Collaborative Online Learning with VR Video: Roles of Collaborative Tools and Shared Video Control

Qiao Jin*
jin00122@umn.edu
University of Minnesota
Minneapolis, MN, USA

Yu Liu*
liu00885@umn.edu
University of Minnesota
Minneapolis, MN, USA

Ruixuan Sun
sun00587@umn.edu
University of Minnesota
Minneapolis, MN, USA

Chen Chen
chen4625@umn.edu
University of Minnesota
Minneapolis, MN, USA

Puqi Zhou
pzhou@gmu.edu
George Mason University
Fairfax, VA, USA

Bo Han
bohan@gmu.edu
George Mason University
Fairfax, VA, USA

Feng Qian
fengqian@umn.edu
University of Minnesota
Minneapolis, MN, USA

Svetlana Yarosh
lana@umn.edu
University of Minnesota
Minneapolis, MN, USA

ABSTRACT

Virtual Reality (VR) has a noteworthy educational potential by providing immersive and collaborative environments. As an alternative but cost-effective way of delivering realistic environments in VR, using 360-degree videos in immersive VR (VR videos) received more attention. Although many studies reported positive learning experiences with VR videos, little is known about how collaborative learning performs on VR video viewing systems. In this study, we implemented two collaborative VR video viewing modes based on the way of group video control, synchronized or shared (Sync mode) and non-synchronized or individual (Non-sync mode) video control, against a conventional VR video viewing setting (Basic mode). We conducted a within-subject study (N = 54) in a lab-simulated remote learning environment. Our results show that collaborative VR video modes (Sync and Non-sync mode) improve users' learning experiences and collaboration quality, especially with shared video control. Our findings provide directions for designing and employing collaborative VR video tools in online learning environments.

CCS CONCEPTS

• **Applied computing** → **Collaborative learning**; **Distance learning**; • **Human-centered computing** → **Virtual reality**.

KEYWORDS

Virtual Reality, educational VR, collaborative learning, 360-degree video, social VR

ACM Reference Format:

Qiao Jin, Yu Liu, Ruixuan Sun, Chen Chen, Puqi Zhou, Bo Han, Feng Qian, and Svetlana Yarosh. 2023. Collaborative Online Learning with VR Video: Roles of Collaborative Tools and Shared Video Control. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3544548.3581395

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '23, April 23–28, 2023, Hamburg, Germany

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581395

1 INTRODUCTION

Learning in immersive Virtual Reality (VR)[1] has many benefits including engagement, motivation, and higher learning outcomes [46]. A holistic multi-stakeholders investigation identified several rationales for why people should use VR in higher education — increasing social presence, accessing otherwise inaccessible learning contexts, remembering visual and spatial knowledge, and supporting embodied learning [53]. These reasons provide a compelling opportunity to support online learning — a new normal in the post-COVID world. However, creating VR learning content may significantly increase the initial investment for a class [25, 56, 69]. It is not easy to simulate realism in VR for instructors without technical backgrounds in 3D model design or program creation. Even for a few VR platforms which have provided pre-designed 3D models and environments, teaching with them still requires training and is hard to meet the need for diversified education purposes [53].

With the proliferation of inexpensive panoramic consumer video cameras and various types of video editing software, using 360-degree videos in VR (VR videos[2]) has attracted more attention as an alternative method for instructors building a realistic and immersive environment. It is a "more user-friendly, realistic and affordable" [26] way to create a realistic digital experience compared to developing a simulated VR environment [92, 97]. The substantial 360-degree video resources and the nature to enable self-paced learning [31] position VR video as a promising tool to fit different educational areas, activities, and settings. However, current VR video viewing systems are not developed as collaborative tools to be used for educational goals, with most of them only providing an individual or asynchronous viewing experience. The lack of collaboration experience when watching the video can potentially downgrade the learning experience [64, 127], as collaboration maybe the key to educational VR's success [53]. This articulates a research gap in the development and empirical investigation of collaboration VR video learning environments.

[1] In this paper, "VR" specifically means the immersive VR that supports rotating the head or with a full 6 Degree-of-Freedom (6DoF), because it creates more immersive experiences and has a noteworthy educational potential [5].

[2] "VR videos" specifically mean in-headset monoscopic 360-degree videos in this work. We chose this format because it has a wide range of applications and is commonly used within the industry [79].

In this study, we investigate two collaborative modes with shared or individual VR video control systems compared with a conventional VR-based learning setting. Each mode contains a video viewing system and an after-video platform for further discussion and collaboration. ***Basic mode*** uses a conventional VR video viewing system together with an existing widely-available online platform (Zoom [19] and Google Slides [38]) [49, 81]. ***Non-sync mode*** includes a collaborative video viewing system with individual control and video timeline and an in-VR platform for after-video discussion. ***Sync mode*** contains the same in-VR after-video platform, but students have shared video control. We investigate the role of collaborative tools and shared video control driven by two core questions:

- **RQ1:** How does VR video delivery via existing technology (Basic mode) compare to collaborative VR video delivery proposed in this study (Sync and Non-Sync mode) on measures of knowledge acquisition, collaboration, social presence, cognitive load and satisfaction?
- **RQ2:** How does individual VR video control (Non-sync mode) compare with shared video control (Sync mode) on measures of knowledge acquisition, collaboration, social presence, cognitive load, and satisfaction?

To answer these questions, we conducted a controlled experiment with 54 college students. Our results showed that visual factual knowledge acquisition was statistically significantly higher in collaborative VR video delivery with moderate and high effect sizes compared to existing technology. However, auditory factual knowledge acquisition was higher in the conventional VR conndition. Three modes performed equally well regarding conceptual knowledge and cognitive load. Our results saw significantly higher scores on collaboration experiences and social presence in Sync mode than in the Basic and Non-sync mode, and significantly higher scores on satisfaction than in the Basic mode. Our qualitative results allowed us to triangulate and enrich these quantitative findings, showing that: 1) a tension existed between time flexibility and communication comfort when watching video together; 2) shared control influenced the perceived usefulness of collaborative tool; and 3) in-VR platform for after-video discussion enhanced visual transmission and engagement. Based on our results, we provide implications for design and research on collaborative VR video viewing in the education area.

To the best of our knowledge, this work is the first study exploring distributed collaborative learning experiences using VR videos as learning material. The contributions of our study include:

- Investigating how three forms of VR video viewing modes affect knowledge acquisition, collaboration, social presence, cognitive load, and satisfaction quantitatively, and conducting a qualitative thematic analysis to triangulate data;
- Conducting a partial replication[3] [45, 47] (a type of scientific work that has long been encouraged but rarely undertaken at HCI venues [40, 47, 122, 123]) of prior video viewing work [79], and extending it with additional outcomes by expanding into the educational context;
- Formulating design implications and insights for collaborative online learning based on VR videos.

[3]The definition of partial replication used in this study is "using deliberate modifications of earlier research, with the aim of testing them in different settings, with different demographic groups of participants, or other operationalizations of variables."[47]

2 RELATED WORK

2.1 Video Based Learning and Collaboration

Video Based Learning (VBL) is "the process of using video technology to acquire knowledge or skills" [33]. It is a powerful method for distance learning due to the widespread availability of online open video resources [85, 126] and its nature to enable self-paced and autonomous learning [31]. VBL is a widely accepted e-learning trend, and it is gaining momentum during the recent pandemic [85]. However, prior work [27] has outlined the several limitations when watching videos alone: students may confront distraction, feel isolated, easily become discouraged from lack interactivity, and fail to generate deeper reflections about the learning content.

Studies show the potential of combining educational video and collaborative learning to mitigate that gap. This combination supports students' intrinsic motivation, which could maximize the use of cognitive resources that are directly related to learning [66, 103]. Research shows that watching videos together positively increases students' attention span and engages them in collaboration [65]. Furthermore, conversation raised from video content is vital for knowledge building, and maybe as or more important than the videos themselves [87] — the different interpretations and connections provide new perspectives [37] and generate important conceptual diversity [82]. Conversations about the video can be helpful during or after watching the video [31].

In order to promote collaboration in VBL for distance learning, Cadiz et al. [11] proposed DCVV (Distributed Collaborative Video Viewing) model, wherein remote students watch educational videos in small groups without a facilitator, with the opportunity to periodically pause the video and discuss it. The DCVV model has been proven to have a positive effect on students' engagement and learning[11]. Enabling the DCVV scenarios requires two components. First, this scenario requires distributed lecture video viewing systems with shared controls, such as play, stop, pause, seek. Second, it should have a communication system for discussion of the video content (including during and after watching the video). Given the advantages of DCVV model, it has the potential to inform the design of an innovative collaborative VBL system that promotes meaningful and engaging learning experiences. We seek to test the potential of DCVV in the educational VR video. One of our modes (see Section 3.1.3) operationalizes this approach for comparison to both baseline and prior work.

2.2 VR Video in Education

Another potential solution to conventional VBL's problems, including difficulty motivating students' interest and promoting their involvement [80, 118], is presented by learning in immersive environments [102]. Many studies have reported that 360-degree VR videos may be one powerful type of educational VR: increasing users' interest and enjoyment [102], making students become more absorbed and engaged in learning activities [92, 102], accessing otherwise inaccessible learning contexts and helping students understand and remember visual knowledge [53].

VR video has been used in many subjects and learning scenarios. Pirker et al. [92] provided an overview of the literature and identified major use cases supported by VR video-based educational activities, including virtual tours [1, 68, 130], recorded processes and procedures [51, 93], recorded situations (e.g., nurse education, safety training) [44, 84], recorded experiences [4], and recorded processes for learning through reply (e.g., teacher education, sports training) [13, 36, 119]. Compared with other computer-generated graphical immersive environments, recording videos and using them with VR players is easier and more cost-effective, since they do not require any advanced development and design skills [92]. Educators can readily create a realistic VR experience with an omnidirectional (360-degree) camera or directly find open resources online [53].

Given the many positive perceptions of educational VR video, prior work calls for understanding feasibility and actual impact on learning outcomes. Some prior studies have reported shortcomings such as inconsistent learning outcomes [52], increased cognitive load [92] and distraction and poor concentration when applying VR video in educational area [111]. Although there are theoretical grounds [58] showing that collaborative learning can reduce cognitive load and yield better learning outcomes, previous reviews [92, 111] have also argued that classic VR video viewing systems struggle to meet students' needs for social interaction and collaboration. To build on these prior studies and work towards addressing these challenges, we chose to investigate the role of collaborative tools and shared control systems for VR collaborative video viewing , and selected to measure considering both the knowledge acquisition and factors highly relative to learning and collaboration: collaboration (e.g., [21, 112]), social presence (e.g., [17, 48, 88, 91]), cognitive load (e.g., [23, 88]), and satisfaction (e.g., [14, 112]). Section 4.4.2 elaborates the rationale for the inclusion of each of these factors.

2.3 VR Collaboration in Education

As VR becomes a more widely accessible technology, numerous VR learning applications have emerged in education domains such as healthcare(e.g., [2, 94, 108, 121]), engineering (e.g.,[32, 54, 100, 105]), science (e.g.,[50, 62, 75]) and general-purpose education(e.g.,[9, 86]). Collaborative learning is perceived as an effective way to increase learners' mutual concerns, motivation, and social presence in learning systems [61, 99]. Many studies have demonstrated that VR can improve the quality of collaboration with the visualizations and immersive environments it provides [90]. However, most studies on collaborative VR in the education area mainly focus on 3D environments instead of videos [8]. For example, Drey et al. [23] used a cooperative learning approach, adding a teacher role to their system to foster pair learning.

Though there have been some prior explorations of collaborative video tools in VR (e.g., [60, 63, 79, 110, 120]), few of them focus on an educational context. One example is CityCompass [55], in which two users practice second-language together in a wayfinding scenario through the interactive VR video city landscapes. This system focused on interactive VR videos, providing the viewers the ability to jump to another video clip together. For those systems that target VR video, the most relevant work to our investigation is CollaVR [79], which introduces a multi-user VR video system that targets collaborative video reviewing for filmmaking. While CollaVR was implemented and tested outside of an educational context, it proposed three design features for VR collaboration (awareness visualization, view sharing and note taking) that may also be relevant in the learning context. CollaVR was evaluated with two small groups (two to three people) where participants sat at computer desks in the same physical location. The results showed that, compared with a general VR video player without any collaboration tools, this system more effectively supported multiple users to collaborate, discuss and review VR video together. While CollaVR did not focus explicitly on an educational context, we seek to extend it to the education domain by comparing their approach with one integrating an explicit learning model (i.e., DCVV).

Fortunately, some of the collaborative VR techniques proposed in this prior work [79] may have potential benefits for operationalizing DCVV or other collaborative approaches in VR. For this work, we recreated the core collaborative features of CollaVR [79] and compared it with the same baseline condition as a partial replication [47]. It is an attempt to expand and generalize a prior study in the educational area by repeating some factors (see 3.1.2) but introducing deliberate changes such as settings, participants and operationalization of variables. This partial replication allowed us to extend the elements considered in the evaluation to learning outcomes, investigate them with different demographic groups of participants, and conduct an explicit comparison of this collaboration approach both against a baseline and against an approach that operationalizes the DCVV model.

3 SYSTEM DESIGN AND IMPLEMENTATION

To answer our RQs, we developed three system modes based on different VR video-based learning systems: Basic mode, Non-Sync mode and Sync mode. Since a portion of the discussion in the DCVV approach occurs after watching the video [11], we included an after-video discussion platform in each mode. For supporting after-video discussion in Basic mode, we used existing widely-available online platforms (Zoom and Google Slides) [49, 74, 81, 114]. Specific differences in functionality and implementation between the three modes' viewing systems and related after-video platform designs are detailed in Table 1.

We implemented our systems with Unity and C# on Oculus Quest I [4]. The multi-user features were implemented with Photon Platform [89]. The connection and server were maintained by Photon. The 3D models and avatars used in the system were from Unity Asset store. We used Google AI [39] to implement speech-to-text function.

3.1 VR Video Viewing Systems

3.1.1 Basic Mode. We developed Basic mode using existing widely-available technologies for online learning (Figure 1a). It is the baseline condition that was defined in the prior work [79], which allows us to conduct a partial replication of prior work on collaborative VR video. It consisted of software and hardware that could be currently obtainable in a standard classroom setting from a practical perspective: a VR headset with a Basic VR video player (with only play, pause and independent timeline control functions), and a pen and

[4]Our system and code are available at https://github.com/GeorgieQiaoJin/CollaborativeVRVideoPlayer

Table 1: Function comparison of three modes (Basic, Non-sync and Sync). Each mode has a video viewing system and an after-video platform for further discussion and collaboration. Basic mode uses a conventional VR video viewing system together with an existing widely-available online platform (Zoom and Google Slides). Non-sync mode includes a collaborative video viewing system with individual video control and an in-VR platform for after-video session. The comparison of Basic mode and Non-sync mode constitute a partial replication of prior work [79]. Sync mode contains the same in-VR platform and a DCVV system operationalizing and implementing the DCVV model [11], in which people have shared process control under the synchronized video timeline.

		Basic Mode	Non-sync Mode	Sync Mode
Video Control		Individual control	Individual control	Shared control
Awareness	Voice chat	Voice chat via Zoom	Spatialized voice chat based on viewport's direction	Spatialized voice chat based on avatar's location
	Embodied visualization	None	None	3D avatar
	Activity visualization	None	Screenshot indicator/ speech-to-text note indicator/ recording status indicator/ teammates' location	Screenshot indicator/ speech-to-text note indicator/ recording status indicator/ teammates' cursor
	Viewport visualization	None	Only visualized under follow function	Always displayed on the video
View-sharing		None	Peek/peek in full window/follow	None
Note-taking		Pen and paper	Screenshot/speech-to-text note/drawings	
After-video Discussion Platform		Zoom and Google Slides	In-VR discussion room	

paper (for note-taking). As our work focuses on remote education, which is different from the in-person setting used in the prior work we replicated, we provided a laptop with Zoom (for within-group communication throughout the during-video and after-video session) and Google Slides (only used for after-video session) to meet the needs of communication and teamwork. We did not provide any in-VR communication method or note-taking functions in the Basic mode because those functions are usually not provided in the off-the-shelf VR video applications. The main purposes of the Basic mode are 1) to provide a conventional environment with existing technologies to test whether improvements to this status lead to better outcomes for students and 2) to replicate the baseline condition of the prior study.

3.1.2 Non-Sync Mode. The Non-Sync (Non-Synchronous) mode provides users with a collaborative VR video watching experience with individual timeline progress and control, and a variety of functions that promote in-VR communication (Figure 1b). We adopted the following features (constituting a partial replication of prior work with Basic mode [79]): awareness tools (spatialized voice chat, activity visualization, viewport visualization), view sharing tools (peek, peek in full window and follow) and note-taking tools (drawing, speech-to-text note, screenshot).

Awareness: We implemented three awareness tools to provide a natural social environment for users and help them know what others are working on: 1) *Spatialized voice chat*: uses users' viewports as the source of their audio to improve the sense of 3D presence while watching the video. 2) *Activity visualization*: the progress of other users is displayed, screenshots and speech-to-text notes indicator on the progress bar, along with the recording status indicator showing if the user is currently taking a speech-to-text note. 3) *Viewport visualization*: When co-viewing video under the follow function (to be detailed below), gaze directions are visualized with a rectangle box overlaying on the video or an arrow indicator of an out-of-frame gaze.

View sharing: We provided three view-sharing functions, which establish a common context for discussion under a non-synchronous timeline: 1) *Peek*: showing a thumbnail view of other users, 2) *Peek in full window*: giving a full view of other users. 3) *Follow*: following other users on timeline. When the user is following others, the viewport will be shown to indicate the user's current viewing direction.

Note-taking: We provided the following note-taking and note-taking tools: 1) *Screenshot*: capturing the current view and saving it as a picture. 2) *Speech-to-text note*: recording user's speech and transcribing the audio into a text note, along with a screenshot taken at the same time the speech note starts. 3) *Drawings*: allowing users to draw and annotate on the video content. The drawings can be captured by the screenshots as well so they can be combined with the speech-to-text note. To avoid distractions from other users on a different timeline, the drawings are only visible to other users in follow function (see above).

3.1.3 Sync Mode. The Sync (Synchronous) mode was designed based on the DCVV model (Distributed Collaborative Video Viewing) [11] where users have shared VCR controls (play, pause, seek, timestamp) among their group (Figure 1c). Besides the major difference between synchronous and non-synchronous video playback, we also make the following adjustments to the functions in Non-Sync mode to fit the needs of shared control:

Awareness: To enable better awareness under a synchronous timeline, we added embodied visualization and adjusted the prior

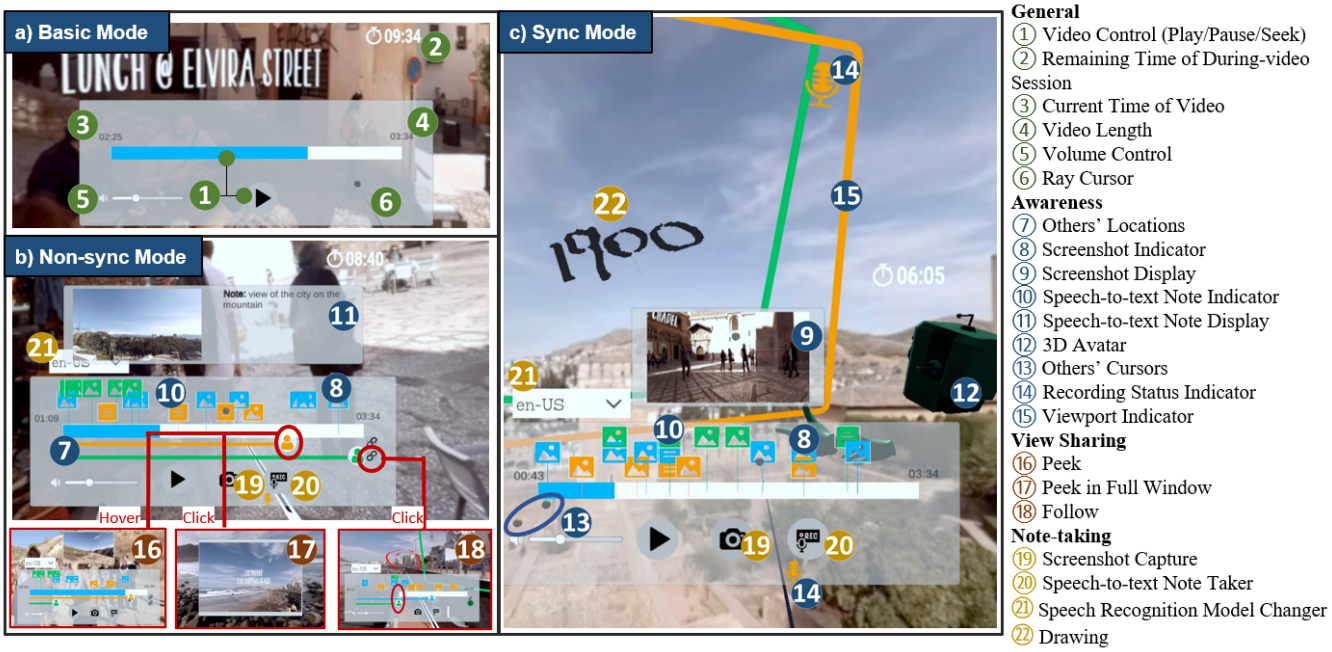

Figure 1: User interface of the three video viewing systems with our design features: a) Basic Mode (a plain VR video player without any collaboration tool, see Section 3.1.1 for more details); b) Non-Sync Mode (a collaborative VR video viewing system with individual timeline control, see Section 3.1.2 for more details); c) Sync Mode (a collaborative VR video viewing system with synchronous video control, see Section 3.1.3 for more details).

awareness tool as follows: 1) *Embodied visualization*: giving 3D avatars in order to increase the awareness between teammates. 2) *Spatialized voice chat*: unlike Non-sync mode, Sync mode uses the position of users' 3D avatars as the audio source, instead of their viewports, to create a more natural communication environment. 3) *Activity visualization*: Since all users shared the same video progress, in Sync mode we do not show the individual process of each user. The other awareness tools such as screenshot and speech-to-text note indicator and recording status indicator are still shown to all users. We additionally visualized teammates' ray cursor with the related avatar's color to share user cursor's movement and operation. 4) *Viewport visualization*: the visualized gaze is always displayed during the video (same as the under follow function of Non-sync mode but still keeps the shared control), since the users are always on the same timeline progress.

View sharing: Because all users have shared progress, they can easily check what other users are viewing with the visualized viewport. Thus, there are no view sharing functions in Sync mode.

Note-taking: For note-taking, the screenshot and speech-to-text note work the same as in Non-Sync mode, and the drawing is shown to users all the time because users are on the same timeline.

3.2 After-Video Discussion Platforms

Based on the DCVV model [11], the discussion about learning content happens during and after the video. However, after-video discussion platforms are not specifically investigated by prior work [79]. We designed a VR discussion room to study how the collaborative watching experience can facilitate the after-video discussion and promote a meaningful learning process. For the after-video discussion platform, we considered two main design aspects: co-presence awareness and shared context, to be detailed below. An example of the in-VR discussion room is shown in Figure 2. To make our conditions comparable and answer our RQ1, Zoom and Google Slides are used as after-video discussion platforms in the Basic mode as the status quo online discussion setting. We selected this setting because it fitted the definition (existing widely-available technologies) of baseline condition to constitute a replication work.

Co-presence awareness: Maintaining users' awareness of other users' social presence is essential for a collaborative environment [129]. Previous work has established that 3D avatars can create a strong sense of presence in virtual environments [6, 41] and gestures provided by the 3D avatars can improve remote instruction performance [57]. In our discussion room, we implemented *3D avatars* with colors representing each user. Users can move freely inside the discussion room. To minimize the risk of falling or hitting in the physical world, we chose to use *teleportation* as means of navigation in the discussion room, so that users do not need to move physically.

Shared context: Providing shared contexts for all users around the discussion topic is one key component for a successful discussion [30, 42]. The screenshots (optionally with attached drawing strokes or with speech-to-text notes) users took during the video were displayed for all users to create a shared after-video discussion context. We allow users to *rescale* and *draw* to annotate the

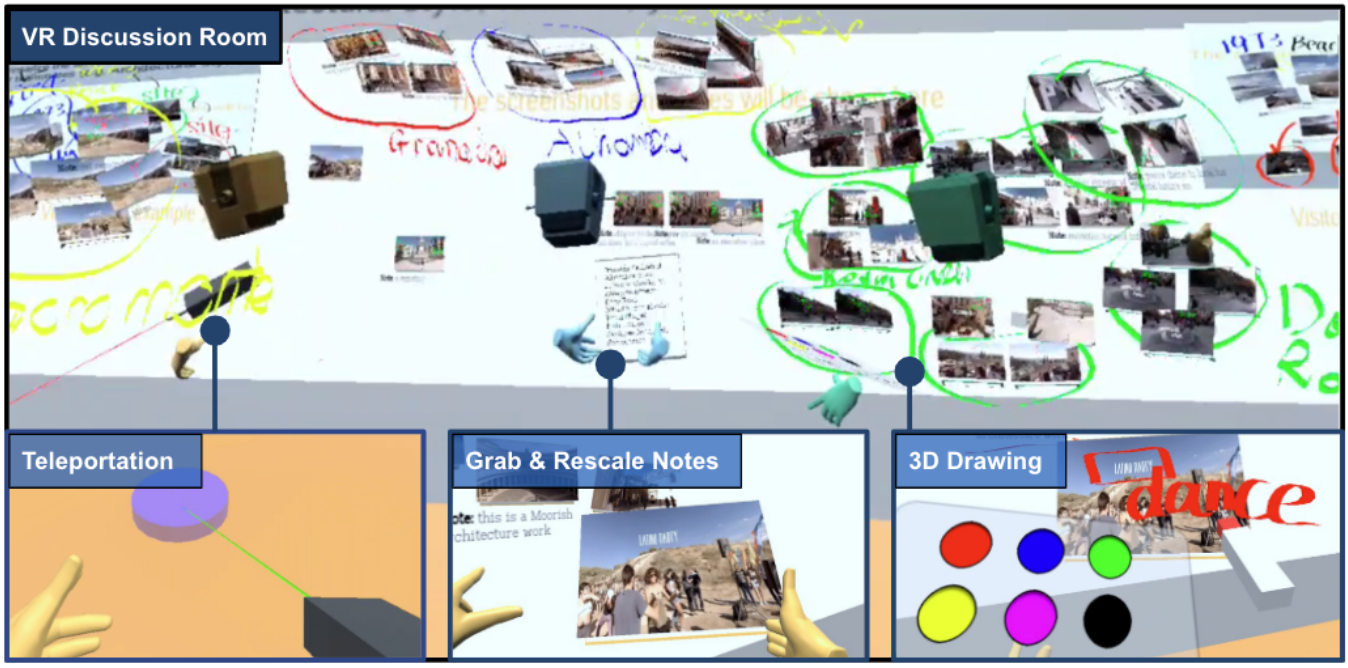

Figure 2: An example of one group (G10) creating the visitor guide with three primary functions we provided in the in-VR discussion room.

screenshots to build connections and synthesize the educational content.

4 METHODS

This study was reviewed and approved by our university's IRB. In order to examine the influence of different types of collaborative technology on the perceptions and experiences of online learning, we conducted a three conditions within-subject experiment with 54 participants (18 groups (trios)). Each trio would work together to finish the same learning task three separate times under different conditions: Basic mode, Non-Sync mode, Sync mode. There were two reasons we selected the within-subject design. First, if the participants only had one condition they wouldn't be able to compare it with other types of conditions as the within-subject design allows participants to compare different conditions and give more insights in the interview [11]. Second, because collaborative learning experiences are significantly influenced by teammates, keeping the same trio could avoid the differences caused by changing teammates. The primary disadvantage of within-subjects design is the carryover effects [109], which means the order of the conditions (i.e., the three different modes) can affect outcomes and results. To deal with that issue, the conditions were provided in a counterbalanced order, while the order of the used videos was always the same. That means each video would be used in three different modes for this study.

Before the formal study, we conducted a pilot study with three groups (two groups with two people and one group with three people). The goals were to 1) test the assigned systems' functionality, 2) improve the system usability, and 3) optimize the study protocol. We found that unlimited time of after-video discussion increased users' fatigue and ineffective collaboration. We also identified some usability issues with note-taking. As a result, we 1) restricted the after-video discussion duration from unlimited time to seven minutes to better manage the study duration; 2) reduced the distraction by muting users when they start the speech recognition note function and separating the channel for the drawing annotation in the Non-Sync mode, and 3) added scale function for screenshots at VR discussion room in order to improve their readability.

4.1 Learning Task and Materials

We chose a virtual field trip as the learning context for this study because the ability to access remote locations is a major reason why people choose to use VR in education [53]. We worked with experts in educational psychology and curriculum and instruction from a mid-western university in the United States during the learning task and learning goals creation, video selection and video reprocessing process.

All 54 participants were told that they would have a virtual field trip in VR with their remote classmates in a small group. For each learning unit, we set the task of designing a visitor guide of the given city for students who missed this field trip in order to motivate their discussions. In order to complete the visitor guide, a group would watch the same VR video simultaneously, and then work together to summarize the attractions and their historical and architectural features after the video. The discussion was allowed during and after the video. The learning goal was to identify key attractions and their historical and architectural features, and to understand the culture and history of the given cities.

The videos used for the virtual field trip came from the same collection "One day In" 360° travel videos [73]. This collection

of 360° travel videos includes more than 70 destinations around the world, depicting the culture, architecture, and attractions in each city. We picked three narrated city-tour videos for the formal study (Granada [70], Porto [71], Seville [72]). We used an extra city tour for the tutorial session. All of them were created by the same aspiring videographer and were almost the same in terms of editing and narration style. We reprocessed the videos in three ways: 1) edited the videos to equalize time duration; 2) balanced the difficulties among the videos by revising the video narrator's wording, and 3) replaced the original background audio with the same AI-generated narrator (an American female voice narrator). Each final video covered historical and architectural features of a European city, lasted about 3.5 minutes and consisted of 14 video clips. The texts used in the videos contain the following word counts: Granada 451, Porto 478, and Seville 514. The resolution of videos is 1920 x 1024 pixels.

4.2 Participants and Settings

A total of 18 groups of three people (54 participants) from a mid-western university in the United States were recruited for this study. We formed teams with three students each because it is a good representation of both pair and small group activities in the classroom. Participants were grouped into these teams based on their time availability to participate in this study and because they could also specify teammates if they wanted to participate in this study with their friends. No prior VR experiences or 360 video watching experiences were required, but we used the answers to VR experience questions in the screening questionnaire to ensure the diversity of the participants' VR-related backgrounds. We elected to run this study with a diverse sample of participants (e.g., major, gender, prior VR experience, familiarity with their teammates) because it increased the confidence so that the results could be generalized to a diverse population of students and therefore provide more valid suggestions to apply our system to a real classroom.

Our participants were between 19 and 31 years old (M = 22.5, SD = 2.65); 26 were females, 27 were males, and one preferred not to disclose gender. All participants were college students (39 undergraduate students and 15 graduate students). 26 participants indicated that they did not know each other before the study, 18 knew one person on their team, and ten knew two people on their team. 14 participants had rich VR experience, 26 had few VR experiences and 14 without any prior VR experience. Prior to the study, 23 participants had experience watching 360 videos using VR headsets, 12 participants watched 360 videos but not using the VR headsets (e.g., laptop, mobile phones), and 19 participants did not watch any 360 videos before. All participants stated that their vision was normal or corrected to normal. Each participant was compensated with $50 gift card.

The study took place in four adjacent rooms with similar settings inside the university (three rooms for participants and one for a researcher as an in-VR tech-facilitator). Each room contained a swivel chair, a desk, one laptop, one VR headset and a camera on a tripod, which captured the whole process.

4.3 Procedure

The study would last three hours for each group in total (See Figure 3). Participants received and filled out a demographic questionnaire and consent form online before the study. After the trio was ready, researchers first led a 5-minute ice-breaking activity with all three participants in the same room. Participants would introduce themselves and talk about their impressions or experiences with VR. Then researchers gave a 5-minute introduction, explaining the goal, procedure, learning task of the study, and then giving an overview of system features they would use.

Next, participants put on the headset and started the system tutorial using a city tour video from the same series (which is not used in the formal study). During the training session, one researcher guided the participants in the assigned VR system. The other two researchers helped the participants learn the physical operation (e.g., finding specific buttons on the controller; adjusting the headset straps). We left adequate training time for participants to get familiar with every function of the three different modes. This process took about 20 to 30 minutes based on participants' prior experiences with VR.

After the training session, the participants were guided individually by a researcher into separate rooms, where they had a 5-10 minute break. Once every participant confirmed they did not have cybersickness and were ready for the next session, they started the knowledge assessment as a pre-study session. Otherwise, we waited until they did not report a strong cybersickness. Once the participants completed the pre-study session (knowledge assessment, see Section 4.4.1), they put on the headset and entered the learning unit. Each learning unit contained two parts: a during-video session and an after-video session. At the first 10-minute during-video session, participants would watch a 3.5-minute video with an assigned VR video system. During this session, the participants could pause, replay, and relocate the video as much as they needed. When in the Basic mode, they communicated with each other via Zoom and took notes using pen and paper (See Figure 4b). In the other two modes, team communication and note-taking were integrated into the VR environment (See Figure 4c). When they entered the after-video session with assigned discussion platform, they had seven minutes to complete the learning task of creating a visitor guide. In Basic mode, this session was done via Zoom and Google Slides, while in sync and Non-Sync mode, it was done via an in-VR discussion room. In addition, a researcher also joined the VR system with participants from a separate room to rectify the tech issues, if any. Another researcher would go to the participants' room to help adjust the physical devices.

Participants started the post-study session when they completed one learning unit. Their knowledge acquisition would be measured with the same knowledge assessment they were tested with before. Then the participants answered the self-reported questionnaires (collaboration experience, social presence, cognitive load and satisfaction, see Section 4.4.2). A total of three learning units (each related to a technology mode) with pre- and post-study sessions were conducted. Between two learning units, there was a 5-10 minute break to make sure they did not have strong cybersickness when reentering the VR. We instructed our participants that they should inform us if they experienced cybersickness. No participant reported severe cybersickness after the break, and all completed the task. Thus we did not define cybersickness as a significant factor in this study. Finally, each participant was interviewed by one

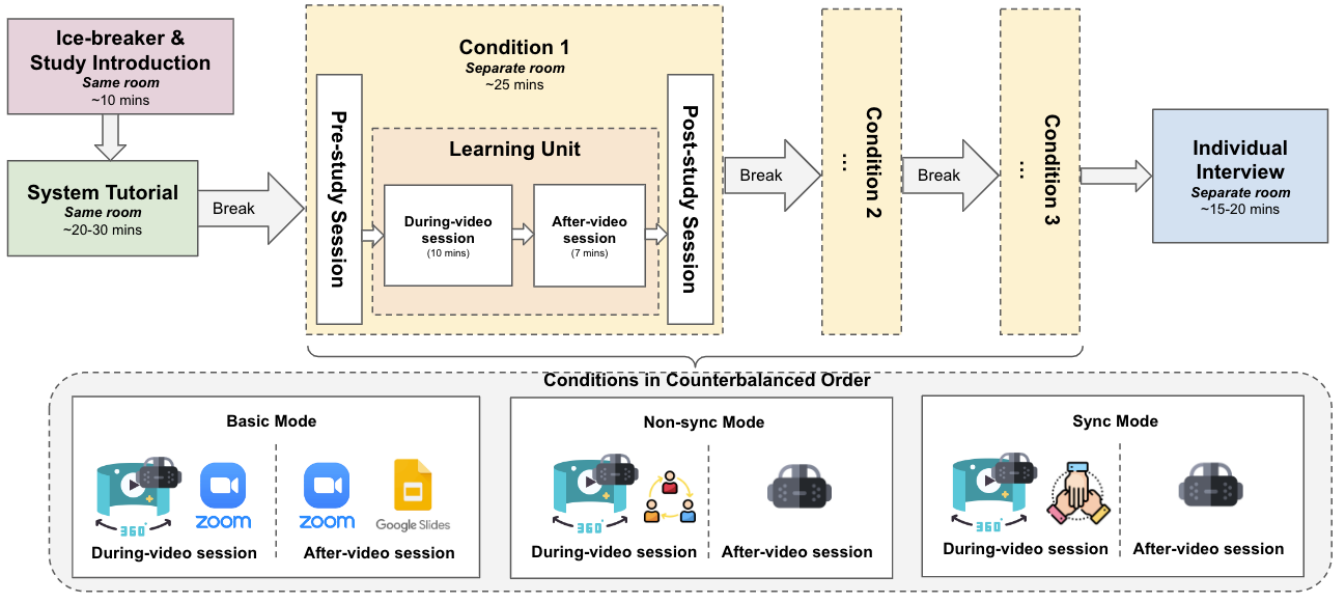

Figure 3: Overview of the study protocol. Each condition contained one specific technology mode (Basic Mode, Non-sync Mode or Sync Mode). Three conditions kept the same procedure and the mode order was alternated in a counterbalanced order. The entire procedure took about three hours to complete.

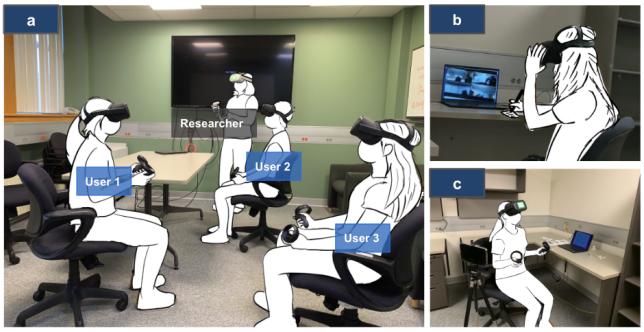

Figure 4: One researcher guided a group to do the system tutorial (a). A user stayed in a separate room to learn from the virtual field trip in the basic mode (b) and the other two collaborative VR modes (c). All participants used the stationary guardian by default.

researcher individually in a separate room (more details about the individual interview see Section 4.4.4).

4.4 Measures and Data Collection

This study used a mixed methods approach. Both qualitative and quantitative data were collected for analysis. Our goal was to gain a holistic view of participants' behaviors and experiences under different technologies through knowledge assessment, self-reported questionnaires, log files, and final interviews to answer our RQs.

4.4.1 Knowledge Assessment. To assess the participants' knowledge acquisition as a result of experiencing three technologies, we developed a multiple-choice test, which was completed before and after each learning unit. This test contained questions regarding the information presented during the video. The test was developed with experts in educational psychology and curriculum and instruction. The assessment included conceptual and factual questions according to Bloom's taxonomy [3]. We measured factual knowledge based on prior educational VR video research [102], separating the questions into auditory and visual based questions of content described in the video.

Our initial version of the assessment had 12 questions for each of the three videos (Granada, Porto, Seville). For each knowledge question, four possible answers were provided with a single correct factual based answer. Prior to the main study, we conducted a pilot study of assessment to examine if the knowledge assessment 1) was understandable for students; 2) had similar difficulty across the three videos' tests and 3) would not achieve a ceiling effect. For the first pilot study, we recruited 10 participants through a university email list for students and personal connections who joined the pilot study voluntarily without any preparation. Participants had 10 minutes per video using a desktop 360-degree video player. They were able to take notes with pen and paper while watching the video, but they could not go back to see the notes when starting the assessment. The mean of the final scores (%) of videos was 45.83% (Granada), 50% (Porto) and 66.67% (Seville). Students were able to understand every question and there was no ceiling effect for the assessment. We averaged the difficulties by removing the higher incorrect responses (at Granada and Porto) or lower incorrect responses (Seville) that did not significantly correlate with the overall score. The final knowledge test, therefore, had 10 questions (4 auditory, 3 visual, and 3 conceptual questions). The mean of the final scores (%) of the remaining questions of each video were 58.33%

(Granada), 54.17% (Porto) and 54.17% (Seville), which attested to the similar difficulty across the learning materials.

4.4.2 Self-Report Questionnaires. The study was designed to assess how different types of technology influenced the participants' experiences on several measures: collaboration, social presence, cognitive load and satisfaction. Prior work has identified those variables as important factors influencing learning and learning outcomes [17, 21, 23, 48, 112]. All questionnaires were done, individually, online, in separate rooms to ensure that participants would not influence each other.

Four variables (collaboration, satisfaction, social presence, and cognitive load) were measured from a total of 25 questions. Two measures, collaboration experiences and satisfaction, were evaluated using question items from So and Thomas' work [112]. The collaboration experience questions were designed to identify how types of technologies influence collaboration. At the same time, measuring satisfaction is an effective way to indicate "the degree of learner reaction to values and quality of learning, and motivation for learning." Social presence survey [59] measures the subjective experience of being present with their teammates. Measuring social presence is essential here because theory suggests that it is a central mechanism that leads to deeper cognitive processing and consequently better learning outcomes [83]. Additionally, cognitive load was measured by [15], including intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL). More specifically, ICL is affected only by the learning content but not by the instructional design. ECL and GCL are caused by instructional design. Unlike ECL, GCL is beneficial for learning, because it is directed to schema acquisition by directing the learner's attention towards relevant learning processes [115]. The Cronbach's α values of questionnaires of the collaboration experience, satisfaction and social presence are 0.860, 0.779 and 0.891 respectively, indicating these questionnaires are reliable. We didn't measure the internal consistency for the cognitive load questionnaire because we reported each item separately in the results.

4.4.3 Observation Notes and Log Files. In addition to the above assessment and self-report scales, we wanted to dive deeper into the measurement of collaboration experience in the during-video session. To do this, we observed and noted each user's discussion time, and logged head-tracking data. Two researchers observed and coded the conversation that was happening between participants during the video into a voice chat transcription based on the recorded video. We captured the duration of during-video discussion and recorded the timestamp. We analyzed the log data to find times where two users or a group shared the same video context, which is a basis for collaboration [79]. We adopted the measured shared focus metric in collaborative VR analytics research [20] to determine users' headset view similarity. Two people's views were considered similar if no angles between two head orientations were more than 45° (half the horizontal field-of-view of the three headsets used in the study) and the difference between their video timeline positions was less than three seconds (determined empirically for the selected video materials). Then the group headset view similarity was computed by dividing the total time three users had similar views by the total task time. Pairwise view similarity used the same method but only calculated with every two participants' head tracking data.

4.4.4 Semi-Structured Interviews. Although our RQs didn't have a specific qualitative component, to increase our confidence in the findings, we triangulated the validated measures with qualitative data from semi-structured interviews [7]. We also use qualitative themes to explain the quantitative results we collected. We elected individual interviews instead of group interviews in order to ensure everyone got the chance to express their feelings and would not be influenced by others. We asked the participants to describe their collaboration strategies and experiences and reflect on each mode. We also asked about their satisfaction and preference for each mode and its related collaborative tools. Interviews were audio recorded and transcribed for analysis (see below). Each interview lasted 15 to 20 minutes.

4.5 Data Analysis

In order to answer the two research questions, we analyzed the quantitative results from the knowledge assessment, self-report questionnaires and log-ins, and the qualitative results from the interviews. For the quantitative data, we conducted one-way ANOVA for those three modes (Basic, Sync, Non-Sync). We ensured that the statistical analyses were performed properly and fitted the requirements and assumptions of the test procedures. For the ANOVA, for example, we applied within-subject statistics, which are used to compare observations of an outcome for the same person being measured at different time points (modes). In order to examine the significant effect of the ANOVA results, we conducted specific condition comparisons using a T-test at the 0.05 significance level with Cohen's d to measure the effect size.

For the qualitative data, we analyzed the interviews following the guidance of data-driven thematic analysis (inspired by the Grounded Theory Method [78]). We first transcribed the recording to textual data. Then the five authors who were involved in the interview process and trained in the Glaser method [35] open-coded the transcriptions and added memos based on their notes from interview sessions. We generated over 2000 open codes through this process based on 54 individual interviews. Two of the five authors (the first two authors) then read the open codes. If there was ambiguity or disagreement in the code/memo, they returned to the transcription to verify or talked with the other three authors. The same two authors then worked together to cluster the open codes using a constant comparison approach operationalized as an affinity map, placing open codes that have similar meanings together. Clusters were iteratively refined until themes emerged. If the first two authors could not reach an agreement about the placement of a particular code or cluster within a theme, they would seek out a third author for discussion, or a fourth author if necessary. Once the iterative analysis process was completed, all authors discussed the significance and novelty of themes to answer our research questions, with the most relevant themes presented in the results section below.

5 RESULTS

5.1 Quantitative Results

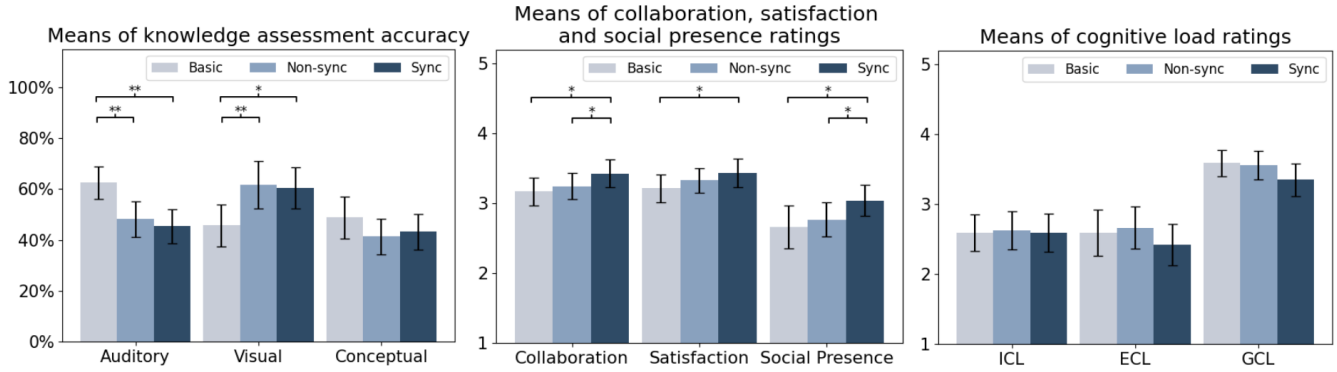

Figure 5: Left: Means of knowledge assessment accuracy (divided by auditory, visual and conceptual questions) by technology mode. Middle: Means of collaboration, satisfaction and social presence ratings by technology mode. Right: Means of cognitive load ratings (divided by intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL)) by technology mode. Error bars show 95% confidence intervals. Asterisk (*) indicates a statistically significant difference between conditions: $p < .05$ (*); $p < .01$ (**).

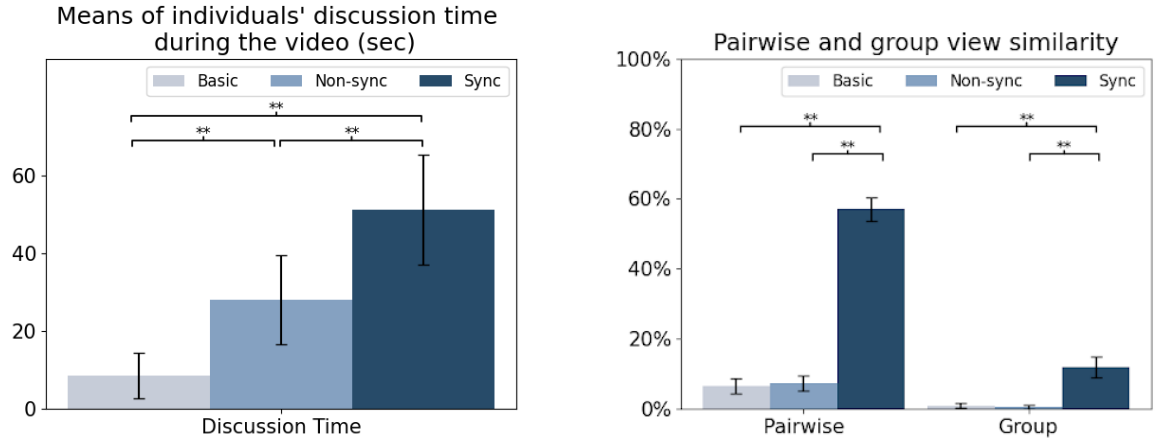

Figure 6: Left: Means of individuals' discussion time during video (N = 54). Right: Means of pairwise view similarity (N = 54) and group view similarity (N = 18) by technology mode. Asterisk (*) indicates a statistically significant difference between conditions: $p < .05$ (*); $p < .01$ (**).

Our ANOVA tests revealed that there was a statistically significant difference in means of auditory and visual knowledge acquisition scores (auditory: F (2, 106) = 8.994, p = .0002, visual: F (2, 106) = 4.839, p = .009), collaboration (F (2, 106) = 3.508, p = 0.03), satisfaction (F (2, 106) = 3.720, p = .02) and social presence ratings (F (2, 106) = 5.202, p = .006) between at least two groups. The results of the T-test for multiple comparisons are reported below.

5.1.1 Knowledge Acquisition. We measured knowledge acquisition based on pre- and post-knowledge assessing questionnaires. Questions in the questionnaires were divided into three knowledge categories: audio-based factual, visual-based factual, and conceptual. The means of the visual knowledge acquisition scores in the two collaborative VR conditions were both statistically significantly higher than the Basic mode[5] ($p_{(B,N)}$ = .004, $d_{(B,N)}$ = -.492, $T_{(B,N)}$ = .305; $p_{(B,S)}$ =.012, $d_{(B,S)}$ =-.491, $T_{(B,S)}$ = 2.605). The observation is not surprising given the inherent benefits of users taking and organizing screenshots during videos. In contrast, the mean of the audio-based knowledge scores was statistically significantly higher in the Basic mode ($p_{(B,N)}$ = 0.002, $d_{(B,N)}$ = .588, $T_{(B,N)}$ = 3.181; $p_{(B,S)}$ < 0.001, $d_{(B,S)}$ = .721, $T_{(B,S)}$ = 3.748).

There was no significant difference for conceptual knowledge. Thus, it seems the existing technology (Basic mode) was better for learning audio-based factual knowledge while our proposed VR delivery (Sync and Non-sync mode) outperformed for memory of visual information. We found no significant difference between the results of Sync and Non-sync mode in the three knowledge categories.

[5]Throughout this section, we denote Basic as B, Non-Sync as N and Sync as S

5.1.2 Collaboration. We measured participants' collaboration experiences in learning units through the self-reported questionnaire. The results of this questionnaire showed that the mean index of collaboration experience (Figure 5) in Basic mode was statistically significantly lower than that of the Sync mode ($p_{(B,S)}$ =.027, $d_{(B,S)}$ =-.354, $T_{(B,S)}$ = 2.269), while it was approximately equal to that of Non-sync mode, with no statistically significant difference. Our results showed that there is a significant difference and a small effect size ($p_{(N,S)}$ = .049, $d_{(N,S)}$= -.26, $T_{(N,S)}$ = 2.008) between the collaboration score of Non-sync ($M_{(N)}$ = 3.24, $SD_{(N)}$ = .69) and Sync mode ($M_{(S)}$ = 3.43, $SD_{(S)}$ = .73), indicating that the Sync mode can provide a better collaboration environment and experience to the participants than Non-sync mode.

We further analyzed participants' collaboration behaviors during the video by coding the observation notes and log files, which included their discussion duration and headset tracking data (See Fig. 6). Both results listed below indicated that two collaborative VR video modes statistically significantly promoted discussion than Basic mode, while only Sync mode statistically significantly promoted both view similarity and discussion time during the video than the Basic mode and Non-sync mode.

Individual's discussion time. Our results show that the mean of the discussion time in Sync mode was statistically significantly longer than the other two modes ($p_{(B,S)}$ < .001, $d_{(B,S)}$ = -.588, $T_{(B,S)}$ = 2.965; $p_{(N,S)}$ < .001, $d_{(N,S)}$ = -.491, $T_{(N,S)}$ = 3.211). At the same time, Non-sync mode was statistically significantly higher than the Basic mode ($p_{(B,N)}$ < .001, $d_{(B,N)}$ = -1.08, $T_{(B,N)}$ = 6.013).

Head tracking data. We used head tracking data to measure the view similarity. For both pairwise and group view similarity, Sync mode produced a statistically significantly higher rate than the Basic mode ($p_{(B,S)}$ < .001, $d_{(B,S)}$ = -4.737, $T_{(B,S)}$ = 26.975; $p_{(S,N)}$ < .001, $d_{(S,N)}$ = -4.728, $T_{(S,N)}$ = 32.892). Both types of view similarity were almost equally low for the Basic and Non-sync mode conditions, and the results were not statistically significantly different ($p_{(B,N)}$ = .676, $d_{(B,N)}$ = .078, $T_{(B,N)}$ = .420).

5.1.3 Social Presence. Social agency theory proposes that social presence during multimedia learning is a central mechanism that leads to deeper cognitive processing and consequently better learning outcomes [83]. While we did not see statistically significant differences between the Basic mode and Non-sync mode in Fig. 5, we observed that the Sync mode yielded higher value with statistical significance compared to that of the Basic mode for the social presence score ($p_{(B,S)}$ = .019, $d_{(B,S)}$ = -.383, $T_{(B,S)}$ = 2.639). Similarly, our results show statistically significant differences between Non-sync and Sync in the social presence with a small effect size ($p_{(N,S)}$ = .009, $d_{(N,S)}$ = -.317, $T_{(N,S)}$ = 2.734). The results indicate that participants feel the most social presence of their teammates in Sync mode.

5.1.4 Cognitive Load. Ideally, the extraneous cognitive load (ECL) should be reduced, the germane cognitive load (GCL) maximized, and the intrinsic cognitive load (ICL) controlled [23]. Our results did not show statistically significant differences with all three categories. However, there was a small effect size ($d_{(B,S)}$ =.31) of Basic mode and Sync mode on GCL, which participants invested more resources and more concentration in the learning process in Basic mode than in the Sync mode. There was a small effect size of Non-sync and Sync mode on ECL ($d_{(N,S)}$ =.222) and GCL ($d_{(N,S)}$ =.255) based on benchmarks suggested by Cohen [18], indicating that participants in Non-sync mode might better concentrate on learning than Sync mode, while Sync mode might be the easier tool to learn with than Non-sync mode.

5.1.5 Satisfaction. User satisfaction metrics are important to improve values and quality of learning, and promote motivation for learning. We measured satisfaction based on the self-reported questionnaire and collected participant preferences in the interview. As shown in Fig. 5, the mean score for satisfaction is (M_B = 3.21, SD_B = 0.72; M_S= 3.43, SD_S = 0.76; M_N = 3.33, SD_N = 0.63), and the difference between the Basic and Sync mode was statistically significant ($p_{(B,S)}$ = .024, $d_{(B,S)}$ = -.296, $T_{(B,S)}$ = 2.317). Results showed Basic mode received the lowest satisfaction. We also asked participants about their favorite mode. 22% of students (N = 12) indicated that their favorite mode was Basic mode, 35% preferred Non-sync mode (N = 18) and 43% preferred Sync mode (N =24). A chi-square test of independence showed that there was no significant difference between the expected results and real results ($X2$ (2, N = 54) = 1.837, p =.399).

5.2 Qualitative Results

We processed qualitative data by the data-driven thematic analysis and numbered the most important themes which depict a picture about the tension of learning pace flexibility and communication comfort, how participants perceived collaborative tool and after-video platforms.

5.2.1 Theme1: A tension existed between learning pace flexibility and communication comfort. We observed a tension between learning pace flexibility and communication comfort when learning from a video together. Participants appreciated the ability to customize their pace of learning in Basic and Non-sync mode but noted the communication suffered by layered sound of voice chat and learning video, and lacked shared content. In contrast, Sync mode was perceived positively on communication comfort but may reduce learning efficiency.

Many participants noted the flexibility of individual control in Basic mode because they could *"pause and repeat the video at their own pace."(14C)* However, 46 out of 54 participants reported that they chose to work individually and only had a few, if not none, discussion while watching the video in Basic mode (this result consisted with our observation notes). The major reason noted by participants that hindered their from communication was *layering sound from multiple devices.* Although Zoom was provided as a communication tool in Basic mode, it was not perceived useful and comfortable to initiate a conversation. Participants heard the video's narrator from others' headsets via Zoom, which distracted them from focusing on their own content. For example, 6C described her frustrated experiences in Basic mode: *"echos that really, really disturbed me. So I had to volume down the voice of the laptop... It isolated me from the group, thus we haven't had much communication"(6C).* Correspondingly, participants stated that *lacking shared content* was another issue reducing their enthusiasm to have further discussion. For example, 6A stated that she didn't talk in Basic mode because *"I didn't know where were they and if they know my questions" (6A).*

In this way, watching VR video with existing technology was more like an individual experience.

Almost all participants admitted that they communicated better in collaborative VR video modes with in-VR communication and the tools to know others' statuses. Non-sync mode still supported *"flexible and customized learning speed like Basic mode"(12A)* but *"got the feeling that I was in a team and worked together with others."(14B)* However, participants noted that the issue of lacking shared content remained in the Non-sync mode. When asked how they felt about the amount of discussion that happened when watching the video in Non-sync mode, many participants said it was insufficient. The reason was similar with Basic mode – *"Everyone was focusing on a different part of the video. They know little about my content. And therefore we cannot start meaningful discussions or get real-time responses."(1C)* For the same reason, they were afraid of *"interrupting other's process"(12A, 14A, 4A)* and also felt *"interrupted if comments from different timeline"(12C).* Agreements were needed with non-synchronized video control in order to obtain an overlapped context. They can either drag the progress bar to the same timestamp or use view sharing tools. Worth noting, despite those ways were feasible in practice, most participants reported that they rarely used them because *"I was too busy focusing on my part to talk"(8C)* and *"it's inconvenient for three of us to immediately arrive at the same location."(6C)*

We received most positive feedback about the communication comfort in Sync mode. It's more natural to ask a question or initiate a conversation when group members *"were all there together and on the same page"(2A).* Many participants mentioned that they were also more likely to get active response because *"my teammates would definitely know the context"(8B)* and *"they can hear everything well because we all paused for now"(6A).* It was worth noting that the shared content was at the expense of time flexibility. Some participants said the Sync mode slowed down their pace because they *"have to be held on to other people, having to pause and go to where other users want to"(12A).* Participants of Group 9 stated that they even didn't finish the video because they paused too many times and had too much conversation. A few participants mentioned that sometimes they were hesitant to pause the video because of the effect on other participants, especially working with strangers (which was also noted in prior video co-viewing studies [11, 116]). But interestingly, most of them didn't think it was a big problem to interfere with their communication and collaboration within their group. One participant noted a potential reason: *"the (teammate's) cursor's position told me that they also want to pause right now. So I just paused."(12A)* .

5.2.2 Theme2: Shared control influenced the perceived usefulness of collaborative tool. Participants reported that how they perceive the usefulness of collaborative tools (awareness, view sharing, note-taking) with or without shared video control was different. First, awareness. Almost all participants highlighted the importance of activity visualization in both Non-Sync and Sync mode as *"knowing other's status was the foundation of collaboration."(13C)* More specifically, activity visualization in Non-sync mode seems more relevant to their collaboration strategy. For example, one participant said *"if I see a large chunk of video without any notes (indicators) here, I will jump into that part to see my teammates missed something."(6B)* Other awareness tools in our study were perceived less useful. One participant commented on the specialized voice that *"hearing others voice clearly was much more significant than knowing where the voice came from."(3A)* Many of our participants said they even didn't notice the voice were specialized by others' viewport or position. For the viewport visualization, participants reported that it improved the co-presence but also *"interfered watching experiences" (1A,1B,2C,5B,18A)* because of too much visual input in Sync mode. Similarly, participants thought embodied visualization in Sync mode was fun but *"eyes were too busy on the video, so most times didn't pay attention to it"(5B).*

Second, view sharing. The three view sharing functions (Peek, Peek in full window and Follow) were only implemented in Non-Sync mode to provide shared content crossing different timeline. Interestingly, participants reported that they didn't use them much. Only 10 out of 54 participants mentioned that they used peek function. Most of they described the reason why they used this tool similar as *"just curious about what they are looking at"(14A)* rather than seeking a specific learning content for discussion. Even fewer participants tried to peek in full window because it *"less easy than peek"(12B)* and "felt weird to look from other's perspective"(2A). Most people didn't use follow function because it interrupted their learning pace. For example, 8C stated, *"peek and follow others would bring me to a new content suddenly, which is hard to catch up."(8C)* One positive example of using follow function was group 3, where they knew *"one person (3A) really good at the topic, so we can follow and see what where she was looking at and listen to her"(3B).*

Third, note-taking. Almost all participants indicated taking screenshots was most helpful and convenient. Participants agreed that screenshots helps them better remember those visual knowledge. Speech-to-text note was also helpful but some participants were less likely to use them in Sync mode than Non-sync mode because *"it took more group time to record and transcribe"(11A).* Participants appreciated speech-to-text notes in Non-Sync mode because they could pause the video without interrupting others. Drawing tool was usually used for writing down simple numbers or words. As 14B noted, *"it's hard to write notes with wrist (controller) because user is used to writing with fingers"(14B).* It was perceived more useful in Sync mode as people used it for pointing out specific visual features in the video and the potential of co-creating sketches within the group. Some users compared the drawing tool with the way they used in the Basic mode and hoped drawing tool could *"as nature as taking notes on pen and paper"(8C).*

5.2.3 Theme3: In-VR platform for after-video discussion enhanced visual transmission and engagement. Most participants agreed that both after-video platforms were helpful for them to solidify the information they got from the video because it *"it was a good way to recall what we learned from video and to create visual."(7B)* Regarding the conventional platform used in Basic mode, participants pointed out a gap of visual transmission "from the headset to the computer" and "within the group". Participants complained about the out-VR platform such as *"I was basically cannot record the visual information from video and then use it in the slides."(13C)* Without screenshots, people found it difficult to see *"if we were talking about the same thing without pictures."(15C)* and had to *"describe the object by words"(16A).* Participants also found that the conventional setting was insufficient to foster an engaging collaboration because of

"lacking real body interaction"(11B) and *"cannot feel much connection with others"(12C).*

In contrast, participants appreciated the in-VR platform's power to transmit and display visual notes and support engaging teamwork. When comparing in-VR platform with out-VR platform, 17B said *"if there were screenshots, you can shift focus from describing the basics to the fact of the object, which makes a difference."(7B)* Participants reported that they had more engaging teamwork on in-VR platform for after video discussion because of the ability to support embodied interaction, including *"spatially organizing the notes"(4B)* and *"building the connections by 3D drawing"(14A).* Some participants also noted the issue of managing stokes and screenshots in the VR discussion room (e.g., *"I could only grab one note at once, which was not as efficient as we can do in slides."(1C)*). Some participants also indicated that the current in-VR platform could be benefited from more intuitive designs. For example, 3A said, *"in-VR discussion room, everyone is the same robot but different color, would be better with a different and more realistic appearance."(3A)*

6 DISCUSSION

Our work presented two collaborative VR video viewing technologies (Sync and Non-sync mode) and compared their effectiveness for online education on five measurements (knowledge acquisition, collaboration, social presence, cognitive load and satisfaction) against an existing technology (Basic mode). In this section, we first summarize our principal results, then reflect on our method with two aspects: running studies of remote VR collaboration and reflecting on replication in HCI. Finally, we identified three design guidelines to better design distributed collaborative VR video viewing systems for educational purposes.

6.1 Principal Results

Our first RQ focused on *VR video delivery via existing technology (Basic mode) compare to collaborative VR video delivery proposed in this study (Sync and Non-Sync mode) on measures of knowledge acquisition, collaboration, social presence, cognitive load and satisfaction.* Our quantitative results showed that the collaborative VR-based systems (Non-Sync and Sync) both achieved statistically significantly higher scores on the measures of visual knowledge acquisition, collaboration, social presence, and satisfaction, compared to the baseline system, with moderate to high effect sizes. The three systems did not show differences based on the level of cognitive load and conceptual knowledge acquisition. We used qualitative themes to triangulate and explain the quantitative results. Participants (Section 5.2.1) reported the potential reasons, such as lack of shared context and current technical obstacles (e.g, echos), for lower scores of Basic mode on collaboration and satisfaction. They also appreciated the in-VR platform's power to transmit and display visuals for after-video discussion (Section 5.2.3), which explained the potential reason for lower scores of Basic mode on the measures of visual knowledge acquisition.

Our second RQ considered *how individual VR video control (Non-Sync mode) compares with shared video control (Sync mode) on measures of knowledge acquisition, collaboration, social presence, cognitive load, and satisfaction.* Our study found that the shared control in Sync Mode significantly increased the ease of collaboration and sense of social presence. There were no statistically significant differences between the two modes on other evaluation metrics (knowledge acquisition, cognitive load, and satisfaction). The qualitative results confirmed that better collaboration experiences with shared control in the Sync mode were due to better communication comfort (Section 5.2.1). The results also revealed the tension of communication comfort and learning pace flexibility (Section 5.2.1) and the control method would influence the perceived usefulness of collaborative tools (Section 5.2.2).

6.2 Method reflection

6.2.1 Running Studies of Remote VR Collaboration. Our study focused on a remote collaborative learning scenario and was conducted in a lab-based environment, where researchers and participants were co-located (but in separate rooms). Our choice of using a simulated remote environment instead of an "in the wild" study is based on the following considerations: 1). *Avoid potentially biased participant groups.* In a general class, students would be more likely to have a diverse technological and demographic background. However, if we were conducting "in the wild" VR studies online, one requirement for participants is that they need to already have VR headsets and related experience[104]. Prior research [95] also mentions that when recruiting online VR participants, the majority of participants may be young men, which will introduce bias in the demographic distribution. 2). *Control the infrastructure differences.* Different infrastructures (e.g., network condition, headset) may introduce one more dimension of interference, especially for a collaboration setting. For example, users with lower bandwidth or more fluctuating networks may have a worse collaboration experience. Therefore, we set up three rooms with similar physical environments and the same network coverage to reduce the influences caused by infrastructure. 3). *Facilitate collaboration settings.* To maintain a smooth procedure and effectively manage time, researchers need to respond to participants' questions and issues in real time. In an online setting, facilitating tutorials and troubleshooting can be time-consuming because of inaccessible headsets and no-backup devices. To avoid that issue, we chose a lab-based remote setting, which had one researcher in each separate room for troubleshooting. Extra devices are stored on-site in the case of hardware issues. Future researchers should also consider the above trade-off between an "in the wild" study and a simulated in-lab study to run a collaborative VR study.

6.2.2 Reflecting on Replication in HCI. The importance and necessity of replication in science have been well-established, though rarely put into practice in HCI [122]. The comparison of Basic mode and Non-sync mode of our work performed a partial replication of prior work [79]. We re-implemented and used deliberate modifications of earlier research, with the aim of testing it on education domains with different participant groups and measures. More specifically, we adapted awareness, view sharing tools and note-taking tools from prior work and added in-VR discussion platform for after-video discussion. The original work focuses on in-person filmmaking and reviewing in an in-person setting while our system targets educational purposes, mainly in the remote setting. For the evaluation, the original work involved 5 participants, divided into two groups, and used a within-subject design, comparing the designed system with a baseline condition. Both the original work

and our work collected users' discussion time, view similarity and self-perception of collaboration through the log files and questionnaires as evaluation measurements. However, in our investigation, we significantly expanded the depth and scope of prior work. We included 54 participants and applied the within-subject method comparing three modes. We evaluated learning outcome, satisfaction, social presence and cognitive load from users and included a semi-structured interview to triangulate the quantitative data. Some of our results of RQ1 confirmed what was previously found in the original work (e.g. in contrast to Basic mode, participants using the Non-sync mode have better collaboration experiences and more discussion time). However, we did find some differences in results that were introduced by the background difference. For example, the view sharing tools were beneficial in the prior work but are not appreciated in the educational settings (see 5.2.2). Accordingly, the original work showed great improvement in view similarity compared to the baseline (similar to our Basic mode), whereas our replicated work and Basic mode did not have a statistically different similarity score. In running this study, we found several challenges in doing replication work in HCI. First, the original system is not open-source and was not available upon request. We were forced to re-implement the original system. Second, some of the system implementation details were not compatible with the current implementation environment. We recommend that the HCI community consider how we could better scaffold and infrastructure system replication work.

6.3 Implications for Design and Research on Collaborative VR Video Viewing

6.3.1 Balancing the trade-off between learning pace flexibility and communication comfort based on teaching needs. Although we didn't see a significant difference in knowledge acquisition between the two collaborative VR video viewing systems, our results in 5.1 and 5.2.1 show that shared control brings a better communication and collaboration experiences and higher social presence, while individual control provides better learning pace flexibility. The expectations for time flexibility and collaboration experience might differ for diverse educational activities and learning scenarios. Therefore, VR collaborative applications should decide whether or not to use shared control based on specific educational purposes. More specifically, we suggest using shared video control in the application if an educational activity requires more discussion and teamwork or if students have a more homogeneous knowledge background and learning ability. We conjecture that in some situations, students would have a more similar learning pace and therefore would not require going through the material at different speeds. Those situations might be ones where students have the same level of knowledge to begin with. The non-shared control system should be preferred when students have a more diverged background of learning ability. From our interview (see 5.2.2), we found that Non-sync mode may have the potential to support Distributed Tutored Video Instruction (DTVI) [24], in which a small group would work with a tutor or student facilitator. For example, with Non-sync mode, the facilitator could check other students' working progress, find the appropriate points on the video to ask others to follow their view, and then initiate discussion topics.[34]

6.3.2 Providing note-taking methods appropriate to the modality of interaction and collaboration. From theme 2 (see 5.2.2), participants noted the importance of note-taking tools for remembering knowledge. Participants' willingness and preference for different note-taking techniques depend on the interaction modality and participants' video collaboration patterns. Taking text-based notes as an example, participants can use either the speech-to-text tool or drawing tool with collaborative VR video delivery. However, many participants noted that neither of those tools was as comfortable or flexible as using pen and paper, especially when writing down a long sentence. Some techniques and approaches [28, 113, 117, 125] have the potential for future applications to improve note-taking experiences in VR. Future research should further investigate whether such interaction modalities can help improve the note-taking experience in a VR video viewing process.

The findings in theme 3 (see 5.2.3) show the necessity of a smooth note transmission from the video to the after-video discussion platform to support collaboration. Participants indicated the difficulty of the current in-VR platform in navigating and managing the notes. We emphasize that note-taking for the purpose of having a collaborative discussion is perhaps different than just note taking for somebody who's doing it for their own purposes. In this case, it was important to be able to share these notes and easily navigate between multiple people's notes. When taking notes specifically for collaborative discussion, these features are necessary for ease of sharing. Future applications should provide a way to transform, navigate and manage the group's notes in collaborative VR for further learning or collaborative activities. Prior works [29, 67, 76, 124, 128] have studied how to better navigate and organize the notes with conventional online learning. A promising direction might be looking into whether those strategies and designs could work well with collaborative VR video-based learning settings.

6.3.3 Supporting awareness visualization but avoid distraction. Awareness visualization enables social interaction by visualizing others' statuses. Our results (See 5.2.2) show it helped participants create a shared context that promotes discussion. However, we found participants might feel distracted by dynamic awareness tools such as viewport. This function serves as a good indicator of where others are focusing but also forces users to interact with more visual input. Because people are more sensitive to dynamic visuals [106] and thus it increases the split-attention effect [77] (in which students split their attention between multiple visual sources of information). Based on studies of the split-attention effect, students learned better when the instructional material did not require them to divide their attention [12]. Therefore, applications should be aware of the split-attention effect and reduce visual distraction in collaborative VR video viewing systems. One potential solution might be simply providing options to hide or fade the dynamic visual elements. We also suggest that applications seek more light-weighted methods to visualize gaze without adding too much visual overload or distraction [96].

Our results show that the embodied visualization didn't distract people (even got ignored) while watching the video. Compared to the viewport, it is more static because participants all used stationary boundaries in this study, and we only visualized their head and gesture movements in the current avatar. However, it might not

be a good example of avoiding distraction because tension exists between promoting social presence and avoiding visual distraction brought by awareness visualization. One limitation of our system is that we chose a particular representation of the avatars, which was very abstract. It is possible that people did not attend to the embodied visualizations because they were not designed to be attractive, realistic, or particularly informative about the people behind them. They also provided no possibility for embodied awareness such as eye and mouth state, facial expression and full-body movements. In other words, these avatars might not have been useful enough for people to attend to. Prior studies noted the realism of an avatar could influence social interaction quality [88, 101]. Thus, future researchers could explore participants' visual preferences and priority on collaborative VR video viewing systems (e.g., using eye-tracking technique [16, 98]) to better guide this area.

6.4 Limitations and Future Work

One main limitation of this study is the Hawthorne effect[107] of the lab-based study. Although we used mixed methods, including qualitative and quantitative approaches, to do the data triangulation in order to reduce the Hawthorne effect, we acknowledge that lab-based studies have lower ecological validity because people might behave differently when they were influenced by observation and measurement. We chose to do a lab-based study because we wanted to control the comparable conditions and to be able to examine the effects of the variables pointed out in our research question. Future work might examine the systems compared to the status quo in a more ecologically valid setting.

We elected for a within-subjects design in which each participant was exposed to all conditions. This approach might have the demand characteristic [10, 43], which means participants might behave in ways that they think are desirable. For example, participants were informed that the purpose of the study was to investigate collaborative VR video tools. This knowledge may have made informed users more likely to report negative aspects of the basic mode because of its lack of collaborative tools. It is hard to completely eliminate the demand characteristic in the HCI area because the stimuli are easily distinguishable by both the participant and the researcher [22]. One potential suggestion to help minimize its impact on future studies is to take precautions such as using a double-blind study.

Since different lengths of time spent on tasks may lead to different ways of using the systems, another potential limitation of this study is that short tasks were used. For example, one participant said *"I didn't use view sharing features (in Non-sync mode) because the video is short. I can get their current contexts by simply dragging the process bar. That may change if the video is longer."* Despite this potential complication, we chose a short task in this study because there were practical considerations of VR collaborative studies that made length limitations necessary. This study requires three participants to join together, which increases the logistical schedule difficulty. In addition, this study was relatively long and tiresome due to the system tutorial and repeated self-reports and knowledge tests, raising the possibility of getting cybersickness and fatigue caused by VR. A longer learning task may exacerbate these effects and reduce the data validity. We encourage other researchers to replicate aspects of this work with longer tasks or use different time allocations of during- and after-video sessions.

As our goal was to seek a better practice and provide clear recommendations on which collaborative VR video viewing technology to use, we combined the during- and after-video session as one learning unit since this structure is more like an actual classroom setting and aligns with general video-based learning collaborative practice [11, 24]. However, due to limited test participants and the combination of during- and after-video settings, some results may not be significant. This may "dilute" the effects or feelings caused by collaborative video design compared to measuring variables directly after watching the video. In order to provide more insights specifically about how they used collaborative video tools, we used log files and interviews to examine their during-video experiences.

We chose three 360 city tour videos as our learning material with a visitor guide creation task in this study. This reduced content validity because this instrument didn't adequately cover the entire domain related to the educational VR videos and possible learning activities. Variations in learning content, learning task, narration, and editing method, might cause different discussion and collaboration patterns. However, we still believe that our results can provide guidance for instructors to choose what types of technology would be most helpful to meet their teaching goals. It could also provide guidance to developers so that they can revise the current technology from a more pedagogical perspective. In order to better understand the impact of content choice, future research should replicate this study design with other learning materials and learning tasks to get more valid results.

Because our study was conducted in a single session in the lab, novelty effects may have reduced ecological validity. For those participants who hadn't used VR before, watching video and communicating with people in VR was quite a novel experience. Users explicitly reacted to the novelty of the technology. For example, one participant noted that he prefers the Basic mode because *"I am already very familiar with Zoom and Google Slides, so I don't worry about remembering how to use the tool."* As collaborative VR is seen as a novel technology for higher education, this novelty may influence the perceived differences between conventional and full VR settings. Sustained use of collaborative VR across multiple sessions may lead to different perceptions and preferences, and future studies should investigate such sustained engagement.

7 CONCLUSION

Collaborative learning and immersive environments have both been proven to have benefits for learning. As a less demanding way to create immersive environments, we chose VR video as the learning base and implemented two VR video viewing systems for distributed collaborative learning. From a within-subject study, we examined the role of collaborative tools and shared video control by comparing the existing technology (Basic mode), shared video control system (Sync mode) and individual video control system (Non-sync mode) on measures of knowledge acquisition, collaboration, social presence, cognitive load and satisfaction. Our results showed collaborative VR systems and shared video control did increase collaboration and social presence, yet the effects on learning were mixed. We saw the satisfaction of shared video control condition (Sync mode) was statistical significance better than

the conventional settings while was equally well with Non-sync mode. There was no statistically significant differences found in cognitive load. Based on our findings, we came up with three major implications for designing a better collaborative video viewing system for educational purposes. However, this research still requires confirmation from field studies and other subjects or educational scenarios. Future research should address the question of whether tutors are necessary, and the related design and practice problem of how instructors manage classes with multiple groups.

ACKNOWLEDGMENTS

We would like to thank our participants for generously contributing their time and valuable insights. We also appreciate the constructive feedback and suggestions provided by our colleagues at Grouplens Lab and anonymous reviewers, which helped improve this work. This research is supported by NSF grants 2212298, 2106090, and 1915122.

REFERENCES

[1] Rahman Abidin, Nunuk Suryani, et al. 2020. Students' Perceptions of 360 Degree Virtual Tour-Based Historical Learning About the Cultural Heritage Area of the Kapitan and Al-Munawar Villages in Palembang City. *International Journal of Social Sciences and Management* 7, 3 (2020), 105–112.

[2] Salsabeel FM Alfalah, Jannat FM Falah, Tasneem Alfalah, Mutasem Elfalah, Nadia Muhaidat, and Orwa Falah. 2019. A comparative study between a virtual reality heart anatomy system and traditional medical teaching modalities. *Virtual Reality* 23, 3 (2019), 229–234.

[3] Lorin W Anderson and David R Krathwohl. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

[4] Dan Archer and Katharina Finger. 2018. Walking in another's virtual shoes: Do 360-degree video news stories generate empathy in viewers? (2018).

[5] Penelope Atsikpasi and Emmanuel Fokides. 2021. A scoping review of the educational uses of 6DoF HMDs. *Virtual Reality* (2021), 1–18.

[6] Jeremy N Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14, 4 (2005), 379–393.

[7] K Louise Barriball and Alison While. 1993. Collecting data using a semi-structured interview: a discussion paper. *Journal of advanced nursing* 18, 10 (1993), 328–335.

[8] Steve Benford, Chris Greenhalgh, Tom Rodden, and James Pycock. 2001. Collaborative virtual environments. *Commun. ACM* 44, 7 (2001), 79–85.

[9] Abbie Brown and Tim Green. 2016. Virtual reality: Low-cost tools and resources for the classroom. *TechTrends* 60, 5 (2016), 517–519.

[10] Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1657–1666.

[11] Jonathan J Cadiz, Anand Balachandran, Elizabeth Sanocki, Anoop Gupta, Jonathan Grudin, and Gavin Jancke. 2000. Distance learning through distributed collaborative video viewing. In *Proceedings of the 2000 ACM conference on computer supported cooperative work*. 135–144.

[12] Paul Chandler and John Sweller. 1992. The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology* 62, 2 (1992), 233–246.

[13] Chen-Wei Chang, Shih-Ching Yeh, Mengtong Li, and Eason Yao. 2019. The introduction of a novel virtual reality training system for gynecology learning and its user experience research. *IEEE Access* 7 (2019), 43637–43653.

[14] David Checa and Andres Bustillo. 2020. Advantages and limits of virtual reality in learning processes: Briviesca in the fifteenth century. *Virtual Reality* 24, 1 (2020), 151–161.

[15] Gabriele Cierniak, Katharina Scheiter, and Peter Gerjets. 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior* 25, 2 (2009), 315–324.

[16] Viviane Clay, Peter König, and Sabine Koenig. 2019. Eye tracking in virtual reality. *Journal of eye movement research* 12, 1 (2019).

[17] Susan Copley Cobb. 2009. Social presence and online learning: A current view from a research perspective. *Journal of Interactive Online Learning* 8, 3 (2009).

[18] J Cohen. 1992. A power primer Psychological Bulletin 112: 155–159. *IE and Dyadic Adjustment* 1 (1992).

[19] Zoom Video Communications. 2011. Zoom. Retrieved Feb 15, 2023 from https://zoom.us/

[20] Maxime Cordeil, Tim Dwyer, Karsten Klein, Bireswar Laha, Kim Marriott, and Bruce H Thomas. 2016. Immersive collaborative analysis of network connectivity: CAVE-style or head-mounted display? *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 441–450.

[21] Camila Cortez, Miguel Nussbaum, Gerardo Woywood, and Ricardo Aravena. 2009. Learning to collaborate by collaborating: a face-to-face collaborative activity for measuring and learning basics about teamwork 1. *Journal of Computer Assisted Learning* 25, 2 (2009), 126–142.

[22] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. " Yours is better!" participant response bias in HCI. In *Proceedings of the sigchi conference on human factors in computing systems*. 1321–1330.

[23] Tobias Drey, Patrick Albus, Simon der Kinderen, Maximilian Milo, Thilo Segschneider, Linda Chanzab, Michael Rietzler, Tina Seufert, and Enrico Rukzio. 2022. Towards Collaborative Learning in Virtual Reality: A Comparison of Co-Located Symmetric and Asymmetric Pair-Learning. In *CHI Conference on Human Factors in Computing Systems*. 1–19.

[24] John Dutra, James F Gibbons, Robert L Pannoni, Michael J Sipusic, Randall B Smith, and William R Sutherland. 1999. Virtual Collaborative Learning: A Comparison between Face-to-Face Tutored Video Instruction (TVI) and Distributed Tutored Video Instruction (DTVI).

[25] Noureddine Elmqaddem. 2019. Augmented reality and virtual reality in education. Myth or reality? *International journal of emerging technologies in learning* 14, 3 (2019).

[26] Marie Evens, Michaël Empsen, and Wouter Hustinx. 2022. A literature review on 360-degree video as an educational tool: towards design guidelines. *Journal of Computers in Education* (2022), 1–51.

[27] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, Ching Liu, and Hao-Chuan Wang. 2022. Understanding the Effects of Structured Note-taking Systems for Video-based Learners in Individual and Social Learning Contexts. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–21.

[28] Zhangjie Fu, Jiashuang Xu, Zhuangdi Zhu, Alex X Liu, and Xingming Sun. 2018. Writing in the air with WiFi signals for virtual reality devices. *IEEE Transactions on Mobile Computing* 18, 2 (2018), 473–484.

[29] Marco Furini. 2018. On introducing timed tag-clouds in video lectures indexing. *Multimedia Tools and Applications* 77, 1 (2018), 967–984.

[30] Susan R Fussell, Robert E Kraut, and Jane Siegel. 2000. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. 21–30.

[31] Fani Galatsopoulou, Clio Kenterelidou, Rigas Kotsakis, and Maria Matsiola. 2022. Examining Students' Perceptions towards Video-Based and Video-Assisted Active Learning Scenarios in Journalism and Communication Courses. *Education Sciences* 12, 2 (2022), 74.

[32] Ronak Dipakkumar Gandhi and Dipam S Patel. 2018. Virtual reality–opportunities and challenges. *Virtual Reality* 5, 01 (2018).

[33] Michail Giannakos, Konstantinos Chorianopoulos, Marco Ronchetti, Peter Szegedi, and Stephanie Teasley. 2014. Video-based learning and open online courses. (2014).

[34] James F Gibbons, WR Kincheloe, and KS Down. 1977. Tutored videotape instruction: a new use of electronics media in education. *Science* 195, 4283 (1977), 1139–1146.

[35] Barney G Glaser. 1992. *Basics of grounded theory analysis: Emergence vs forcing*. Sociology press.

[36] Bernadette Gold and Julian Windscheid. 2020. Observing 360-degree classroom videos–Effects of video type on presence, emotions, workload, classroom observations, and ratings of teaching quality. *Computers & Education* 156 (2020), 103960.

[37] R Goldman. 2006. Orion, an Online Collaborative Digital Video Data Analysis Tool: Changing Our Perspectives as an Interpretive Community. *Video Research in the Learning Sciences, R. Goldman, R. Pea, B. Barron, and SJ Derry, eds* (2006), 507–520.

[38] Google. 2006. Google Slides. Retrieved Feb 15, 2023 from https://slides.google.com

[39] Google. 2015. Google AI. Retrieved Feb 15, 2023 from https://ai.google/

[40] Saul Greenberg and Bill Buxton. 2008. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 111–120.

[41] Scott W Greenwald, Zhangyuan Wang, Markus Funk, and Pattie Maes. 2017. Investigating social presence and communication with embodied avatars in room-scale virtual reality. In *International Conference on Immersive Learning*. Springer, 75–90.

[42] William A Hamilton, Nic Lupfer, Nicolas Botello, Tyler Tesch, Alex Stacy, Jeremy Merrill, Blake Williford, Frank R Bentley, and Andruid Kerne. 2018. Collaborative live media curation: Shared context for participation in online learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[43] Michael A Harvey and Carl N Sipprelle. 1976. Demand characteristic effects on the subtle and obvious subscales of the MMPI. *Journal of Personality Assessment* 40, 5 (1976), 539–544.

[44] Sean Hauze, Helina Hoyt, James Marshall, James Frazee, and Philip Greiner. 2018. An evaluation of nursing student motivation to learn through holographic mixed reality simulation. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. IEEE, 1058–1063.

[45] Clyde Hendrick. 1990. Replications, strict replications, and conceptual replications: are they important? *Journal of Social Behavior and Personality* 5, 4 (1990), 41.

[46] Paula Hodgson, Vivian WY Lee, Johnson Chan, Agnes Fong, Cindi SY Tang, Leo Chan, and Cathy Wong. 2019. Immersive virtual reality (IVR) in higher education: Development and implementation. In *Augmented reality and virtual reality*. Springer, 161–173.

[47] Kasper Hornbæk, Søren S Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is once enough? On the extent and content of replications in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3523–3532.

[48] Carol Hostetter. 2013. Community matters: Social presence and learning outcomes. *Journal of the Scholarship of Teaching and Learning* 13, 1 (2013), 77–86.

[49] Amanda Hurlbut. 2021. Videos or Zoom? Using a Flipped Classroom Approach to Facilitate PST Online Learning. In *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 175–184.

[50] Wu-Yuin Hwang and Shih-Shin Hu. 2013. Analysis of peer learning behaviors using multiple representations in virtual reality and their impacts on geometry problem solving. *Computers & Education* 62 (2013), 308–319.

[51] Santiago González Izard, Juan A Juanes, Francisco J García Peñalvo, Jesús Mª Estella, Mª Ledesma, and Pablo Ruisoto. 2018. Virtual reality as an educational and training tool for medicine. *Journal of medical systems* 42, 3 (2018), 1–5.

[52] Lasse Jensen and Flemming Konradsen. 2018. A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies* 23, 4 (2018), 1515–1529.

[53] Qiao Jin, Yu Liu, Svetlana Yarosh, Bo Han, and Feng Qian. 2022. How Will VR Enter University Classrooms? Multi-stakeholders Investigation of VR in Higher Education. In *CHI Conference on Human Factors in Computing Systems*. 1–17.

[54] Qiao Jin, Yu Liu, Ye Yuan, Lana Yarosh, and Evan Suma Rosenberg. 2020. VWorld: an immersive VR system for learning programming. In *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*. 235–240.

[55] Pekka Kallioniemi, Tuuli Keskinen, Jaakko Hakulinen, Markku Turunen, Jussi Karhu, and Kimmo Ronkainen. 2017. Effect of Gender on Immersion in Collaborative IODV Applications. Association for Computing Machinery, New York, NY, USA.

[56] Sam Kavanagh, Andrew Luxton-Reilly, Burkhard Wuensche, and Beryl Plimmer. 2017. A systematic review of virtual reality in education. *Themes in Science and Technology Education* 10, 2 (2017), 85–119.

[57] David S Kirk and Danaë Stanton Fraser. 2017. The effects of remote gesturing on distance instruction. In *Computer Supported Collaborative Learning 2005: The Next 10 Years!* Routledge, 301–310.

[58] Paul A Kirschner, John Sweller, Femke Kirschner, Jimmy Zambrano R, et al. 2018. From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning* 13, 2 (2018), 213–233.

[59] Karel Kreijns, Paul A Kirschner, Wim Jochems, and Hans Van Buuren. 2011. Measuring perceived social presence in distributed learning groups. *Education and Information Technologies* 16, 4 (2011), 365–381.

[60] Kartikaeya Kumar, Lev Poretski, Jiannan Li, and Anthony Tang. 2022. Tourgether360: Collaborative Exploration of 360° Videos Using Pseudo-Spatial Navigation. 6, CSCW2 (2022).

[61] Marjan Laal and Seyed Mohammad Ghodsi. 2012. Benefits of collaborative learning. *Procedia-social and behavioral sciences* 31 (2012), 486–490.

[62] Joseph LaViola. 2000. MSVT: A virtual reality-based multimodal scientific visualization tool. In *Proceedings of the third IASTED international conference on computer graphics and imaging*. 1–7.

[63] Gun A Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2018. A user study on mr remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 153–164.

[64] Nan Li, Himanshu Verma, Afroditi Skevi, Guillaume Zufferey, Jan Blom, and Pierre Dillenbourg. 2014. Watching MOOCs together: investigating co-located MOOC study groups. *Distance Education* 35, 2 (2014), 217–233.

[65] Nan Li, Himanshu Verma, Afroditi Skevi, Guillaume Zufferey, and Pierre Dillenbourg. 2014. *MOOC learning in spontaneous study groups: Does synchronously watching videos make a difference?* Technical Report. PAU Education.

[66] Chin-Wen Liao, Ching-Huei Chen, and Sie-Jhih Shih. 2019. The interactivity of video and collaboration for learning achievement, intrinsic motivation, cognitive load, and behavior patterns in a digital game-based learning environment. *Computers & Education* 133 (2019), 43–55.

[67] Ching Liu, Juho Kim, and Hao-Chuan Wang. 2018. ConceptScape: Collaborative concept mapping for video learning. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[68] David M Markowitz, Rob Laha, Brian P Perone, Roy D Pea, and Jeremy N Bailenson. 2018. Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in psychology* 9 (2018), 2364.

[69] Benjy Marks and Jacqueline Thomas. 2022. Adoption of virtual reality technology in higher education: An evaluation of five teaching semesters in a purpose-designed laboratory. *Education and information technologies* 27, 1 (2022), 1287–1305.

[70] Orbitian Media. 2017. One day in Granada: 360° Virtual Tour with Voice Over. Retrieved Feb 15, 2023 from https://www.youtube.com/watch?v=_13l5A7lc2g&t=2s

[71] Orbitian Media. 2017. One day in Porto: 360° Virtual Tour with Voice Over. Retrieved Feb 15, 2023 from https://www.youtube.com/watch?v=vxBasS20MDs&t=40s

[72] Orbitian Media. 2017. One day in Seville: 360° Virtual Tour with Voice Over. Retrieved Feb 15, 2023 from https://www.youtube.com/watch?v=BwO7PlBOFjk&t=10s

[73] Orbitian Media. 2023. One day in: 360° travel videos. https://www.youtube.com/playlist?list=PLHiCdB8YTO76Gv843e8rdcza-FJbYCYSj

[74] Shahid Minhas, Tasaddaq Hussain, Abdul Ghani, Kiran Sajid, and L Pakistan. 2021. Exploring students online learning: A study of zoom application. *Gazi University Journal of Science* 34, 2 (2021), 171–178.

[75] Rachel Mintz, Shai Litvak, and Yoav Yair. 2001. 3D-virtual reality in science education: An implication for astronomy teaching. *Journal of Computers in Mathematics and Science Teaching* 20, 3 (2001), 293–305.

[76] Toni-Jan Keith Palma Monserrat, Shengdong Zhao, Kevin McGee, and Anshul Vikram Pandey. 2013. NoteVideo: Facilitating Navigation of Blackboard-Style Lecture Videos. Association for Computing Machinery, New York, NY, USA.

[77] Roxana Moreno and Richard E Mayer. 1999. Visual presentations in multimedia learning: Conditions that overload visual working memory. In *International conference on advances in visual information systems*. Springer, 798–805.

[78] Michael Muller. 2014. Curiosity, creativity, and surprise as analytic tools: Grounded theory method. In *Ways of Knowing in HCI*. Springer, 25–48.

[79] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. CollaVR: collaborative in-headset review for VR video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 267–277.

[80] Heather L O'Brien and Elaine G Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology* 59, 6 (2008), 938–955.

[81] Elizaveta Osipovskaya and Elena Burdovskaya. 2019. Presentation Software Tools in Higher Educational Setting. *ARPHA Proceedings* 1 (2019), 1487.

[82] Elinor Ostrom. 2008. The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. *Perspectives on Politics* 6, 4 (2008), 828–829.

[83] Murat Oztok and Clare Brett. 2011. Social presence and online learning: A review of the research. (2011).

[84] Karoliina Paalimäki-Paakki, Mari Virtanen, Anja Henner, MT Nieminen, and M Kääriäinen. 2021. Patients', radiographers' and radiography students' experiences of 360° virtual counselling environment for the coronary computed tomography angiography: A qualitative study. *Radiography* 27, 2 (2021), 381–388.

[85] Debajyoti Pal and Syamal Patra. 2021. University students' perception of video-based learning in times of COVID-19: A TAM/TTF perspective. *International Journal of Human–Computer Interaction* 37, 10 (2021), 903–921.

[86] Dhaval Parmar, Joseph Isaac, Sabarish V Babu, Nikeetha D'Souza, Alison E Leonard, Sophie Jörg, Kara Gundersen, and Shaundra B Daily. 2016. Programming moves: Design and evaluation of applying embodied interaction in virtual environments to enhance computational thinking in middle school students. In *2016 IEEE Virtual Reality (VR)*. IEEE, 131–140.

[87] Roy Pea and Robb Lindgren. 2008. Video collaboratories for research and education: An analysis of collaboration design patterns. *IEEE Transactions on Learning Technologies* 1, 4 (2008), 235–247.

[88] Gustav Bøg Petersen, Aske Mottelson, and Guido Makransky. 2021. Pedagogical agents in educational vr: An in the wild study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[89] Photon. 2013. Photon Engine: Multiplayer Game Development. Retrieved Feb 15, 2023 from https://www.photonengine.com

[90] Catlin Pidel and Philipp Ackermann. 2020. Collaboration in virtual and augmented reality: a systematic overview. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 141–156.

[91] Daniel Pimentel, Sri Kalyanaraman, Yu-Hao Lee, and Shiva Halan. 2021. Voices of the unsung: The role of social presence and interactivity in building empathy in 360 video. *new media & society* 23, 8 (2021), 2230–2254.

[92] Johanna Pirker and Andreas Dengel. 2021. The Potential of 360-Degree Virtual Reality Videos and Real VR for Education-A Literature. (2021).

[93] Yeshwanth Pulijala, Minhua Ma, Matt Pears, David Peebles, and Ashraf Ayoub. 2018. An innovative virtual reality training tool for orthognathic surgery. *International journal of oral and maxillofacial surgery* 47, 9 (2018), 1199–1205.

[94] M Radia, M Arunakirinathan, and D Sibley. 2018. A guide to eyes: ophthalmic simulators. *The Bulletin of the Royal College of Surgeons of England* 100, 4 (2018), 169–171.

[95] Rivu Radiah, Ville Mäkelä, Sarah Prange, Sarah Delgado Rodriguez, Robin Piening, Yumeng Zhou, Kay Köhle, Ken Pfeuffer, Yomna Abdelrahman, Matthias Hoppe, Albrecht Schmidt, and Florian Alt. 2021. Remote VR Studies: A Framework for Running Virtual Reality Studies Remotely Via Participant-Owned HMDs. 28, 6 (2021).

[96] Yitoshee Rahman, Sarker M Asish, Nicholas P Fisher, Ethan C Bruce, Arun K Kulshreshth, and Christoph W Borst. 2020. Exploring eye gaze visualization techniques for identifying distracted students in educational VR. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 868–877.

[97] Maria Ranieri, Damiana Luzzi, Stefano Cuomo, and Isabella Bruni. 2022. If and how do 360° videos fit into education settings? Results from a scoping review of empirical research. *Journal of Computer Assisted Learning* (2022).

[98] Natasha Anne Rappa, Susan Ledger, Timothy Teo, Kok Wai Wong, Brad Power, and Bruce Hilliard. 2019. The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review. *Interactive Learning Environments* (2019), 1–13.

[99] Paul Resta and Thérèse Laferrière. 2007. Technology in support of collaborative learning. *Educational psychology review* 19, 1 (2007), 65–83.

[100] Vicente Román-Ibáñez, Francisco A Pujol-López, Higinio Mora-Mora, Maria Luisa Pertegal-Felices, and Antonio Jimeno-Morenilla. 2018. A low-cost immersive virtual reality system for teaching robotic manipulators programming. *Sustainability* 10, 4 (2018), 1102.

[101] Daniel Roth, Jean-Luc Lugrin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. 2016. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*. IEEE, 277–278.

[102] Michael A Rupp, James Kozachuk, Jessica R Michaelis, Katy L Odette, Janan A Smither, and Daniel S McConnell. 2016. The effects of immersiveness and future VR expectations on subjec-tive-experiences during an educational 360 video. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 2108–2112.

[103] Marija Sablić, Ana Mirosavljević, and Alma Škugor. 2021. Video-based learning (VBL)—past, present and future: An overview of the research published from 2008 to 2019. *Technology, Knowledge and Learning* 26, 4 (2021), 1061–1077.

[104] David Saffo, Sara Di Bartolomeo, Caglar Yildirim, and Cody Dunne. 2021. Remote and Collaborative Virtual Reality Experiments via Social VR Platforms. Association for Computing Machinery, New York, NY, USA.

[105] Alcínia Zita Sampaio and Octávio Martins. 2017. VR model of bridge construction: a didactic application. In *Proceedings of the Virtual Reality International Conference-Laval Virtual 2017*. 1–3.

[106] Charles T Scialfa, Philip M Garvey, Kenneth W Gish, Linda M Deering, Herschel W Leibowitz, and Charles C Goebel. 1988. Relationships among measures of static and dynamic visual sensitivity. *Human Factors* 30, 6 (1988), 677–687.

[107] Philip Sedgwick and Nan Greenwood. 2015. Understanding the Hawthorne effect. *Bmj* 351 (2015).

[108] Jinsil Hwaryoung Seo, Brian Michael Smith, Margaret Cook, Erica Malone, Michelle Pine, Steven Leal, Zhikun Bai, and Jinkyo Suh. 2017. Anatomy builder VR: Applying a constructive learning method in the virtual reality canine skeletal system. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 245–252.

[109] John J Shaughnessy, Eugene B Zechmeister, and Jeanne S Zechmeister. 2000. *Research methods in psychology*. McGraw-Hill.

[110] Samarth Singhal and Carman Neustaedter. 2017. Bewithme: An immersive telepresence system for distance separated couples. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 307–310.

[111] Chareen Snelson and Yu-Chang Hsu. 2020. Educational 360-degree videos in virtual reality: A scoping review of the emerging research. *TechTrends* 64, 3 (2020), 404–412.

[112] Hyo-Jeong So and Thomas A Brush. 2008. Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & education* 51, 1 (2008), 318–336.

[113] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[114] Adam Stefanile. 2020. The transition from classroom to Zoom and how it has changed education. *Journal of social science research* 16 (2020), 33–40.

[115] John Sweller, Jeroen JG Van Merrienboer, and Fred GWC Paas. 1998. Cognitive architecture and instructional design. *Educational psychology review* 10, 3 (1998), 251–296.

[116] Anthony Tang and Omid Fakourfar. 2017. Watching 360° Videos Together. Association for Computing Machinery, New York, NY, USA.

[117] Yang Tian, Chi-Wing Fu, Shengdong Zhao, Ruihui Li, Xiao Tang, Xiaowei Hu, and Pheng-Ann Heng. 2019. Enhancing Augmented VR Interaction via Egocentric Scene Analysis. 3, 3 (2019).

[118] Bruce Torff and Rose Tirotta. 2010. Interactive whiteboards produce small gains in elementary students' self-reported motivation in mathematics. *Computers & Education* 54, 2 (2010), 379–383.

[119] Nicola Walshe and Paul Driver. 2019. Developing reflective trainee teacher practice with 360-degree video. *Teaching and Teacher Education* 78 (2019), 97–105.

[120] Cheng Yao Wang, Mose Sakashita, Upol Ehsan, Jingjin Li, and Andrea Stevenson Won. 2020. Again, together: Socially reliving virtual reality experiences when separated. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[121] Fu Wang, Ying Liu, Min Tian, Yumei Zhang, Shaofeng Zhang, and Jihua Chen. 2016. Application of a 3D haptic virtual reality simulation system for dental crown preparation training. In *2016 8th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, 424–427.

[122] Max Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, and Jeffrey Nichols. 2012. RepliCHI SIG: From a panel to a new submission venue for replication. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 1185–1188.

[123] Max LL Wilson, Paul Resnick, David Coyle, and Ed H Chi. 2013. Replichi: the workshop. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 3159–3162.

[124] Chengpei Xu, Ruomei Wang, Shujin Lin, Xiaonan Luo, Baoquan Zhao, Lijie Shao, and Mengqiu Hu. 2019. Lecture2Note: Automatic Generation of Lecture Notes from Slide-Based Educational Videos. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 898–903.

[125] Kiwon Yeom, Jounghuem Kwon, JooHyun Maeng, and Bum-Jae You. 2015. [POSTER] Haptic Ring Interface Enabling Air-Writing in Virtual Reality Environment. In *2015 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 124–127.

[126] Ahmed Mohamed Fahmy Yousef, Mohamed Amine Chatti, and Ulrik Schroeder. 2014. The state of video-based learning: A review and future perspectives. *International Journal on Advances in Life Sciences* 6, 3 (2014), 122–135.

[127] Carmen Zahn, Karsten Krauskopf, Friedrich W Hesse, and Roy Pea. 2012. How to improve collaborative learning with video tools in the classroom? Social vs. cognitive guidance for student teams. *International Journal of Computer-Supported Collaborative Learning* 7, 2 (2012), 259–284.

[128] Shan Zhang, Xiaojun Meng, Can Liu, Shengdong Zhao, Vibhor Sehgal, and Morten Fjeld. 2019. ScaffoMapping: Assisting concept mapping for video learners. In *IFIP Conference on Human-Computer Interaction*. Springer, 314–328.

[129] Huahui Zhao, Kirk PH Sullivan, and Ingmarie Mellenius. 2014. Participation, interaction and social presence: An exploratory study of collaboration in online peer review groups. *British Journal of Educational Technology* 45, 5 (2014), 807–819.

[130] Jiayan Zhao, Peter LaFemina, Julia Carr, Pejman Sajjadi, Jan Oliver Wallgrün, and Alexander Klippel. 2020. Learning in the field: Comparison of desktop, immersive virtual reality, and actual field trips for place-based STEM education. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 893–902.