presence	5.35	0.92	2.75	7.13	5.38
Spatial Activity - Social Presence	5.47	0.96	2.25	7.13	5.63
Discussion - Loneliness	2.38	0.35	1.63	3.63	2.38
Spatial Activity - Loneliness	2.42	0.40	1.00	3.63	2.38
Discussion - Mental Demand	2.91	1.32	1.00	6.00	3.00
Spatial Activity - Mental Demand	2.94	1.40	1.00	6.00	3.00
Discussion - Physical Demand	1.64	1.04	1.00	5.00	1.00
Spatial Activity - Physical Demand	2.68	1.47	1.00	7.00	2.00
Discussion - Temporal Demand	2.43	1.29	1.00	7.00	2.00
Spatial Activity - Temporal Demand	2.92	1.48	1.00	7.00	3.00
Discussion - Performance	5.05	1.52	1.00	7.00	5.00
Spatial Activity - Performance	5.78	1.62	1.00	7.00	6.00
Discussion - Effort	3.22	1.48	1.00	7.00	3.00
Spatial Activity - Effort	3.31	1.57	1.00	7.00	3.00
Discussion - Frustration	2.27	1.34	1.00	7.00	2.00
Spatial Activity - Frustration	2.30	1.43	1.00	7.00	2.00
Discussion - Closeness	4.21	1.58	1.00	7.00	4.00
Spatial Activity - Closeness	4.37	1.49	1.00	7.00	4.00

Supplemental Material D. Descriptives for Group-Level Variables.

Variable	Mean	SD	Min	Max	Median
Passive Response	5.92	1.41	2.00	7.00	6.50
Active Response	5.42	1.15	3.00	7.00	5.00

Mixed Reality Groups

Irrelevant Discussion	1.55	1.31	1.00	6.00	1.00
Off-Task Technology Use	1.98	1.30	1.00	6.00	1.50
Positive Climate	6.19	1.02	3.00	7.00	6.00
Negative Climate	1.21	0.65	1.00	4.00	1.00
Sensitivity	5.58	1.44	2.00	7.00	6.00
Productivity	6.13	1.33	2.00	7.00	7.00

Supplemental Material E. Correlations of Individual-Level Dependent Variables.

	Discussion - Social Presence	Discussion - Loneliness	Discussion - Mental Demand	Discussion - Physical Demand	Discussion - Temporal Demand	Discussion - Performance	Discussion - Effort	Discussion - Frustration	Discussion - Closeness
Discussion - Social Presence	1								
Discussion - Loneliness	-0.10 [-0.25, 0.06]	1							
Discussion - Mental Demand	0.12 [-0.04, 0.27]	0.24** [0.09, 0.38]	1						
Discussion - Physical Demand	-0.13 [-0.28, 0.03]	0.16* [0.00, 0.31]	0.16* [0.00, 0.31]	1					
Discussion - Temporal Demand	0.01 [-0.15, 0.16]	0.08 [-0.08, 0.23]	0.31*** [0.16, 0.44]	0.22** [0.07, 0.37]	1				
Discussion - Performance	0.32*** [0.17, 0.45]	-0.07 [-0.22, 0.09]	0.02 [-0.14, 0.17]	-0.09 [-0.24, 0.07]	0.02 [-0.13, 0.18]	1			
Discussion - Effort	0.13 [-0.03, 0.28]	0.17* [0.02, 0.32]	0.40*** [0.26, 0.53]	0.25** [0.10, 0.39]	0.24** [0.09, 0.38]	0.05 [-0.11, 0.20]	1		
Discussion - Frustration	-0.15 [-0.30, 0.00]	0.32*** [0.17, 0.45]	0.21** [0.06, 0.35]	0.30*** [0.15, 0.44]	0.24** [0.09, 0.38]	-0.20** [-0.35, -0.05]	0.25** [0.10, 0.39]	1	
Discussion - Closeness	0.40*** [0.26, 0.52]	-0.02 [-0.17, 0.14]	0.02 [-0.14, 0.17]	-0.15 [-0.30, 0.01]	-0.05 [-0.21, 0.10]	0.30*** [0.15, 0.43]	0.16* [0.00, 0.31]	-0.21** [-0.35, -0.06]	1

Mixed Reality Groups

	Spatial Activity - Social Presence	Spatial Activity - Loneliness	Spatial Activity - Mental Demand	Spatial Activity - Physical Demand	Spatial Activity - Temporal Demand	Spatial Activity - Performance	Spatial Activity - Effort	Spatial Activity - Frustration	Spatial Activity - Closeness
Spatial Activity - Social Presence	1								
Spatial Activity - Loneliness	-0.18* [-0.32, -0.02]	1							
Spatial Activity - Mental Demand	0.08 [-0.08, 0.23]	0.19* [0.04, 0.34]	1						
Spatial Activity - Physical Demand	0.09 [-0.07, 0.24]	0.07 [-0.09, 0.22]	0.49*** [0.36, 0.60]	1					
Spatial Activity - Temporal Demand	0.04 [-0.11, 0.20]	0.18* [0.03, 0.33]	0.37*** [0.23, 0.50]	0.38*** [0.24, 0.50]	1				
Spatial Activity - Performance	0.46*** [0.32, 0.57]	-0.01 [-0.16, 0.15]	-0.13 [-0.28, 0.02]	0.05 [-0.11, 0.20]	0.06 [-0.10, 0.21]	1			
Spatial Activity - Effort	-0.03 [-0.19, 0.12]	0.06 [-0.10, 0.21]	0.38*** [0.24, 0.51]	0.47*** [0.34, 0.59]	0.41*** [0.27, 0.53]	0.07 [-0.08, 0.23]	1		
Spatial Activity - Frustration	-0.24** [-0.38, -0.08]	0.34*** [0.20, 0.47]	0.41*** [0.27, 0.53]	0.46*** [0.33, 0.58]	0.34*** [0.19, 0.47]	-0.10 [-0.25, 0.06]	0.34*** [0.19, 0.47]	1	
Spatial Activity - Closeness	0.54*** [0.42, 0.64]	-0.02 [-0.18, 0.13]	0.13 [-0.02, 0.28]	0.07 [-0.09, 0.22]	0.05 [-0.10, 0.21]	0.21** [0.06, 0.35]	0.07 [-0.09, 0.22]	-0.25** [-0.39, -0.10]	1

Supplemental Material F. Correlations of Group-Level Dependent Variables.

Variable	Passive Response	Active Response	Irrelevant Discussion	Off-Task Technology Use	Positive Climate	Negative Climate	Sensitivity
Passive Response	1						

Mixed Reality Groups

Active Response	0.58*** [0.28, 0.77]	1					
Irrelevant Discussion	-0.72*** [-0.85, -0.48]	-0.56** [-0.76, -0.25]	1				
Off-Task Technology Use	-0.75*** [-0.85, -0.59]	-0.40* [-0.66, -0.05]	0.46* [0.12, 0.70]	1			
Positive Climate	0.75*** [0.59, 0.85]	0.73*** [0.51, 0.86]	-0.54** [-0.75, -0.23]	-0.62*** [-0.77, -0.41]	1		
Negative Climate	-0.65*** [-0.79, -0.45]	-0.39* [-0.66, -0.04]	0.63*** [0.35, 0.80]	0.71*** [0.54, 0.83]	-0.51*** [-0.69, -0.26]	1	
Sensitivity	0.79*** [0.65, 0.88]	0.60*** [0.31, 0.79]	-0.65*** [-0.82, -0.39]	-0.70*** [-0.82, -0.52]	0.79*** [0.65, 0.88]	-0.59*** [-0.75, -0.36]	1
Productivity	0.61*** [0.39, 0.76]	0.75*** [0.54, 0.87]	-0.78*** [-0.89, -0.59]	-0.50*** [-0.69, -0.26]	0.69*** [0.50, 0.81]	-0.35* [-0.58, -0.07]	0.65*** [0.45, 0.79]