



Article

# MirrorCampus: A Synchronous Hybrid Learning Environment That Supports Spatial Localization of Learners for Facilitating Discussion-Oriented Behaviors

Shota Sawada <sup>1,\*</sup> , SunKyoung Kim <sup>2</sup> , Masakazu Hirokawa <sup>3</sup> and Kenji Suzuki <sup>4</sup>

<sup>1</sup> Empowerment Informatics Program, University of Tsukuba, Tsukuba 3058573, Japan

<sup>2</sup> Institute of Library, Information and Media Science, University of Tsukuba, Tsukuba 3058550, Japan; sunkyoungk@acm.org

<sup>3</sup> Data Science Laboratories, NEC Corporation, Kawasaki 2118666, Japan; hirokawa\_m@ieee.org

<sup>4</sup> Institute of Systems and Information Engineering, University of Tsukuba, Tsukuba 3058573, Japan; kenji@ieee.org

\* Correspondence: sawasho@ai.iit.tsukuba.ac.jp; Tel.: +81-90-2054-9550

**Abstract:** A growing number of higher-education institutions are implementing synchronous hybrid delivery, which provides both online and on-campus learners with simultaneous instruction, especially for facilitating discussions in Active Learning (AL) contexts. However, learners face difficulties in picking up social cues and gaining free access to speaking rights due to the geometrical misalignment of individuals mediated through screens. We assume that the cultivation of discussions is allowed by ensuring the spatial localization of learners similar to that in a physical space. This study aims to design a synchronous hybrid learning environment, called Mirror Campus (MC), suitable for the AL scenario that connects physical and cyberspaces by providing spatial localization of learners. We hypothesize that the MC promotes discussion-oriented behaviors, and eventually enhances applied skills for group tasks, related to discussion, creativity, decision-making, and interdependence. We conducted an experiment with five different groups, where four participants in each group were asked to discuss a given topic for fifteen minutes, and clarified that the occurrences of facing behaviors, intervening, and simultaneous utterances in the MC were significantly increased compared to a conventional video conferencing. In conclusion, this study demonstrated the significance of the spatial localization of learners to facilitate discussion-oriented behaviors such as facing and speech.

**Keywords:** hybrid learning environment; synchronous learning; active learning; spatial localization; learning behavior; facing behavior; speech behavior; rubric assessment; applied skills



**Citation:** Sawada, S.; Kim, S.; Hirokawa, M.; Suzuki, K. MirrorCampus: A Synchronous Hybrid Learning Environment That Supports Spatial Localization of Learners for Facilitating Discussion-Oriented Behaviors. *Multimodal Technol. Interact.* **2024**, *8*, 31. <https://doi.org/10.3390/mti8040031>

Academic Editor: Stephan Schlögl

Received: 15 March 2024

Revised: 6 April 2024

Accepted: 8 April 2024

Published: 11 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As synchronous communication tools become more accessible, a growing number of higher-education institutions have begun to implement synchronous hybrid delivery of learning contents [1]. It allows teachers to provide instruction and accept questions from both online and on-campus students simultaneously in a single learning environment. In most cases, a certain amount of learners attend the course on campus, while certain individuals follow the course remotely from a location of their choice [2]. Raes et al. reviewed 47 studies regarding synchronous hybrid learning and revealed cautious optimism about the new style of learning compared to fully online or fully on-site instruction, with its benefits and challenges identified [3]. From the students' viewpoint, this style of learning not only allows students to participate in classes remotely but also exposes them to a wider range of views and ideas through access to expertise outside the institution.

On the other hand, some studies found that online and on-campus students experienced the class differently in the hybrid synchronous situation [4,5]. This indicates the gap in experience and hence the quality of learning between on-campus students and

remote students. In face-to-face communication, it is known that nonverbal cues, such as body and head posture, gestures, facial expressions, and tone of voice, play important roles in communication among people. In particular, facing behaviors (head, gaze) help individuals understand others' intention of speech [6], and thus contribute to achieving smooth turn-taking in discussion [7]. To achieve this, having spatial localization among individuals (i.e., being aware of others' location) is crucial; however, in the conventional synchronous hybrid learning environment the geometrical relationship between on-campus and online students is not preserved, which forces students to be aware of the orientation and positioning of cameras and speak into microphones [8]. This results in students being less able to pick up on the social cues of the other students to gain free access to speaking rights [8].

This problem has a serious impact on learning, particularly on Active Learning (AL), where interactive engagement fostered through discussions among learners is crucial [9,10]. AL was defined by Freeman as "engaging students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert" [11]. Previous studies have shown that courses designed to incorporate AL significantly increase learners' depth of understanding in comparison with traditional lecture-based or textbook-centered learning, which results in the development of applied skills such as verbal communication, collaboration, and critical thinking [12]. This can be achieved through various pedagogical methods, including cooperative learning, collaborative learning, discovery learning, experiential learning, problem-based learning, and inquiry-based learning. According to a cross-case analysis of the blended synchronous learning, a central emergent theme from student and teacher observations across all the cases was the importance of designing for AL [13]. Also, according to a recent review paper analyzing 33 articles related to blended synchronous learning, insufficient interaction between online learners and classroom learners has been highlighted as a challenge [5]. The paper suggests encouraging frequent interactions and enabling learners to assume active roles as potential solutions. These suggestions are closely tied to the essence of AL. However, previous studies have simply suggested novel blended synchronous learning environments without considering AL [5,14]. Our study, in contrast, focused on AL, emphasizing the facilitation of interactions between learners by promoting facing behaviors. We also explored the application of the rapidly growing method of virtual reality, not just in hybrid conferences. Thus, we hold the view that to fully harness the benefits of AL in synchronous hybrid learning it is important to encourage effective discussions.

This is an experimental study to examine learning behaviors and outcomes in a specific setup with a within-subjects design. This study aims to design a synchronous hybrid learning environment suitable for an AL scenario that connects physical and cyberspaces by providing spatial localization of learners similar to that in a physical space. Providing spatial localization in this study refers to preparing the environment to geometrically align the positional and directional relationships of individuals. We assume that ensuring spatial localization of individuals allows learners in the environment to utilize nonverbal social cues as face-to-face communication, leading to frequent turn-taking during discussions. This is eventually expected to promote dynamic discussions that result in learning outcomes that may be evaluated or assessed by rubrics. This involves assessing students' achievements based on specific criteria and measuring the attainment of each desired skill.

The literature review identified several directions for future research [3]. One direction pertains to examining the impact on students' learning behaviors and outcomes within a specific AL setting. Our research addressed filling these two gaps and providing empirical data to identify meaningful learning behaviors and outcomes.

## 2. Related Works

The relationship between space and human behavior has been discussed in several studies in terms of the facilitation of behaviors by its spatial characteristics. In the context of sociology, Kendon demonstrated the importance of maintaining a basic corporeal unit of

participation known as a facing formation (F-formation) in social interactions by two or more people [15]. Particularly, in the realm of placemaking for community development, Lang explained that the term “sociopetal” was used by Humphrey Osmond to describe spaces that encourage social gatherings, such as park playgrounds, restaurants, and home dining tables [16]. Scott-Webber stated that the sociopetal space that facilitates face-to-face interaction maximizes direct eye contact as crucial for sustaining engagement, and emphasized that understanding this concept helps come up with certain functional solutions that may be applied especially within the context of learning environments [17]. These concepts have been put forward over time, and this perspective underscores the relevance of establishing an optimal environment to effectively cultivate learning behaviors.

The Social Virtual Reality Environment (SVRE) has gained attention as one of the environments aimed at regulating learning behaviors to improve learning outcomes by its spatial characteristics. For instance, one study revealed that one of the characteristics of the SVRE, namely Mozilla Hubs, could boost participation and engagement in online lessons by adding a sense of “truly being” in a class [18]. Another study also demonstrated that providing learners with tools for virtual object construction in the SVRE, specifically Second Life, could enhance constructionist learning experiences [19]. However, there are few studies investigating the impact of spatial localization on learning behaviors and outcomes. Furthermore, the majority of SVRE systems are designed to be used in a space isolated from other users and are not expected to be used for a synchronous hybrid learning environment where multiple individuals are present in the same physical classroom. The mediating factors for deep and meaningful e-learning in SVREs were accumulated and classified in the Blended Model for Deep and Meaningful E-learning in SVREs [20]. The paper stated that the 3D learning environment is one of the mediating factors in the model, and its characteristics are determined by the adoption or absence of teacher perceptions, certain learning theories, philosophies, or pedagogic frameworks. The proposed environment helps to advance 3D learning environments by ensuring the spatial localization of learners in a way that does not disrupt the normal flow of conversations and provides instructors with an alternative environment for conducting discussions in synchronous hybrid learning.

### 3. Research Question and Hypothesis

Our research question is as follows: *Does a synchronous hybrid learning environment for supporting the spatial localization of learners influence the cultivation of discussions in an AL scenario?* This study hypothesizes that the synchronous hybrid learning environment promotes discussion-oriented behaviors such as facing and speech and eventually enhances applied skills across four categories: discussion, creativity, decision-making, and interdependence, as detailed in Section 6.3.

We hold the belief that the preserving spatial localization of learners makes a difference compared to conventional videoconferencing, especially in terms of the angles of facing behaviors. Our previous study explored the impact of learner’s embodiment and movements during discussion, which are represented on an avatar in a social VR space, and clarified that turn-taking in the space occurred more than on a monitor, leading to a better discussion [21]. Ooko et al. proposed a method of judging a user’s conversational engagement based on head pose data [22]. They found that the amplitude of head movement and rotation had a moderate positive correlation with the level of conversational engagement. Based on these studies, a facing behavior is considered a direct indicator of the increased frequency of learners conveying the intentions that they are willing to engage in a conversation.

Additionally, we expect that the increased frequency of the facing behaviors leads to more active engagement in a conversation. Maroni et al. examined the rhythm and the management of classroom interaction as an important constituent of a teaching–learning process, focusing on some specific aspects of turn-taking: overlapping, interruptions, and pauses [23]. This study revealed that pupils’ participation in interaction being delib-

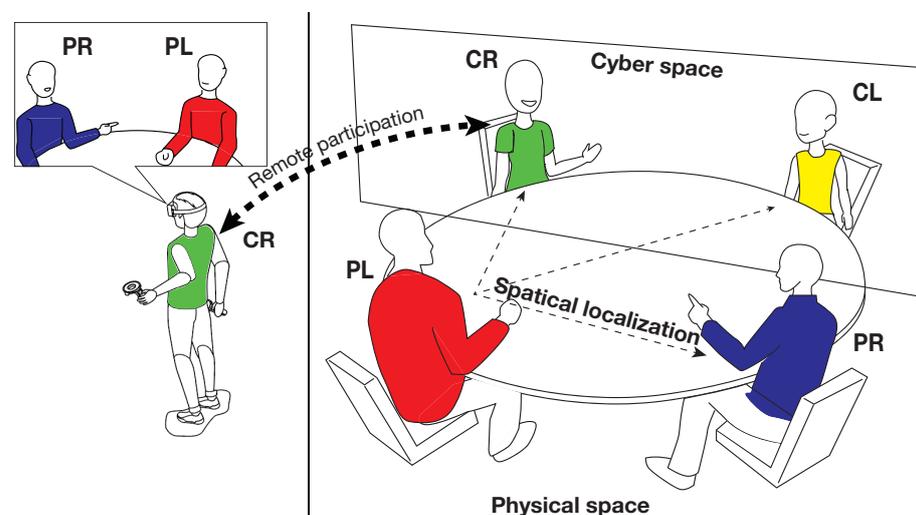
erately encouraged might increase failed interruptions and overlapping. The encouraged behavior was at times interpreted by the teacher as a moment of confusion, making them gain control over the conversation. They also found that pause duration was correlated with the next speaker. Therefore, speech behavior is deemed as an indirect indicator of heightened active engagement in a conversation.

#### 4. Methods

##### 4.1. Overview

We proposed a synchronous hybrid learning environment suitable for the AL scenario, called Mirror Campus (MC), which connects physical and cyberspaces by providing spatial localization of learners similar to that in a physical space. The system requirements are that (1) both local and remote learners can engage in discussions face-to-face, and that the environment (2) is authentically “sociopetal” from the perspective of each learner, where the positional and directional relationships of the individuals are geometrically aligned in a similar manner to those in a physical space, and (3) possesses a fundamental mechanism for detecting specific critical factors of facing and speech behaviors. With the requirements fulfilled, the environment enables the learners to easily discern the others’ intentions or identify who is addressing or attempting to speak to whom and facilitates face-to-face interaction.

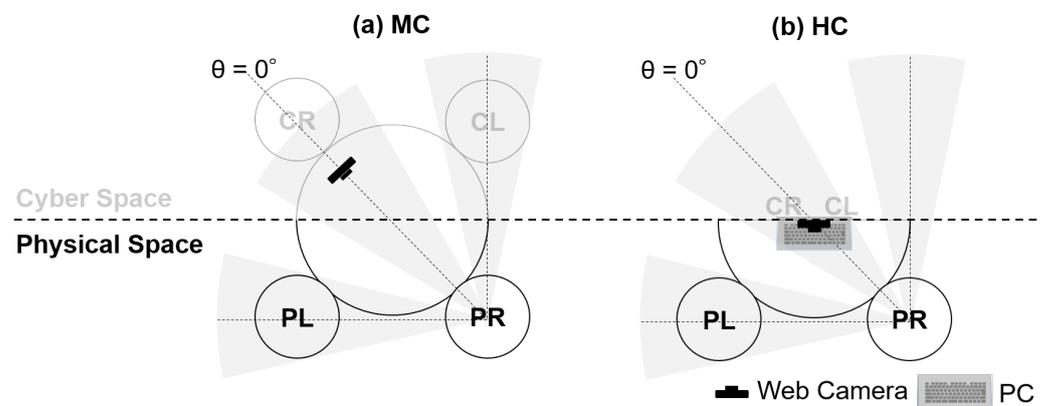
The overview of the proposed system is shown in Figure 1. The synchronous hybrid learning environment was designed as a simple classroom setting, integrating a physical space consisting of a real semicircular table and two real chairs with a cyberspace comprising a virtual semicircular table and two virtual chairs. Learners in remote rooms can access the synchronous hybrid learning environment by equipping a Head-Mounted Display (HMD) on their heads to interact with other learners in a local room. In this figure, a situation is described in which two learners in a room, referred to as “Physical Right (PR)” and “Physical Left (PL)”, are in discussion with two other learners in different rooms, denoted as “Cyber Right (CR)” and “Cyber Left (CL)”. Accordingly, the learners in both spaces can have a discussion while seeing each other in the synchronous hybrid learning environment, which satisfies requirement (1). For requirement (2), we partitioned the synchronous hybrid learning environment by a plane upon which the streaming video from the other space is displayed, referred to as a “mirror” in this study. This mirror can achieve increased spatial localization consistency by aligning the position, posture, and size of learners in the space with those in the other space. For requirement (3), we predefined the factors of facing and speech behaviors detected in the environment, which are described in the following paragraphs.



**Figure 1.** Overview of the synchronous hybrid learning environment. PR, PL, CR and CL stand for “Physical Right”, “Physical Left”, “Cyber Right” and “Cyber Left”, respectively.

#### 4.2. Facing Behavior

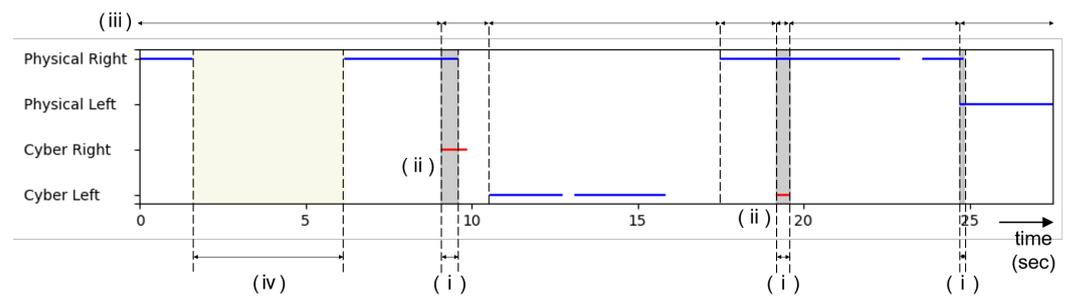
As for the facing behavior, Hachisu et al. defined face-to-face behavior as a physical state in which two people's faces are within  $\pm 20^\circ$  of each other's facing direction because humans can pay attention to things only with eye gaze if they are inside that range, and analyzed three-participant group interactions [24]. We took into consideration that a larger number of participants would shorten the range of the facing behavior and defined it as the state in which one is facing another or the peripheral area within  $\pm 15^\circ$ . The specific threshold setting in the MC is illustrated on the left in Figure 2. For PR, facing behaviors in the range of  $-15^\circ \leq \theta \leq 15^\circ$  were considered to be directed towards CR,  $30^\circ \leq \theta$  towards CL, and  $\theta \leq -30^\circ$  towards PL. Those in the intermediate ranges of  $15^\circ \leq \theta \leq 30^\circ$  and  $-30^\circ \leq \theta \leq -15^\circ$  were regarded as being towards the participant faced until the very last moment. This is because we interpreted exceeding the thresholds as simply averting their gaze from the participant's face during speech.



**Figure 2.** Threshold settings of facing behaviors. The figure indicates specific threshold settings of PR's facing behaviors in two conditions: (a) Mirror Campus (MC) condition and (b) Hybrid Conference (HC) condition. PR's facing behaviors in the range of  $-15^\circ \leq \theta \leq 15^\circ$  were considered to be directed towards CR,  $30^\circ \leq \theta$  towards CL, and  $\theta \leq -30^\circ$  towards PL. Those in the intermediate ranges of  $15^\circ \leq \theta \leq 30^\circ$  and  $-30^\circ \leq \theta \leq -15^\circ$  were regarded as being towards the participant faced until the very last moment.

#### 4.3. Speech Behavior

As for the speech behavior, we regarded a natural unit of speech bounded by breaths or pauses as an utterance and measured four aspects of the speech behavior: simultaneous utterance, intervening utterance, turn-taking, and silent time. For instance, Figure 3 presents the speech behavior in a four-participant group discussion. The horizontal lines in the figure represent the speech intervals of each participant. Simultaneous utterance was defined as an overlap of utterances, as shown in (i). In cases of overlapping with three participants, the simultaneous utterance was counted twice. In (ii), intervening utterances consisted of relatively short utterances that lacked substantial contents. These included back-channelings, repetitions of previous utterances, simple clarifications, laughter, and anticipated failures where a participant attempted to speak but was timed poorly, causing an overlap with the current speaker's utterance, followed by the participant giving up speaking. The concept of turn-taking has been defined in diverse ways in previous studies. In our case, we established the definition of a turn as the duration initiating when one participant began speaking until the person stopped and another participant began speaking. In the example in (iii), the order of participants holding a turn was PR, CR, CL, PR, CL, PR, and PL, resulting in seven turn-takings. Silent time, depicted in (iv), represented the duration in which all participants remained silent for at least five seconds, indicating passivity in the conversation. Silence lasting less than five seconds was excluded since it was considered a waiting period for the next speaker.

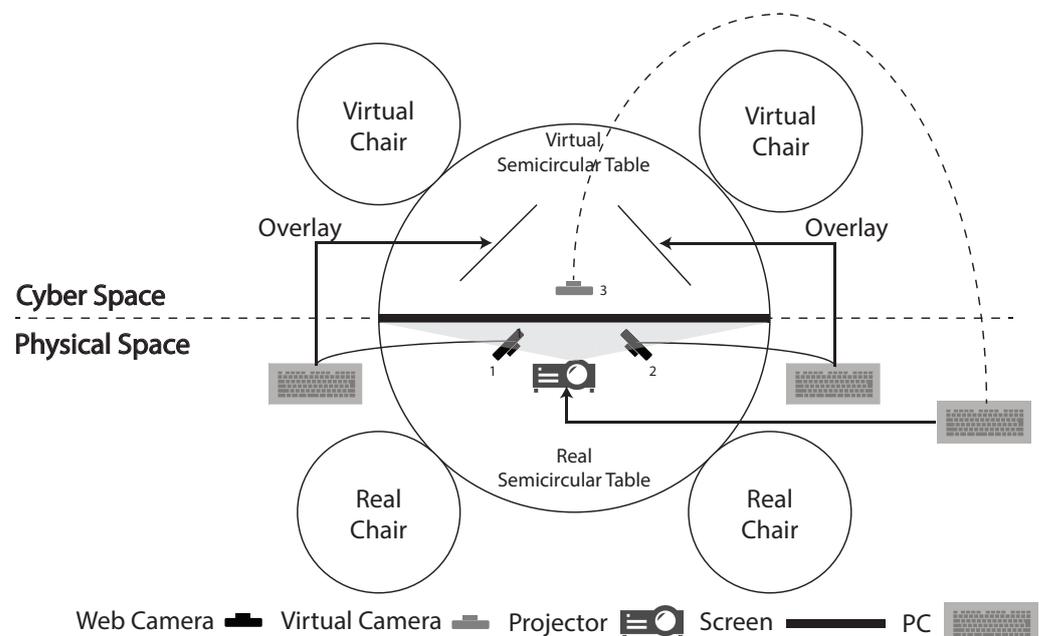


**Figure 3.** Indicators of utterance. The figure presents four indicators of utterances: (i) simultaneous utterance, (ii) intervening utterance, (iii) turn-taking, and (iv) silent time. The horizontal lines represent the speech intervals of each participant, where intervening utterances (ii) are highlighted in red, while the other speech intervals are in blue. The order of the participants holding a turn was PR, CR, CL, PR, CL, PR, and PL, resulting in seven turn-takings, as shown in (iii).

### 5. System Configuration

#### 5.1. Synchronous Hybrid Learning Environment from Physical Space

A semicircular table, two chairs, a short focus projector, and two web cameras (No. 1, No. 2) with 150° fisheye lenses are installed in the physical room, as shown in Figure 4. Learners are seated on the chairs arranged around the table. Regarding the visual exchange of the streaming videos, the video captured by one’s front-facing web camera (No. 1, No. 2) within the physical space is displayed on one’s window of a video conferencing platform called Zoom, while the video captured by a Unity camera (No. 3) within the cyberspace is projected on the wall of the room by the projector, serving as the mirror. Regarding verbal exchange of the streaming audios, they exchange all the voices through a social VR platform called VRChat. The audio within the physical space is picked up by the built-in microphone of the PC logged in VRChat, while the audio is received through the built-in speaker of the projector at the center of the room.



**Figure 4.** Deployment of hardware in the synchronous hybrid learning environment.

#### 5.2. Synchronous Hybrid Learning Environment from Cyberspace

The cyberspace was created using VRChat SDK3 [25] and was accessible from a variety of VR devices. It was equipped with a virtual semicircular table and two virtual chairs

similar to those in the physical space, as shown in Figure 4. We created male and female avatars using Ready Player Me [26], which is a tool to easily generate multiple avatars based on photographs. It is important to note that the avatars were embodied avatars [27] that reflected the participants' body movements, facing behaviors, and mouth movements. Two HMDs (Meta Quest 2, Meta, Menlo Park, CA, USA) were prepared for CR and CL to access the cyberspace. A user wore the HMD and was seated on the virtual chair arranged around the virtual table as an avatar. Regarding the visual exchange of the streaming videos, the video captured by the Unity camera (No. 3) positioned in front of the two avatars in the cyberspace was displayed on the screen in the physical space using a projector. On the other hand, for the CR, the application window of Zoom was overlaid onto the cyberspace using XSOverlay [28] after switching the window to the speaker view, while pinning the individual window of PR and making it full-screen, serving as the mirror. Similarly, the CL carried out the same process on the window of PL. Regarding verbal exchange of the streaming audios, they exchange all the voices through VRChat using the microphone and speaker provided with the HMDs.

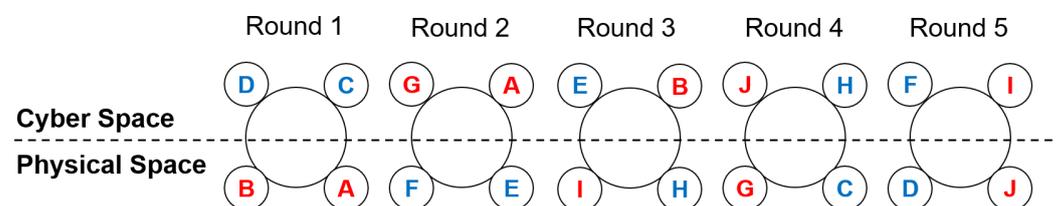
## 6. Experiment

### 6.1. Participants

We recruited ten gender-balanced participants (five males and five females) who met the inclusion criteria of being aged between eighteen and twenty-two years, native Japanese speakers, and having no prior experience with VR sickness. All of the participants had not joined our previous experiment and reported either no experience or infrequent usage of VR devices.

### 6.2. Grouping Design of the Participants

Each of the participants took part in two rounds of the experiment, and in each round four participants as a group were asked to discuss a given topic under three conditions: Full Real (FR), Hybrid Conference (HC), and Mirror Campus (MC), as mentioned in Section 6.4. The participants are shown as A to J in Figure 5, with their roles randomly assigned in each round, and the number of both genders was balanced: two females (indicated as red letters) and two males (indicated as blue letters). Webb et al. found that there is little gender difference in small-group interaction and achievement among high-achieving groups of high-school students compared to low-achieving groups [29]. Thus, gender differences in learning behaviors and outcomes through discussions were deemed negligible among participants at their age. Additionally, it should be noted that each of them experienced both spaces and discussed with completely different participants in the two rounds to avoid the development of social conditions.



**Figure 5.** Combination of participants in each round. The names (letters) of female participants are highlighted in red, while those of male participants are in blue.

### 6.3. Discussion Topics

Several taxonomies of group activities have been developed [30]. In this study, we adopted McGrath's group task circumplex model to select the discussion topics [31]. This model considers the entire process of group activities, including the required performance for individuals cognitively and behaviorally, as well as the range of group interdependence from collaboration to conflict. McGrath's model divides all group activities into eight types based on the achievement goals of a group task: (1) planning tasks (generating plans),

(2) creativity tasks (generating ideas), (3) intellectual tasks (solving problems with correct answers), (4) decision-making tasks (deciding issues without right answers), (5) cognitive conflict tasks (resolving conflicts of viewpoints), (6) mixed-motive tasks (resolving conflicts of interest), (7) competitive tasks (resolving conflicts of power), and (8) psycho-motor tasks (executing performance tasks). Considering the characteristics of free discussion, which involve generating ideas and narrowing them down in a group, we particularly focused on creativity, decision-making, and cognitive conflict tasks.

Following McGrath's group task circumplex model, we prepared the following five open-ended questions. These questions are designed to require participants to utilize their creativity, decision-making, and resolving cognitive conflict by the current circumstances of discussions. In other words, it is anticipated that the dynamics of the discussion context will be constantly fluctuating, resulting in changing the necessary abilities as well as the level of interdependence with other participants. Additionally, the questions were selected by considering the participants' higher level of familiarity with the topics.

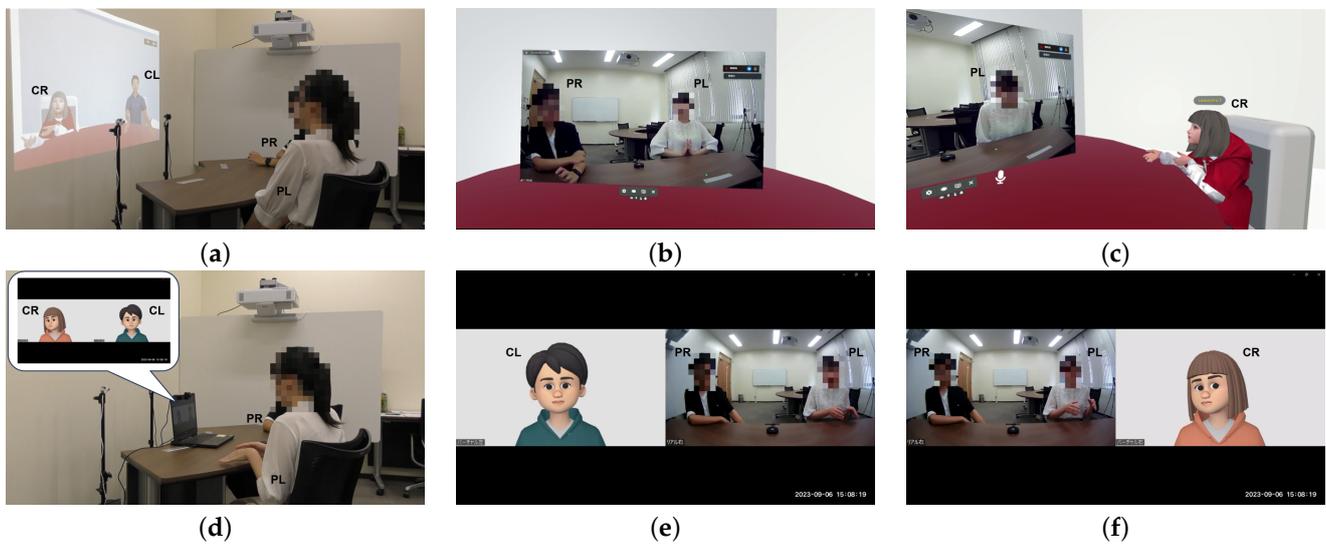
- Optimal approaches for enhancing English conversational skills of Japanese individuals
- Key initiatives for the new urban development in a city of Japan
- Efficient strategies for utilizing multiple social networking services
- Maximizing the quality of life during the COVID-19 pandemic self-isolation period
- Essential competencies required of humans in the era of artificial intelligence

#### 6.4. Experiment Procedure

Each group was given seventy-five minutes to finish the round, ranging from listening to the explanation to answering the questionnaire. Initially, all the participants (PR, PL, CR, and CL) were asked to sign an agreement regarding the purpose, contents, privacy protection, and ethical considerations of the experiment after listening to the explanation for ten minutes. They then engaged in a fifty-minute discussion on a specific topic, as shown in Section 6.3, with other participants in three rooms prepared for this experiment, one of which was used as a local room while the other two were used as remote rooms. The specific topics were disclosed immediately before the commencement of each round, affording participants no opportunity for prior preparation. The discussion took place under three conditions, FR, HC, and MC, each lasting for five minutes. Under the FR condition, all the participants in the local room faced each other in the seating order as shown in Figure 5. Under the HC condition, they all moved to each designated room and looked at Zoom on a laptop without any equipment. Under the MC condition, PR and PL shifted their gaze to VRChat on a wall while CR and CL accessed the cyberspace equipped with HMDs and hand controllers. The experiment was conducted in the order of the MC and HC conditions for the first, third, and fifth rounds, and that for the other two rounds was swapped to control the order effect. After the discussion under each condition, they spent an additional five minutes assessing their discussion using a score sheet of rubric assessment, as shown in Appendix A Tables A1 and A2. Lastly, they worked on responding to a questionnaire regarding the usability of the proposed system in comparison with the other environments for fifty minutes.

Figure 6 shows overviews of the activity under the MC condition as (a) to (c), while those under the HC condition as (d) to (f). In (a), PR and PL were looking at CR and CL as avatars projected on a monitor, (b) and (c) represent the first-person view of the CL in cyberspace, and (c) is the moment of the CL performing a facing behavior to the other avatar, CR. In both spaces, they were all surrounding the apparent circular table and discussing with each other. On the other hand, the moment (d) is identical to (a) with the sole distinction being that CR and CL were displayed on a monitor. As can be seen in the speech balloon, we generated male and female embodied avatars similar to the appearance of those in the MC using a built-in feature of the Zoom application [32] to maintain as many variables as possible under the MC condition, other than the target variable, the spatial localization of learners. The embodied avatars reflected the participants' facing behaviors, facial expressions, and mouth movements. (e) and (f) show images that were displayed on

monitors for CR and CL. The video of PR and PL was captured by a web camera attached to the display of a PC, while the ones of CR and CL were captured by the built-in cameras of the PCs. It should be noted that we hid the Zoom's self-view windows of the participants and arranged the other windows taking into consideration that they were in the same order as those under the MC condition. Ratan et al. revealed that viewing self-video during video conferencing could potentially lead to negative self-focused attention, contributing to virtual meetings or Zoom fatigue [33]. Thus, we mitigated the fatigue by hiding the self-view windows under the HC condition. Additionally, regarding the detection of the facing behaviors, the threshold setting under the MC condition was applied to those under the HC condition, as illustrated on the right in Figure 2. It is important to note that while there were variations in camera placement angles between the MC and HC conditions, the employment of a fisheye camera lens effectively mitigated any disparities in the relative positions and rotations between the cameras and the participants.



**Figure 6.** Overview of the activity under each condition. (a) View of physical space under the MC condition; (b) First-person view of CL seeing physical space under the MC condition; (c) First-person view of CL performing a facing behavior to CR under the MC condition; (d) View of physical space under the HC condition; (e) Display on monitor for CR under the HC condition; (f) Display on monitor for CL under the HC condition.

## 6.5. Measurements

### 6.5.1. Facing Behavior

Initially, we captured five-minute segments from the recorded videos, triggered by cues indicating the beginning of the discussion. We then adjusted the frame cropping to ensure optimal visibility of each participant's face. This process resulted in the creation of six video clips for each round: PR and PL under the MC condition, as well as each participant under the HC condition. The extraction of head pose data employed OpenFace [34], focusing on the "pose\_RY" feature. For CR and CL under the MC condition, data extraction was performed using log files exported from VRChat. VRChat API [35] was utilized to extract the yaw values of head pose data for five minutes based on timestamp information, recorded as "eulerAngles". In both cases, data were recorded in 25 Hz with values expressed in radians and precision up to three decimal places. Additionally, the reference directions were adjusted so that the front directions when seated corresponded to zero radians. Subsequently, we analyzed a total of forty datasets (four roles  $\times$  two conditions  $\times$  five rounds) of facing behaviors. A backward moving average was applied to smooth noise from the waveform with a window size of 1 s (twenty-five data points) to avoid eliminating the components of the waveform representing facing behaviors. After that,

the determination of those who were being faced under the MC condition was performed based on the definition of facing behaviors shown in Section 3.

#### 6.5.2. Utterance

First, we captured a five-minute segment from the recorded audio in the same way as the videos to create a total of forty datasets. For the speech data of PR and PL under the HC and MC conditions, we utilized the voice data obtained from the channels of a directional microphone, which included each participant's voice separated. As for CR and CL under the MC condition, we used screen recording data from a Windows PC, and for CR and CL under the HC condition, individual recording data from Zoom. Next, we extracted the start and end times of their speech and created the script of the speech contents using Faster Whisper [36], developed by Guillaume Klein et al. Faster Whisper is a model that accelerates the OpenAI speech recognition library, Whisper, and can achieve higher transcription accuracy by adjusting parameter values. Particularly, it effectively eliminates silent intervals between the start and end times of the speech to avoid repetitions in speech because the Silero Voice Activity Detection (VAD) model [37] was implemented as "vad\_filter". Moreover, the duration of these intervals can be adjusted using the "min\_silence\_duration\_ms" in the VAD Options class, which was set to 200 ms in this study. Subsequently, based on this analysis we manually confirmed the start and end times and any missing speech content, such as back-channelings, in order to enhance data accuracy. Finally, following the definitions outlined in Section 3 we counted the numbers of simultaneous utterances, intervening utterances, and turn-takings, and measured the duration of silent time.

#### 6.5.3. Rubric Assessment

As shown in Section 3, we aimed for the participants to gain proficiency in four essential areas: engaging in meaningful discussions, fostering creativity, improving their decision-making skills, and enhancing their ability to collaborate effectively. Barkley et al. recommend using rubrics to help assess to what extent learning has occurred and provide several examples of rubrics [38]. The rubrics for assessing discussion and creativity [38] were revised based on those available online [39] and those offered by Brookhart [40]. The rubrics for assessing practical thinking and teamwork [38] were created by faculty working with the Association of American Colleges and Universities (AAC&U) [41]. We cited the rubrics for assessing the four categories of the abilities to create a score sheet of rubric assessment, as shown in Appendix A Tables A1 and A2. Taking into account the feature of the activity, we eliminated several criteria in which the participants were required to work on something before or after the discussion, refer to sources, or take actions during the discussion: "Preparation" in discussion, "Variety of Sources" and "Overall Novelty and Values" in creativity, "Implement Solution" and "Evaluate Outcomes" in practical thinking, and "Individual Contributions Outside of Team Meetings" and "Fosters Constructive Team Climate" in teamwork. Therefore, three criteria for discussion, three for teamwork, four for problem-solving, and two for creativity were prepared for the activities, and the participants were assigned to assess their own activities on a scale of four points in each criterion.

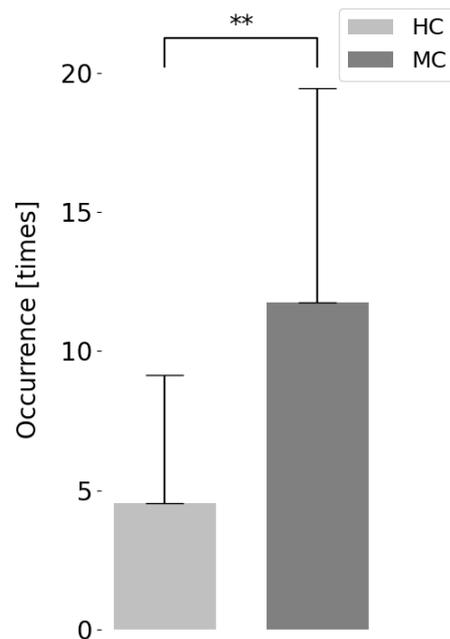
#### 6.5.4. Questionnaire

The questionnaire solicited the participants' free-form responses primarily to five key inquiries: (1) the general user experience with the MC condition, (2) the ease of conversation in comparison to the FR condition and the rationale behind it, (3) the ease of conversation in comparison to the HC condition and the underlying reasons, (4) the primary concern regarding the appearance of avatars under the MC condition, and (5) the ease of conversation when comparing the physical space and the cyberspace of the MC condition, along with the associated reasons. The forms of the questionnaire were collected from all the participants.

## 6.6. Results

### 6.6.1. Facing Behavior

Figure 7 provides a comparison of facing behaviors (transitions of those who were being faced) for all the participants between the HC and MC conditions. We calculated the average occurrence of facing behaviors for each space by participants and conducted an intra-individual comparison using the Wilcoxon signed-rank test to assess the differences in the average occurrences between the MC and HC conditions. The result indicates that the number of facing behaviors under the MC condition was significantly higher than that under the HC condition ( $p = 0.005$ ,  $r = 0.886$ ,  $Z = -2.803$ ).



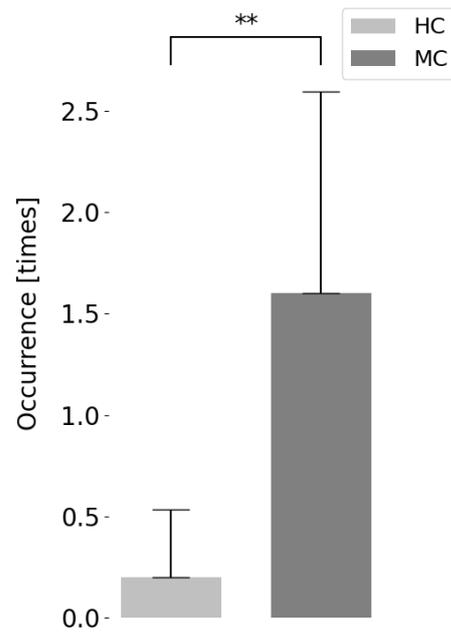
**Figure 7.** Comparison of average occurrence of facing behaviors under the HC and MC conditions. \*\* represents  $p < 0.01$ .

### 6.6.2. Utterance

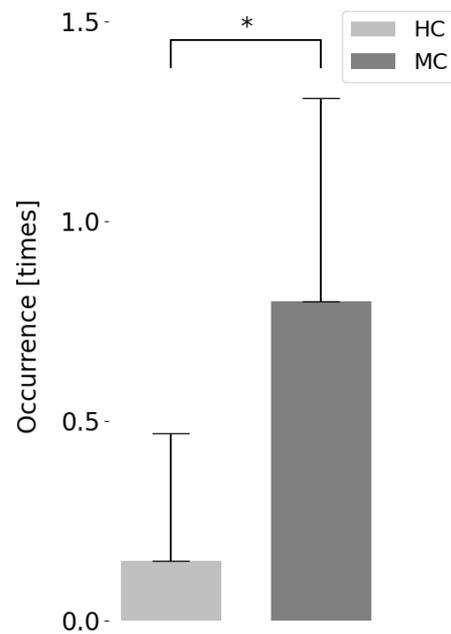
Figure 8 depicts a comparison of simultaneous utterances, while Figure 9 represents that of intervening utterances. Both figures provide a visual analysis of all the participants between the HC and MC conditions. We calculated the average occurrence of simultaneous and intervening utterances for each space by participants and conducted an intra-individual comparison using Wilcoxon signed-rank test to assess the difference in the average occurrences between the MC and HC conditions. Simultaneous utterances ( $p = 0.008$ ,  $r = 0.889$ ,  $Z = -2.666$ ) and intervening utterances ( $p = 0.028$ ,  $r = 0.899$ ,  $Z = -2.201$ ) under the MC condition were significantly higher than those under the HC condition. These results revealed that the numbers of simultaneous and intervening utterances under the MC condition were statistically significantly higher than those under the HC condition.

On the other hand, Figure 10 illustrates a comparison of turn-takings for all the rounds between the HC and MC conditions, while Figure 11 exhibits that of silent time in the same way. Due to the small sample size (five rounds for each condition), we did not apply any statistical analysis. Each figure shows boxplots to visualize the five-number summary provided in Table 1 and to evaluate the results on a median basis. The median occurrence of turn-takings was 12 under the HC condition and 14 under the MC condition, while the median duration of silent time was 7.7 s under the HC condition and 7.1 s under the MC condition. Additionally, the magnitude relationship of the median scores between the two conditions was also consistent with that of all other metrics (Min, 25th percentile, 75th percentile, Max) for both turn-takings and silent time. This implies that the number of

turn-takings was larger and the duration of silent time was smaller under the MC condition than those under the HC condition.



**Figure 8.** Comparison of average occurrence of simultaneous utterances under the HC and MC conditions. \*\* represents  $p < 0.01$ .



**Figure 9.** Comparison of average occurrence of intervening utterances under the HC and MC conditions. \* represents  $p < 0.05$ .

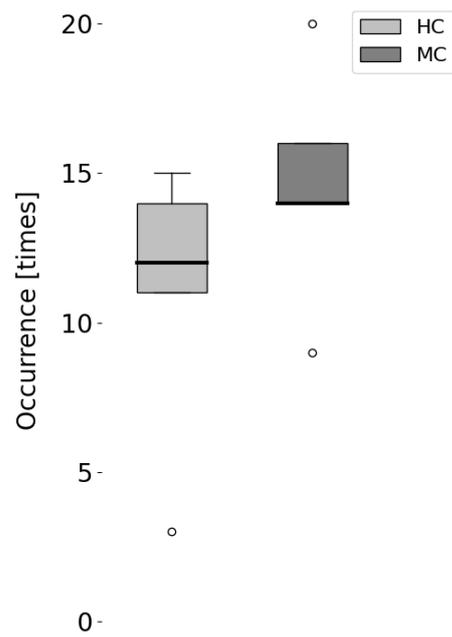


Figure 10. Comparison of median occurrence of turn takings under the HC and MC conditions.

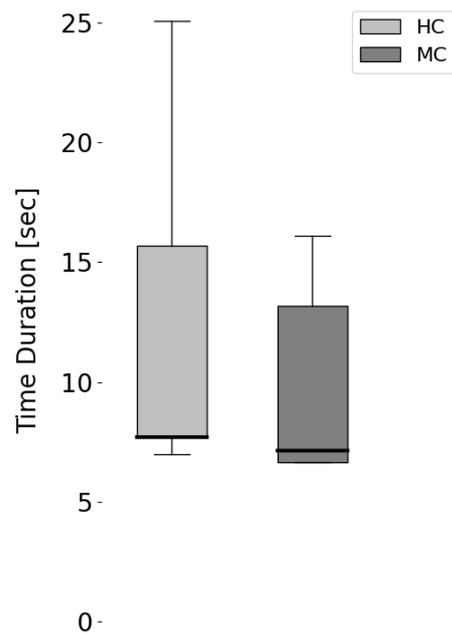


Figure 11. Comparison of median duration of silent time under the HC and MC conditions.

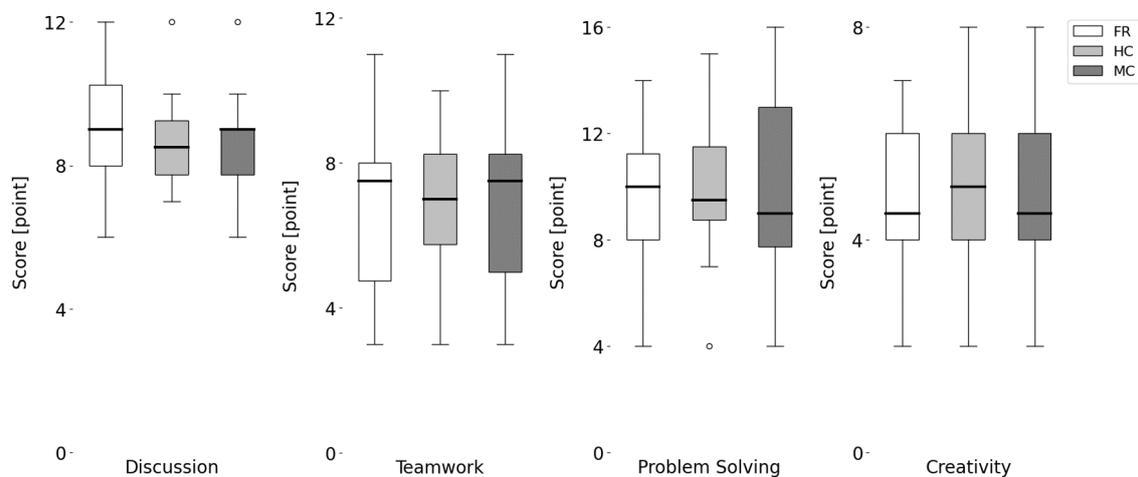
Table 1. Five-number summary on the turn takings and silent time.

Aspect of Speech Behavior	Condition	Min	25%	Median	75%	Max
Turn Takings	HC	11	11	12	14	15
	MC	14	14	14	16	16
Silent Time	HC	7.0	7.7	7.7	15.7	25.1
	MC	6.6	6.6	7.1	13.2	16.1

The unit for turn takings is in times, while that for silent time is in seconds. HC and MC stand for Hybrid Conference and Mirror Campus, respectively.

### 6.6.3. Rubric Assessment

Figure 12 shows a comparison of the median scores derived from all the participants' rubric assessments by categories under the FR, HC, and MC conditions. Given that the assessment is on an ordinal scale, we performed the Friedman test to compare the three conditions. The results show that there were no statistically significant differences between the MC and HC conditions, the HC and FR conditions, and the FR and MC conditions in all the categories: Discussion ( $p = 0.225, T = 2.981$ ), Teamwork ( $p = 0.867, T = 0.286$ ), Problem Solving ( $p = 0.545, T = 1.216$ ), and Creativity ( $p = 0.677, T = 0.780$ ). Therefore, the figure provides a boxplot instead to visualize the five-number summary provided in Table 2 for greater information. Although there was a difference in the median between the two conditions for each category, the magnitude relationship of the median scores was not consistent with that of all other metrics (Min, 25th percentile, 75th percentile, Max).



**Figure 12.** Comparison of rubric assessment scores among the Full Real (FR), HC, and MC conditions.

**Table 2.** Five-number summary and scoring ratio on the rubric assessment scores.

Rubric Criterion	Full Score	Condition	Min	25%	Median	75%	Max	Scoring Ratio
Discussion	12	FR	6	8	9	10.25	12	0.75
		HC	7	7.75	8.5	9.25	10	0.71
		MC	6	7.75	9	9	10	0.75
Teamwork	12	FR	3	4.75	7.5	8	11	0.63
		HC	3	5.75	7	8.25	10	0.58
		MC	3	5	7.5	8.25	11	0.63
Problem Solving	16	FR	4	8	10	11.25	14	0.63
		HC	7	8.75	9.5	11.5	15	0.59
		MC	4	7.75	9	13	16	0.56
Creativity	8	FR	2	4	4.5	6	7	0.56
		HC	2	4	5	6	8	0.63
		MC	2	4	4.5	6	8	0.56

The unit for any numerical value but the scoring ratio is in points. FR stands for Full Real.

Additionally, the table includes the scoring ratio, calculated as the median divided by the full score for each category. In each of the conditions, the ratio in Discussion was more than 0.7, while that in Teamwork, Problem Solving, and Creativity was 0.55 to 0.65, which

indicates that the participants enhanced more applied skills in Discussion than in the other categories through the experiment.

#### 6.6.4. Questionnaire

Regarding the advantages of the MC condition, it was mentioned that the natural and realistic communication enhanced the fluidity of discussions, and there was a sense of immersion in both visuals and sound. As for the disadvantages, it was noted that the weight of VR goggles could be distracting, and certain audio lags could impede speech.

In comparison to the HC condition, several participants provided comments suggesting that the MC condition, where they sat across the table from each other, allowed for a sense of depth, particularly in cyberspace. They also appreciated the ability to use body movements to change their perspective. Conversely, the preference for the HC condition stemmed from the familiarity with communication on Zoom and the acknowledgment that Zoom avatars effectively represented eye contact.

When compared to the FR condition, one of them found the MC condition preferable because there were no other individuals present, creating a more comfortable environment. However, most of them preferred the FR condition, citing the lack of facial expressions and eye contact in VRChat avatars, which made it difficult to gauge the other participants' reactions.

Furthermore, when asked about which space, the physical space or cyberspace, was easier to converse on, their responses were evenly divided. As for improvements on the cyberspace, suggestions included making the video of the physical space more integrated with the cyberspace for increased immersion and providing information on what one's own avatar looked like.

## 7. Discussion

### 7.1. Learning Behaviors and Outcomes in Active Learning

In this experiment, the number of facing behaviors by participants was significantly larger under the MC condition than those under the HC condition. This indicates that the participants were more willing to engage in conversation when conducting the activity under the MC condition. Additionally, the number of simultaneous and intervening utterances by participants were also significantly larger under the MC condition. In light of these results, the participants under the MC condition actually engaged in and interrupted the conversation, and their utterances were sometimes overlapped. Moreover, the number of turn-takings by rounds was larger, and the duration of silent time for more than five seconds by rounds was smaller under the MC condition. Although statistical significance was not tested due to the limited number of samples, the increased number of interruptions led to the larger number of turn-taking and the smaller amount of silent time. Therefore, the MC promoted discussion-oriented behaviors such as facing and speech in the experiment.

These findings are also substantiated by the results of the questionnaire, which indicate that the advantages of the MC over the HC were as follows:

- The fluidity of discussions
- A sense of depth and immersion in both visuals and sound
- The usability of body movements to change individuals' perspective

Accordingly, the MC environment approached discussions closer to those in a physical space in these terms.

The primary difference between the MC and HC conditions was whether or not the positional and directional relationships of the individuals were geometrically aligned to be similar to those in a physical space by the mirror. Which element of the relationships significantly affected the results has not been examined. However, we carefully considered the position, posture, and size of the learners in both spaces to be aligned with each other. Therefore, these outcomes resulted from the increased spatial localization consistency.

In the rubric assessment, in contrast, we could not claim that there was a difference between the HC and MC conditions in all categories, indicating that the learning outcomes

have not at least diminished. It is important to note that the results suggest that there is no significant difference, not only between the HC and MC conditions but also between the FR and HC conditions, as well as the FR and MC conditions. Martin et al. reviewed nine studies and revealed that synchronous online learning did not statistically significantly improve students' cognitive and affective outcomes compared with traditional face-to-face learning [42]. However, the authors also acknowledged that previous studies [43,44] indicated that traditional face-to-face learning is generally perceived to be a more effective method of learning and instruction, which could vary depending on differences in students' targeted outcomes and the context of learning. According to Kemp et al., students strongly favored conducting class discussions face-to-face, stating that they felt more engaged than online discussions [45]. In light of these studies, traditional face-to-face learning could be more effective, specifically on social outcomes in the context of discussions. Hence, the absence of a significant difference between the FR and HC conditions, particularly in engagement-oriented categories such as discussion and teamwork, might imply that students became more accustomed to using conventional video conferencing systems following the onset of the COVID-19 pandemic in 2019, or that the five rounds of the five-minute discussion in each condition among the participants meeting for the first time was not considered sufficient to achieve learning outcomes.

Given that the score sheet of the rubric assessment was systematically produced, we believe that it is necessary to further investigate the learning outcomes under each condition when extending the duration and number of trial repetitions. Taking into account the observed greater enhancement of applied skills in Discussion across all the conditions, to foster development in the other categories it may be imperative to explore alternative tasks beyond the free discussion of answering open-ended questions.

### 7.2. Limitations

Several limitations could potentially impact the results, especially regarding the participants' appearances. Due to technical constraints, the participants in cyberspace engaged in discussions not as their normal selves, but as the designated avatars. The disparities in their appearances between the two spaces could potentially influence the results. If the participants appeared as their normal selves in the HC condition, their facial expressions and eye contact, which were lacking in the Zoom avatars, might make it easier to gauge the other participants' reactions. Thus, if such is the case, their normal selves might mitigate the negative impact of the geometrical misalignment of the learners in the HC condition. Additionally, under the MC condition they used a VRChat avatar that mirrored their body movements, whereas in the HC condition they were supplied with a Zoom avatar focused more on facial expressions and eye movements. The differences in their appearances and the movability of their body parts between the two conditions could also have an impact on the results. Even if technological advances in the future allow us to participate as one's normal selves or as realistic 3D models under the MC condition, according to the Proteus effect we cannot necessarily say that avatars resembling oneself would always be the most effective. The Proteus effect suggests that individuals in a virtual environment adjust their behaviors based on the characteristics of their respective avatars [46]. Leveraging the Proteus effect to deliberately influence their behaviors is called avatar-focused gamification [47], which has been explored as a fruitful approach to improving educational experiences and outcomes. For instance, Oyanagi et al. investigated the effect of an artist-like avatar on the score of creativity on brainstorming through an experiment, and some of the participants reported that the avatar's appearance affected their thinking during the task execution [48]. Accordingly, it may be necessary to customize the appearance of avatars in line with the learning outcomes you attempt to improve.

Moreover, the following limitations were also considered concerning the number of participants and the nature of the discussions. We recruited ten participants and conducted five rounds of the experiment, which was insufficient for robust statistical analyses, especially regarding the median occurrence of turn-takings and the median duration of silent

time for the five rounds. The shortage of substantial support from strict mathematical foundations in the experiment was considered a limitation. In addition, the participants were assigned a single task, involving a fifteen-minute free discussion on a specific topic, with each individual discussing two different topics through the experiment. The nature of the tasks may have influenced the learning behaviors, and more collaborative tasks including note-taking or generating outputs could yield different results. The brief duration of the discussions may also have been too limited to allow for substantial development.

## 8. Conclusions

In this paper, we proposed a synchronous hybrid learning environment for AL scenarios that connects physical and cyberspaces by providing spatial localization of learners similar to that in a physical space. We clarified that, in terms of discussion-oriented behaviors such as facing and speech, the proposed environment significantly increased the occurrences of facing behaviors, intervening, and simultaneous utterances compared to conventional video conferencing. With regards to the learning outcomes achieved through discussions, we found no noticeable differences in the rubric assessment scores between the conditions, indicating that they were not declined. This study has the potential to contribute to the placemaking of discussions in the AL scenario.

We anticipate that the importance of the learning environment, where both local and remote students can feel as if they are engaging in discussions in a single physical environment, will be verified in real school settings. Based on the questionnaire responses, however, it was found that the weight of the VR goggles could be distracting, and certain audio delays could hinder speech, which potentially discourages the adoption of the learning environment in real school settings. To address these issues of HMDs, employing a video conferencing system for remote learners that meets the three requirements outlined in the method could serve as a potential alternative. For instance, such a system might feature a large display to assist remote learners in aligning geometrically with local learners, while also being equipped with a camera to capture and transmit the facing behaviors of remote learners. Additionally, it appeared that being an avatar and differences in controlled behaviors of the VRChat and Zoom avatars often had an impact on the ease of decision-making in one's behaviors and reactions to the other participants' behaviors. On the flip side, considering participants' personality traits when selecting their roles (a human in the physical space or an avatar in the cyberspace) may hold the potential to lead to better discussions. We hope that these improvements in future research will advance the development of a synchronous hybrid learning method that fosters learning outcomes close to a full real space in real school settings.

**Author Contributions:** Conceptualization, S.S. and K.S.; methodology, S.S., S.K., M.H. and K.S.; software, S.S., M.H. and K.S.; validation, S.S., S.K., M.H. and K.S.; investigation, S.S., S.K., M.H. and K.S.; data curation, S.S., S.K., M.H. and K.S.; writing—original draft preparation, S.S.; writing—review and editing, S.K., M.H. and K.S.; supervision, K.S.; project administration, K.S.; funding acquisition, K.S. All authors have read and agreed to the published version of the manuscripts.

**Funding:** This work was supported by the Japan Science and Technology Agency (JST) for the Core Research for the Evolutional Science and Technology (CREST) research project on Social Signaling (JPMJCR19A2), and the Student Research Fellowship of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan.

**Institutional Review Board Statement:** The study was approved by the Internal Ethics Review Board of Institute of Systems and Information Engineering, University of Tsukuba (2022R719-1, 22 August 2023).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** Thanks to the staff and students at the Artificial Intelligence Laboratory at the University of Tsukuba for their support in preparing for and conducting the experiment.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funding organization had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. NEC Corporation has no financial or non-financial interests that could be construed as a conflict of interest in relation to the research presented in this manuscript.

### Abbreviations

The following abbreviations are used in this manuscript:

AL	Active Learning
CA	California
CL	Cyber Left
CR	Cyber Right
FR	Full Real
HC	Hybrid Conference
HMD	Head-Mounted Display
MC	Mirror Campus
PC	Personal Computer
PL	Physical Left
PR	Physical Right
SDK	Software Development Kit
SVRE	Social Virtual Reality Environment
USA	United States of America
VAD	Voice Activity Detection
VR	Virtual Reality

### Appendix A

We created a score sheet of the rubric assessment. Appendix A Table [A1](#) shows the criteria and standards of discussion and teamwork categories, while Appendix A Table [A2](#) shows those of the problem-solving and creativity categories. The way of selecting the criteria and standards is specifically described in Section [6.5.3](#), and all the statements are sourced from [[39–41](#)], as quoted in [[38](#)].

**Table A1.** Description of the Criteria and Standards of Discussion and Teamwork Categories in Rubric Assessment.

Scores	4	3	2	1
Listening	Listens carefully and respectfully to classmates all of the time.	Listens carefully and respectfully to classmates most of the time.	Listens carefully and respectfully to classmates some of the time.	Spaces out a lot during discussion and/or interrupted the speaker.
Speaking	Contributes several meaningful comments to the whole group discussion based on evidence from the text, without dominating the discussion.	Contributes some meaningful comments to the whole group discussion based on evidence from the text, without dominating the discussion.	Contributes one meaningful comment to the whole group discussion based on evidence from the text, without dominating the discussion.	Does not contribute to the group discussion at all or alternately dominated the discussion.
Depth of Thought	All questions and comments show deep understanding and original, profound thought.	Some questions and comments show deep understanding and original, profound thought.	A few questions and comments show deep understanding and original, profound thought.	Questions and comments do not show very deep, original thinking.
Contributes to Team Meetings	Helps the team move forward by articulating the merits of alternative ideas or proposals.	Offers alternative solutions or courses of action that build on the ideas of others.	Offers new suggestions to advance the work of the group.	Shares ideas but does not advance the work of the group.
Facilitates the Contributions of Team Members	Engages team members in ways that facilitate their contributions to meetings by both constructively building upon or synthesizing the contributions of others as well as noticing when someone is not participating and inviting them to engage.	Engages team members in ways that facilitate their contributions to meetings by constructively building upon or synthesizing the contributions of others.	Engages team members in ways that facilitate their contributions to meetings by restating the views of other team members and/or asking questions for clarification.	Engages team members by taking turns and listening to others without interrupting.
Responds to Conflict	Addresses destructive conflict directly and constructively, helping to manage/resolve it in a way that strengthens overall team cohesiveness and future effectiveness.	Identifies and acknowledges conflict and stays engaged with it.	Redirects focus toward common ground, toward task at hand (away from conflict).	Passively accepts alternate viewpoints/ideas/opinions.

**Table A2.** Description of the Criteria and Standards of Problem Solving and Creativity Categories in Rubric Assessment.

Scores	4	3	2	1
Define Problem	Demonstrates the ability to construct a clear and insightful problem statement with evidence of all relevant contextual factors.	Demonstrates the ability to construct a problem statement with evidence of most relevant contextual factors, and problem statement is adequately detailed.	Begins to demonstrate the ability to construct a problem statement with evidence of most relevant contextual factors, but problem statement is superficial.	Demonstrates a limited ability in identifying a problem statement or related contextual factors.
Identify Strategies	Identifies multiple approaches for solving the problem that apply within a specific context.	Identifies multiple approaches for solving the problem, only some of which apply within a specific context.	Identifies only a single approach for solving the problem that does apply within a specific context.	Identifies one or more approaches for solving the problem that do not apply within a specific context.
Propose Solutions/Hypotheses	Proposes one or more solutions or hypotheses that indicates a deep comprehension of the problem. Solution/hypotheses are sensitive to contextual factors as well as all of the following: ethical, logical, and cultural dimensions of the problem.	Proposes one or more solutions or hypotheses that indicates comprehension of the problem. Solutions/hypotheses are sensitive to contextual factors as well as one of the following: ethical, logical, or cultural dimensions of the problem.	Proposes one solution or hypothesis that is “off the shelf” rather than individually designed to address the specific contextual factors of the problem.	Proposes a solution or hypothesis that is difficult to evaluate because it is vague or only indirectly addresses the problem statement.
Evaluate Potential	Solutions Evaluation of solutions is deep and elegant (for example, contains thorough and insightful explanation) and includes, deeply and thoroughly, all of the following: considers history of problem, reviews logic/ reasoning, examines feasibility of solution, and weighs impacts of solution.	Evaluation of solutions is adequate (for example, contains thorough explanation) and includes the following: considers history of problem, reviews logic/ reasoning, examines feasibility of solution, and weighs impacts of solution.	Evaluation of solutions is brief (for example, explanation lacks depth) and includes the following: considers history of problem, reviews logic/reasoning, examines feasibility of solution, and weighs impacts of solution.	Evaluation of solutions is superficial (for example, contains cursory, surface level explanation) and includes the following: considers history of problem, reviews logic/ reasoning, examines feasibility of solution, and weighs impacts of solution.
Variety of ideas	Large number of important and appropriate ideas that span multiple contexts or disciplines.	Ideas represent important and appropriate concepts from different contexts or disciplines.	Ideas are predictable and/ or from the same or similar contexts or disciplines.	Few ideas; ideas are very obvious.
Combination of ideas	Ideas are combined in markedly original and surprising ways to solve a problem, address an issue, or make something new.	Ideas are combined in original ways to solve a problem, address an issue, or make something new.	Ideas are combined in ways that are derived from the thinking of others (for example, of the authors in sources consulted).	Ideas are copied or restated from sources.

## References

- Roseth, C.; Akcaoglu, M.; Zellner, A. Blending synchronous face-to-face and computer-supported cooperative learning in a hybrid doctoral seminar. *Techtrends Tech. Trends* **2013**, *57*, 54–59. [\[CrossRef\]](#)
- Hastie, M.; Hung, I.; Chen, N.; Kinshuk A blended synchronous learning model for educational international collaboration. *Innov. Educ. Teach. Int.* **2010**, *47*, 9–24. [\[CrossRef\]](#)
- Raes, A.; Detienne, L.; Windey, I.; Depaepe, F. A systematic literature review on synchronous hybrid learning: Gaps identified. *Learn. Environ. Res.* **2020**, *23*, 269–290. [\[CrossRef\]](#)
- Li, K.; Wong, B.; Kwan, R.; Wu, M. Learning in a hybrid synchronous mode: Experiences and views of university students. *Int. J. Innov. Learn.* **2023**, *34*, 197–207. [\[CrossRef\]](#)
- Wang, Q.; Huang, Q. Engaging online learners in blended synchronous learning: A systematic literature review. *IEEE Trans. Learn. Technol.* **2024**, *17*, 594–607. [\[CrossRef\]](#)
- Bakx, I.; Turnhout, K.; Terken, J. Facial Orientation During Multi-party Interaction with Information Kiosks. In Proceedings of the Human-Computer Interaction INTERACT '03: IFIP TC13 International Conference On Human-Computer Interaction, Zurich, Switzerland, 1–5 September 2003. [\[CrossRef\]](#)
- Duncan, S.; Fiske, D. Face-To-Face Interaction: Research, Methods, and Theory. 2015. Available online: [https://books.google.co.jp/books?hl=ja&lr=lang\\_en|lang\\_ja&id=o7XMCgAAQBAJ&oi=fnd&pg=PA1&dq=Starkey+Duncan+Jr+and+Donald+W.+Fiske,+Face-to-face+interaction:+research,+methods+and+theory,+Hillsdale,+New+Jersey:+lawrence+Erlbaum,+1977.&ots=McmoWE8r1J&sig=Y6uOaIQtKwcpFFM6aEwqgpliZdE](https://books.google.co.jp/books?hl=ja&lr=lang_en|lang_ja&id=o7XMCgAAQBAJ&oi=fnd&pg=PA1&dq=Starkey+Duncan+Jr+and+Donald+W.+Fiske,+Face-to-face+interaction:+research,+methods+and+theory,+Hillsdale,+New+Jersey:+lawrence+Erlbaum,+1977.&ots=McmoWE8r1J&sig=Y6uOaIQtKwcpFFM6aEwqgpliZdE) (accessed on 7 April 2024).
- Cunningham, U. Teaching the disembodied: Othering and activity systems in a blended synchronous learning situation. *Int. Rev. Res. Open Distrib. Learn.* **2014**, *15*, 33–51. [\[CrossRef\]](#)
- Johnson, D.; Johnson, R.; Smith, K. Cooperative learning returns to college what evidence is there that it works? *Chang. Mag. High. Learn.* **1998**, *30*, 26–35. [\[CrossRef\]](#)
- Hmelo-Silver, C. Problem-based learning: What and how do students learn? *Educ. Psychol. Rev.* **2004**, *16*, 235–266. [\[CrossRef\]](#)
- Freeman, S.; Eddy, S.; McDonough, M.; Smith, M.; Okoroafor, N.; Jordt, H.; Wenderoth, M. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8410–8415. [\[CrossRef\]](#)
- Fink, L. Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses. 2013. Available online: [https://books.google.co.jp/books?hl=ja&lr=lang\\_en|lang\\_ja&id=cehvAAAAQBAJ&oi=fnd&pg=PR7&dq=Fink,+L.+D.+2013.+Creating+significant+learning+experiences.+An+integrated+approach+to+designing+college+courses.+John+Wiley+%26+Sons.&ots=GDpIuU9qEL&sig=7u-sFzU7EvOnOelAolOuFlISyUf#v=onepage&q=Fink%2C%20L.%20D.%20.%20\(2013\).%20Creating%20significant%20learning%20experiences%3A%20An%20integrated%20approach%20to%20designing%20college%20courses.%20John%20Wiley%20%26%20Sons.&f=false](https://books.google.co.jp/books?hl=ja&lr=lang_en|lang_ja&id=cehvAAAAQBAJ&oi=fnd&pg=PR7&dq=Fink,+L.+D.+2013.+Creating+significant+learning+experiences.+An+integrated+approach+to+designing+college+courses.+John+Wiley+%26+Sons.&ots=GDpIuU9qEL&sig=7u-sFzU7EvOnOelAolOuFlISyUf#v=onepage&q=Fink%2C%20L.%20D.%20.%20(2013).%20Creating%20significant%20learning%20experiences%3A%20An%20integrated%20approach%20to%20designing%20college%20courses.%20John%20Wiley%20%26%20Sons.&f=false) (accessed on 7 April 2024).
- Bower, M.; Dalgarno, B.; Kennedy, G.; Lee, M.; Kenney, J. Design and implementation factors in blended synchronous learning environments: Outcomes from a cross-case analysis. *Comput. Educ.* **2015**, *86*, 1–17. [\[CrossRef\]](#)
- Carruana Martin, A.; Alario-Hoyos, C.; Delgado Kloos, C. A Study of Student and Teacher Challenges in Smart Synchronous Hybrid Learning Environments. *Sustainability* **2023**, *15*, 11694. [\[CrossRef\]](#)
- Kendon, A. Conducting Interaction: Patterns of Behavior in Focused Encounters. 1990. Available online: [https://books.google.co.jp/books?hl=ja&lr=lang\\_en|lang\\_ja&id=7-8zAAAAIAAJ&oi=fnd&pg=PA1&dq=Conducting+interaction:+Patterns+of+behavior+in+focused+encounters&ots=oAiabTWYC-&sig=aW6c1muxBuoQSSIUTPt2BudnQ](https://books.google.co.jp/books?hl=ja&lr=lang_en|lang_ja&id=7-8zAAAAIAAJ&oi=fnd&pg=PA1&dq=Conducting+interaction:+Patterns+of+behavior+in+focused+encounters&ots=oAiabTWYC-&sig=aW6c1muxBuoQSSIUTPt2BudnQ) (accessed on 7 April 2024).
- Lang, J. Creating Architectural Theory, The Role of The Behavioral Sciences in Environmental Design. 1987. Available online: [https://books.google.co.jp/books/about/Creating\\_Architectural\\_Theory.html?id=IHLwQgAACAAJ&redir\\_esc=y](https://books.google.co.jp/books/about/Creating_Architectural_Theory.html?id=IHLwQgAACAAJ&redir_esc=y) (accessed on 7 April 2024).
- Scott-Webber, L. Environmental Behavior Research and the Design of Learning Spaces. 2004. Available online: [https://www.academia.edu/24124687/Environmental\\_Behavior\\_Research\\_and\\_the\\_Design\\_of\\_Learning\\_Spaces](https://www.academia.edu/24124687/Environmental_Behavior_Research_and_the_Design_of_Learning_Spaces) (accessed on 7 April 2024).
- Chessa, M.; Solari, F. The sense of being there during online classes: Analysis of usability and presence in web-conferencing systems and virtual reality social platforms. *Behav. Inf. Technol.* **2021**, *40*, 1237–1249. [\[CrossRef\]](#)
- Girvan, C.; Tangney, B.; Savage, T. SLurtles: Supporting constructionist learning in second life. *Comput. Educ.* **2013**, *61*, 115–132. [\[CrossRef\]](#)
- Mystakidis, S.; Berki, E.; Valtanen, J. Deep and meaningful e-learning with social virtual reality environments in higher education: A systematic literature review. *Appl. Sci.* **2021**, *11*, 2412. [\[CrossRef\]](#)
- Sawada, S.; Kim, S.; Hirokawa, M.; Suzuki, K. Effect of Using Embodied Avatars on Turn-taking during Conversational Activities in a Social VR Space. *IEEE Int. Conf. Eng. Technol. Educ.* **2021**, *1*, 934–937. [\[CrossRef\]](#)
- Ooko, R.; Ishii, R.; Nakano, Y. Estimating a user's conversational engagement based on head pose information. In Proceedings of the Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, 15–17 September 2011; pp. 262–268. [\[CrossRef\]](#)
- Maroni, B.; Gnisci, A.; Pontecorvo, C. Turn-taking in classroom interactions: Overlapping, interruptions and pauses in primary school. *Eur. J. Psychol. Educ.* **2008**, *23*, 59–76. [\[CrossRef\]](#)
- Hachisu, T.; Pan, Y.; Matsuda, S.; Bourreau, B.; Suzuki, K. FaceLooks: A smart headband for signaling face-to-face behavior. *Sensors* **2018**, *18*, 2066. [\[CrossRef\]](#) [\[PubMed\]](#)
- Setting up the SDK. Available Online: <https://creators.vrchat.com/sdk/> (accessed on 6 November 2023).

26. Integrate a Character Creator into Your Game or App in Days. Available online: <https://readyplayer.me/> (accessed on 6 November 2023).
27. Fahlenbrach, K.; Schröter, F. Embodied avatars in video games: Audiovisual metaphors in the interactive design of player characters. In *Embodied Metaphors in Film, Television, and Video Games*; Routledge: London, UK, 2015; pp. 251–268. [CrossRef]
28. XSOverlay. Available online: <https://store.steampowered.com/app/1173510/XSOverlay/?> (accessed on 6 November 2023).
29. Webb, N.; Kenderski, C. Gender differences in small-group interaction and achievement in high-and low-achieving classes. *Gen. Infl. Classr. Interact.* **1985**, *1*, 209–236. [CrossRef]
30. Hackman, J. Effects of task characteristics on group products. *J. Exp. Soc. Psychol.* **1968**, *4*, 162–187. [CrossRef]
31. Straus, S.; McGrath, J. Does the medium matter? The interaction of task type and technology on group performance and member reactions. *J. Appl. Psychol.* **1994**, *79*, 87. [CrossRef]
32. Quick Guide to Creating Your Personalized Zoom Avatar. Available online: <https://www.askdaveytaylor.com/quick-guide-to-creating-your-personalized-zoom-avatar/> (accessed on 29 October 2023).
33. Ratan, R.; Miller, D.; Bailenson, J. Facial appearance dissatisfaction explains differences in zoom fatigue. *Cyberpsychol. Behav. Soc. Netw.* **2022**, *25*, 124–129. [CrossRef] [PubMed]
34. Amos, B.; Ludwiczuk, B.; Satyanarayanan, M. OpenFace: A General-Purpose Face Recognition Library with Mobile Applications. 2016. Available Online: <http://cmusatyalab.github.io/openface/> (accessed on 8 November 2023).
35. VRChat API. Available online: <https://udonsharp.docs.vrchat.com/vrchat-api/> (accessed on 8 November 2023)
36. Faster Whisper Transcription with CTranslate2. Available online: <https://github.com/guillaumekln/faster-whisper> (accessed on 8 November 2023).
37. Team, S. Silero VAD: Pre-Trained Enterprise-Grade Voice Activity Detector (VAD), Number Detector and Language Classifier. 2021. Available online: <https://github.com/snakers4/silero-vad> (accessed on 8 November 2023).
38. Barkley, E.; Major, C. Learning Assessment Techniques: A Handbook for College Faculty. 2015. Available online: [https://books.google.co.jp/books?hl=ja&lr=lang\\_en|lang\\_ja&id=0pstCwAAQBAJ&oi=fnd&pg=PR5&ots=b0QhYZ0HMv&sig=b0pnm6HZhxi2JgecG5oqcawYBQE](https://books.google.co.jp/books?hl=ja&lr=lang_en|lang_ja&id=0pstCwAAQBAJ&oi=fnd&pg=PR5&ots=b0QhYZ0HMv&sig=b0pnm6HZhxi2JgecG5oqcawYBQE) (accessed on 7 April 2024).
39. Discussion Rubric. Available online: <https://www.edutopia.org/pdfs/stw/edutopia-stw-assessment-9th-grade-humanities-discussion-rubric.pdf> (accessed on 29 October 2023).
40. Brookhart, S. Assessing Creativity. *Educ. Leadersh.* **2013**, *70*, 28–34. [CrossRef]
41. Value Rubrics. Available online: <https://www.aacu.org/initiatives/value-initiative/value-rubrics> (accessed on 29 October 2023).
42. Martin, F.; Sun, T.; Turk, M.; Ritzhaupt, A. A meta-analysis on the effects of synchronous online learning on cognitive and affective educational outcomes. *Int. Rev. Res. Open Distrib. Learn.* **2021**, *22*, 205–242. [CrossRef]
43. Siler, S.; VanLehn, K. Learning, interactional, and motivational outcomes in one-to-one synchronous computer-mediated versus face-to-face tutoring. *Int. J. Artif. Intell. Educ.* **2009**, *19*, 73–102. [CrossRef]
44. Francescucci, A.; Rohani, L. Exclusively synchronous online (VIRI) learning: The impact on student performance and engagement outcomes. *J. Mark. Educ.* **2019**, *41*, 60–69. [CrossRef]
45. Kemp, N.; Grieve, R. Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. *Front. Psychol.* **2014**, *5*, 1278. [CrossRef]
46. Yee, N.; Bailenson, J. The Proteus effect: The effect of transformed self-representation on behavior. *Hum. Commun. Res.* **2007**, *33*, 271–290. [CrossRef]
47. Ratan, R.; Klein, M.; Ucha, C.; Cherchiglia, L. Avatar customization orientation and undergraduate-course outcomes: Actual-self avatars are better than ideal-self and future-self avatars. *Comput. Educ.* **2022**, *191*, 104643. [CrossRef]
48. Oyanagi, A.; Narumi, T.; Lugrin, J.; Aoyama, K.; Ito, K.; Amemiya, T.; Hirose, M. The Possibility of Inducing the Proteus Effect for Social VR Users. *Int. Conf. -Hum. Interact.* **2022**, *13518*, 143–158. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.