International Society of the Learning Sciences

Interaction Forms of Collaborative VR Video Learning: An Exploratory Study

Qiao Jin, Ishan Sodhi, Angie Chen, Svetlana Yarosh
jin00122@umn.edu, sodhi012@umn.edu, chen7313@umn.edu, lana@umn.edu
Department of Computer Science and Engineering, University of Minnesota

Abstract: Collaborative and immersive environments both contribute to increasing engagement in video-based learning (VBL). This exploratory study delves into the dynamics of interaction forms within collaborative VR video-based learning and evaluates how shared and individual video control influence these interactions. Through a within-subject study involving 18 groups of three participants (N = 54), we aim to address three research questions: 1) identifying interaction forms in collaborative VR video-based learning, 2) comparing interaction forms between shared and individual video control, and 3) examining the relationship between interaction forms and knowledge acquisition and satisfaction. Our findings provide valuable insights for technology developers, researchers, and educators looking to enhance VBL systems and instructional approaches, as well as gain a deeper understanding of how social interactions impact knowledge acquisition and satisfaction in VR video-based learning environments.

Introduction

Video-based learning (VBL) has been shown to be an effective educational approach (Kay, 2012), leveraging visual and auditory elements for enhanced learning process. However, VBL for online classes faces challenges such as isolation and reduced motivation (Fang et al., 2022). In response to these challenges, collaborative video viewing, such as the Distributed Collaborative Video Viewing (DCVV) model, has emerged as a pedagogical strategy with broad-reaching benefits (Cadiz et al., 2000). These approaches have been demonstrated to enhance students' engagement and improve their overall learning experiences. Further, the infusion of Virtual Reality (VR) technology into VBL (e.g., VR videos[(1)]) has been shown to amplify students' intrinsic motivation, bolster engagement in learning activities (Fang et al., 2022). Additionally, VR environments naturally offers a shared digital space that allows students to immerse themselves in learning contexts, aiding students in constructing and retaining visual knowledge (Kallioniemi et al., 2017).

Previous research has connected the effectiveness of collaborative learning to the types of interactions taking place among learners (Vuopala et al., 2016). Additionally, research has shown that technology design can influence the types of interactions that occur in virtual learning environments. Inspired by this, our study seeks to examine the various forms of interaction within different collaborative modes and their impact in the context of VR video-based learning. In pursuit these objectives, we formulate the following research questions:

- *RQ1:* What forms of interaction occur during collaborative VR video-based learning?
- *RQ2:* How do forms of interaction vary between the shared control (Sync mode) and individual control (Non-sync mode) in collaborative VR video viewing systems?
- *RQ3*: How do forms of interaction relate to the measures of knowledge acquisition and satisfaction?

We conducted a within-subject experiment with 18 groups and utilized video analysis to answer our research questions. The study reveals how collaborative interactions unfold and impact on knowledge acquisition and satisfaction during VR video-based field trips, shedding light on the influence of shared and individual video control techniques on interaction patterns. Our findings advance the understanding of collaborative learning within VR videos, offering preliminary insights that can inform both technology design and practical instructional strategies in the field of computer-supported collaborative learning (CSCL).

Related work

Video-based learning (VBL) can be an effective pedagogical approach (Kay, 2012) as it promotes greater sensory engagement and interactivity. However, some limitations of VBL have been proposed in prior work: students risk distraction, loss of motivation, or isolation without communication with an instructor or peers, potentially hindering learning engagement (Fang et al., 2022). Watching videos together positively increases the attention span and engagement of students (Li et al., 2014), and fosters conversation, allowing students to share perspectives and generate conceptual diversity (Goldman, 2014). These conversations are vital for knowledge building, and some literature suggests that they may be as or more important than the videos themselves (Pea & Lindgren, 2008). One collaborative VBL model is the Distributed Collaborative Video Viewing (DCVV) model proposed by Cadiz et al. (2000) wherein small groups of students watch educational videos without an instructor, periodically pausing to discuss, which was shown to have positive effects on student engagement and learning.

Alternatively, problems found with VBL such as poor intrinsic motivation and engagement (Torff & Tirotta, 2010), may be solved by learning in an immersive environment. One possible immersive environment is VR video which is more cost-effective and accessible (Jin et al. 2022) compared to other computer-generated graphical immersive environments. VR videos have been shown to increase students' intrinsic motivation, engagement in learning activities (Fang et al., 2022), allows students to access otherwise inaccessible learning contexts, and helps students build and retain visual knowledge (Kallioniemi et al. 2017). VR video can be applied successfully in many learning scenarios (Pirker & Dengel, 2021) include virtual tours, recorded processes and procedures, recorded situations, recorded experiences, etc. Although some literature exists on collaboration within VR video-based learning (Jackson & Fagan, 2000; Jin, et al. 2023), there has been little investigation into the forms of interaction between students in collaborative VR video-based learning. It has been shown that the forms of interaction which take place during a collaborative learning session can have a significant effect on its success (Vuopala et al., 2016). Certain forms of interaction significantly contribute to the quality of collaborative learning (e.g., interaction was more group-related than task-related) (Lebie et al., 1996).

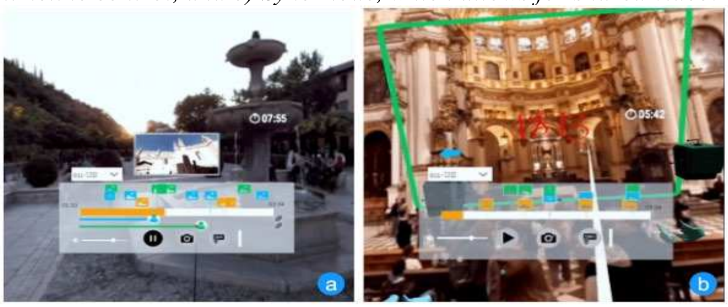
Therefore, we specify our research to VR video-based learning environment, seeking to investigate the influence of technology design on various forms of interaction and their impact on learning outcomes. More specifically, our study aims to explore interaction forms, their effects on knowledge acquisition and satisfaction, and how technology influences these interactions. This is important for educators, engineers, and researchers in designing learning environments that encourage these positive interactions.

Collaborative VR video viewing systems

We implement two collaborative VR video viewing systems (Figure 1) to explore how different video controls influence the forms of interaction in collaborative VR video-based learning. The Non-sync mode in collaborative VR video watching offers individual timeline control with collaborative tools enhancing in-VR communication. It includes *awareness* tools like spatialized voice chat, activity visualization displaying user progress and notes, and viewport visualization showing gaze directions (only under following function). *View sharing* tools comprise peek and full window options, and a follow function for synchronized viewing. *Note-taking* tools feature drawing, speech-to-text notes, and screenshot capabilities, with visibility adjustments for different timelines.

In contrast, the Sync mode, based on the DCVV model, allows shared control over the video playback. This mode enhances *awareness* through embodied visualization with 3D avatars and spatialized voice chat using avatar positions. It simplifies activity visualization by omitting individual progress displays but maintains note indicators. The visualized gaze is constant due to shared timeline progress, and the ray cursor of teammates is color-coded for better interaction. Unlike the Non-sync mode, view sharing functions are not required due to shared progress, and *note-taking* tools such as screenshots and speech-to-text notes are consistent, with drawings visible to all users at all times. The design considerations and technical details of systems are described in our previous work (Jin et al., 2023).

Figure 1
Collaborative VR video viewing systems: a) Non-sync mode, which enables individual timeline control, and b) Sync mode, which allows for shared video.



Methods

Participants and learning settings

This study is a part of a larger research project investigating collaborative VR video tools. 54 participants in 18 groups of three from a Midwest U.S. university involved in this study. Teams were formed based on availability and participants' choices for teammates. Participants were aged 19 to 31 (M = 22.5, SD = 2.65). There were 26 females, 27 males, and one undisclosed gender. Of the participants, 26 were unfamiliar with their teammates, 18 knew one teammate, and ten knew two teammates. 14 had significant VR experience, 26 had limited experience, and 14 were VR newcomers. 23 had watched 360 videos using VR headsets, 12 watched them on non-VR devices, and 19 had no 360-video exposure. All had normal or corrected-to-normal vision and received a $50 gift card for their participation.

The study took place in four adjacent university rooms—three for participants and one for an in-VR tech facilitator. Each room included a swivel chair, desk, laptop, VR headset, and a tripod-mounted camera capturing the process. In this study, we chose a virtual field trip as the learning context to leverage VR's capacity for accessing remote locations. A group would be watching VR videos together and discussing the city's attractions, history, and architecture. The videos, from the "One day in" 360° travel collection[2], depicted various global destinations. We used two narrated city-tour videos, edited for equal length and difficulty, and narrated by an AI-generated voice. Each video, highlighting historical and architectural aspects, lasted about 3.5 minutes and contained 14 clips, with resolutions of 1920 x 1024 pixels.

Procedure

The study is primarily divided into four parts: ice-breaking and introduction, training, and learning units (each featuring one mode). Within each learning unit, there are components for pre-knowledge assessment, a video session, and post-knowledge acquisition assessment, along with a satisfaction questionnaire. After a brief ice-breaking activity (~5 minutes), where three participants introduced themselves and shared VR-related impressions or experiences, an introduction (~5 minutes) by researchers outlined the study's objectives, procedures, learning goal, with an overview of system features. Following this, participants wore headsets for a system tutorial. Training duration ranged from 20 to 30 minutes based on participants' prior VR experience. After training, participants took a 5-minute break to ensure they were free from cybersickness. Once all participants were ready to enter the learning unit, the knowledge assessment as a pre-learning session began. Then, participants entered the system to watch the video. In this session, participants watched a 3.5-minute video and could pause, replay, and relocate it as needed within 10 minutes. Post-study, knowledge acquisition was assessed, followed by self-reported questionnaires of satisfaction. Two learning units (Sync and Non-sync mode) were conducted in a counterbalanced order, separated by a 5-minute break based on the motion sickness participants experienced.

Measures

The knowledge assessment was designed to assess students' knowledge acquisition as result of experiencing two technologies. The test consisted of 10 multi-option questions regarding the information presented in the videos according to Bloom's taxonomy (Anderson & Krathwohl, 2001) and prior educational VR video research (Radia et al., 20018), including four questions based on auditory information, three visual information, and three conceptual questions. The test was piloted two times and adjusted in order to avoid a ceiling effect.

Participants took an independent self-report questionnaire about their satisfaction after watching the video. We selected satisfaction as a metric since it is a significant indicator reflected "the degree of learner reaction to values and quality of learning, and motivation for learning" (Saffo et al., 2021). There were 11 questions in this validated questionnaire (So & Brush, 2008), aimed at assessing participant satisfaction with both the learning activities and the extent to which the system design aligned with their learning expectations. Because those measures were all collected individually, and we used an average of three participants as the final score of the measure for this group.

Data analysis

We chose a qualitative approach to gain insights into the qualities and variations in forms of interaction across different modes. A total of 36 videos from 18 groups were collected during the study. Two researchers watched and coded five videos together based on a prior coding scheme and process (Vuopala et al., 2016). Throughout the analysis, there are three main categories: Task-related, Group-related, and Off-task interactions and related subcategories and codes. The coding scheme was iteratively updated to better fit the study's learning content, rather than the original's. We used the new coding scheme to code another five videos and conducted the inter-rater reliability test. The final agreements from the Cohen's Kappa reached around 70%. Finally, we coded the rest of the videos individually, as well as counted the *occurrence* and the *proportion* of each item in coding scheme within each group to answer the RQ1. For RQ2, a one-way repeated measures ANOVA was conducted to check for significant effects of video control on the occurrence of coding scheme items in different levels. For RQ3, we used multiple regression analyze (Mason & Perreault, 1991) to identify which forms of interaction have a significant impact on the selected measures, and how much of an impact they have. We added "video technology mode" as one feature because it might influence the selected measures and used Dummy Coding to process categorical variables in multiple regression. We did the feature selection using correlation matrix with all codes in different levels to determine which features were dependent to improve the accuracy of our model and avoid overfitting. After selecting the features based on the correlation matrix, we then performed a multiple regression model to analyze the relationship between the selected features and the response measures.

Results

RQ1: What forms of interaction occur during collaborative VR video-based learning?

We collected a total of 1350 codes from 36 videos to analyze the forms of interaction and their proportions during collaborative VR video-based learning. Our coding scheme (Table 1) was adapted from a previous study (Vuopala et al., 2016). We found that *task-related interactions* were predominant, accounting for more than half of the interactions (50.89%). *Group-related interactions* (48.74%) were nearly as prevalent, highlighting the social and organizational aspects of collaboration. *Off-Task interactions* were minimal (0.37%), indicating occasional deviations from the primary focus. In terms of subcategories, *socio-emotional expressions* made up 25.49% of interactions, emphasizing the importance of emotional aspects in group dynamics. *Coordination of group activities* accounted for 23.25% of interactions, showcasing a structured approach to collaboration. Then, *answer or comment* (19.63%), *question* (19.04%), and *new knowledge* (12.22%) had relatively lower frequencies during VBL. At the code level, *declaratory comments* (12.59%), *organizing ongoing activities* (10.44%), and *expressing cohesion* (10.3%) were the most prevalent interaction forms.

Table 1

Coding Scheme for Forms of Interaction and Their Proportions during Video-based Learning

Main category	Sub category	Code	Coding rule	Example
Task-related interaction (50.89%)	Answer or comment (19.63%)	Declaratory comment (12.59%)	Agrees, states, repeats	"Okay. That sounds good."
		Comment with explanation (7.04%)	Explains, justifies, clarifies	"Maybe just rewatch, because we can discuss a little bit when revisit the content…"
	Question (19.04%)	Clarifying question (7.04%)	Clarifies previous question or asks for clarification	"What did you say about this building?"
		New question (6.44%)	Brings new question into the discussion	"What's the year of this museum?"
		Suggestion (5.56%)	States or suggests and waits for comments.	"You guys wanna watch the last few?"
	New knowledge (12.22%)	Content-based (10.15%)	Brings new topic based on the video content	"The Portuguese tiles are the bits of reference to the walls."
		Experience-based (2.07%)	Brings new topic based on experience or opinion	"This was on the pre-questionnaire."
Group-related interaction (48.74%)	Socio-emotional expressions (25.49%)	Expressing cohesion (10.3%)	Helping, rewarding, acknowledging	"That's a good point. Very cool."
		Accompanying (9.04%)	Expressing presence, mumbling	"...Moorish... fort… (repeat the video content)"
		Decreasing tension (6.15%)	Laughing, joking	"Ah, Gothic oyster, that's close enough but I cannot eat. (laugh)"
	Coordination of group	Organizing ongoing activities (10.44%)	Planning and organizing current group activities	"Ok, I'm gonna pause it and you take notes of the things she said about this."

	activities (23.25%)	Organizing upcoming activities (4.74%)	Planning and organizing future group activities	“I don't know exactly where anybody wants to go but just skip back to someplace you need to review.”
		Technological issues (4.15%)	Technological challenges, use of technology	“I was trying to delete my notes because I realized that it did not interpret what I said at all.”
		Evaluating group work (2.59%)	Evaluating group work	“Our notes so far probably missed a good chunk of information”
		Reporting current activities (1.33%)	Reporting, acknowledging current self-activities to group	“Um, I’m writing it down.”
Off-task interaction (0.37%)	Off-task (0.37%)	Off-task (0.37%)	Topics that are not related to course content or group work.	“Wee I'm floating!”

RQ2: How do forms of interaction vary between the shared control (Sync) and individual control (Non-sync) in collaborative VR video viewing systems?

We investigated the difference of occurrences and proportion of interaction forms under two collaborative conditions (Sync and Non-sync mode) in different levels of coding scheme.

Occurrence of interaction forms

The occurrence reflects the number of times an interaction takes place. The results showed that the mean of *task-related interaction* occurrences in Non-sync mode was statistically significantly lower than that of the Sync mode (p < 0.01, d =1.09). Within *task-related interactions*, the subcategory of *new knowledge* occurrences had a statistically significant difference (p = 0.011, d =0.84), with the sync mode having a higher mean than the Non-sync mode. When further broken down, the code *content-based* has a statistically significant (p = 0.01, d = 0.89) difference between the frequency of occurrence of the two modes, with Sync having a higher frequency. Likewise for the subcategory *question*, with the Sync mode mean with a higher statistically significant (p < 0.01, d = 1.00) mean number of occurrences than the Non-sync mode. The codes within this subcategory followed suit, with *new question* (p < 0.01, d = 1.14) and *suggestion* (p < 0.01, d = 1.28) both having a statistically significant higher mean number of occurrences in the Sync mode over the Non-sync mode. Another code within this subcategory was *clarifying question*. While this code did not have a strong statistically significant difference (p = 0.06, d = 0.71), we still saw that it had a higher mean number of occurrences in Sync as compared to Non-sync mode. The subcategory *answer or comment* shared similar values (p < 0.01, d = 1.02), with the mean number of occurrences for Sync mode statistically significantly higher than the Non-sync mode. Within this subcategory, we saw that both codes *declaratory comment* (p = 0.016, d = 0.98) and *comment with explanation* (p = 0.04, d = 0.61) had a statistically significant difference between the two modes. Again, we saw that both codes had a higher mean frequency of occurrence in the Sync mode when compared to the Non-sync mode. No other subcategories nor codes within task-related interactions saw any other significant differences.

Although we didn’t find the statistical significance in *group-related interaction* between the two modes, the results show that Sync mode brings a higher frequency *group-related interaction* (p = 0.056, d = 0.65). Within the category, the subcategory *coordination of group activities* has a significantly significant (p = 0.022, d = 0.83) higher mean number of occurrences for the Sync compared to the Non-sync mean. Similarly, the code *organizing ongoing activities* from this subcategory had a statistically significant difference (p = 0.012, d = 1.00) in the mean frequency of occurrence between the two groups, with Sync mode having the higher mean. While the subcategory *socio-emotional expressions* did not see a statistically significant difference between the two modes, the code *expressing cohesion* within that subcategory did. There was a statistically sufficient difference (p = 0.026, d = 0.80) between the means. Once again, the Sync mode saw a higher mean than the Non-sync mode. Within this category, no other codes nor subcategories had any significant differences. We did not see any significant differences in the *off-task* category.

Proportion of interaction forms

Proportion refers to the fraction or ratio of the number of times an interaction occurs relative to the total number of interactions. The results also demonstrate that the proportion of *task-related interactions* in Non-sync mode was statistically significantly lower than that of the Sync mode (p = 0.022, d = 0.62). Breaking this down, the subcategory *answer or comment* has a significant difference (p = 0.021, d = 0.88) between the Sync and Non-sync mode, with Sync being higher. Looking at the individual codes within this subcategory, *declaratory comment* is the only one with a significant difference (p = 0.017, d = 0.85), Sync having a higher mean proportion than Non-sync. Whilst the subcategory *questions* show no significant difference in mean proportion, the code *suggestion* also shows a significant difference (p = 0.052, d = 0.75), with Non-sync having lower mean proportion of occurrences than Sync mode. The categories *group-related interaction* and *off-task interaction* did not show any significant changes. No other subcategories nor codes show any significant differences.

RQ3: How do forms of interaction relate to the measures of knowledge acquisition and satisfaction?

Knowledge acquisition

For *auditory knowledge*, we didn't find statistically significant independent variables in main categories and subcategories for auditory knowledge acquisition over occurrence. For the occurrence of codes, it was found that occurrence of *content-based* (β = 0.0195, p = 0.024) and *reporting current activities* (β = 0.1229, p = 0.012) significantly predicted audio-based knowledge acquisition with positive effects. This results came from the regression model with weak statistical significance, $F(11, 22) = 1.906$, $p = 0.095$, $R^2 = 0.232$. For the proportion, the multiple regression analysis did not reveal any statistically significant relationships between the audio-based knowledge acquisition and the proportions of main categories and subcategories. For the proportion in code level, it was found that *organizing upcoming activities* (β = 5.948, p = 0.008) and Sync mode (β = 0.1736, p = 0.042) significantly predicted audio-based knowledge acquisition, suggesting a positive impact. These results came from a non-statistical significance regression model, $F(17, 16) = 1.615$, $p = 0.172$, $R^2 = 0.632$.

For *conceptual knowledge*, we didn't find statistically significant independent variables in main categories and subcategories for these conceptual knowledge acquisition over occurrence. For the occurrence of all codes, it was found that the *experience-based* (β = -0.08, p = 0.012) significantly predicted conceptual knowledge acquisition, suggesting negative effects. This is resulted from a non-statistical significance regression model, $F(11, 22) = 1.228$, $p = 0.326$, $R^2 = 0.38$. For the proportion, we didn't find statistically significant independent variables in main categories and subcategories. It's worth noting that in subcategories over proportion, $F(7, 26) = 1.690$, $p = 0.155$, $R^2 = 0.313$, *question* (β = 0.7941, p = 0.035) had a positive effect, and *coordination of group activities* (β = -0.4593, p = 0.040) had a negative effect, both statistical significantly influencing conceptual knowledge acquisition. When considering all codes, the model did not reach statistical significance, $F(17, 16) = 1.795$, $p = 0.124$, $R^2 = 0.656$. However, it was found that *clarifying question* (β = 1.3765, p = 0.040), *declaratory comment* (β = 1.3235, p = 0.033), *organizing ongoing activities* (β = -1.3235, p = 0.011) significantly predicted conceptual knowledge acquisition for the proportions of all codes.

For *visual knowledge,* we didn't find statistically significant for the regression model nor independent variables in three levels influencing the occurrence or proportion of virtual knowledge acquisition.

Satisfaction

There is no statistically significant difference in regression models on the occurrence of main category and sub category for predicting satisfaction. We found that the regression for satisfaction over occurrence of all code was statistically significant, $F(11, 22) = 4.455$, $p = 0.0014$, $R^2 = 0.69$. It was found that the occurrence of codes *content-based* (β = 0.0441, p = 0.004), *decreasing tension* (β = 0.0591, p = 0.046), *accompanying* (β = 0.0754, p = 0.005) significantly predicted satisfaction, suggesting a positive effect on satisfaction. However, *technological issues* (β = -0.0591, p = 0.046), *reporting current activities* (β = -0.3927, p < 0.001) and Non-sync mode (β = -0.2572, p = 0.040) indicated a negative impact on satisfaction. No independent variables were found to significantly predict satisfaction in any of the proportion models.

Conclusion and discussion

By designing and evaluating two collaborative VR video viewing systems — one with shared video control (Sync mode) and the other with individual control (Non-sync mode) — we have explored the dynamics of collaborative interactions in VR video-based learning and interaction forms' impact on knowledge acquisition and learner satisfaction. This exploratory study provides the foundation for additional research of understanding forms of interaction in CSCL environments in VR and optimizing the design the systems and pedagogical strategies.

Our findings from RQ1 reveal that task-related interactions played a pivotal role in the collaborative VR VBL environment, constituting the majority at 50.89%. This dominance underscores the fundamental nature of task-related discussions within the learning process. Interestingly, group-related interactions closely followed, accounting for 48.74% of the interactions. These interactions emphasize the significance of the social and organizational dimensions inherent in collaborative learning, indicating learners' active engagement in group dynamics, which is similar to prior CSCL interaction forms research (Vuopala et al., 2016). When we delved into the subcategories and codes of interactions, the results underscore the crucial role of emotional elements (e.g., *socio-emotional expressions*) and structured and organized group activities (e.g., *coordination of group activities*) in group dynamics the overall learning experience. This result implies that when designing collaborative learning environments, supporting task-related interactions and group-related interactions from both instructional and technical perspectives are very important.

RQ2 examines the variations in occurrence and proportion of interaction forms within two collaborative conditions. Based on our results, Sync mode appears to offer several advantages over Non-sync mode in the context of CSCL: it suggests more dynamic knowledge exchange, active questioning and discussion, and better coordination among learners compared to Non-sync mode. The major reason is that Sync mode provides the shared context for participants, so they're more comfortable communicating with each other and spending more time on discussion. When came to group conversation structure (aka. proportion). Sync mode has a higher proportion of task-related interactions compared to Non-sync mode. However, it is interesting that the group-related interaction and off-task interaction proportion of the whole discussion did not show any significant changes between modes. The result implies that when designing collaborative VR learning experiences, emphasizing synchronous communication and a shared context can lead to more knowledge exchange related to task.

From the results of RQ3, fostering content-based interactions, especially those involving new knowledge, appeared to enhance auditory knowledge acquisition. Additionally, effective group coordination through activities like organizing upcoming tasks and reporting current activities positively influenced auditory knowledge acquisition. Moreover, prioritizing the proportion of organizing upcoming activities during discussions was linked to increased audio-based knowledge acquisition. Surprisingly, Non-sync mode, which allows asynchronous interactions, seemed more conducive to auditory knowledge acquisition compared to Sync mode. When it comes to conceptual knowledge, encouraging students to ask questions, particularly clarifying ones, positively impacted learning. However, an excessive focus on group coordination activities appeared to hinder conceptual knowledge acquisition, suggesting that balance is crucial. Regarding satisfaction, content-based interactions were found to have a positive influence, emphasizing the importance of substantive discussions. Social interactions, such as accompanying and reducing tension significantly enhanced satisfaction. However, the technological issues and an overemphasis on reporting current activities negatively affected satisfaction. Interestingly, Sync mode was associated with higher satisfaction levels than Non-sync mode, possibly due to real-time interactions.

This exploratory research still requires confirmation from field studies and other subjects or educational scenarios. The most significant limitation of this study is its small sample size. Due to this limitation, our regression models about knowledge acquisition and satisfaction may not be strong enough to reach statistical significance as a group. While the models may not be significant, we still reported some individual independent variables which were found to be significantly explanatory for a dependent variable. This suggests that while those independent variables may be related to the dependent variable, the entire set of independent variables may not be jointly related. Additionally, it is possible that there may be interaction effects or nonlinear relationships between the independent variables and the dependent variable that are not captured by the current model. It is also possible that the sample size is too small to detect a significant effect of the independent variables on the dependent variable. So, due to the limited sample size, it was difficult to form a convincing conclusion. This work was also subject to some limitations regarding contextual factors which may influence forms of interaction between its participants. Specifically, relationships between team members, the nature of the learning content or tasks, and the duration of the learning session vary across different learning contexts and may play a role in forms of interaction which occur. The findings of this study may not be directly generalizable to situations with distinct team dynamics, content domains, task complexities, or study durations. Further, the within-subject design of this study introduces a potential limitation through the risk of carryover effects, as participants' experiences with one control mode may influence subsequent ones, potentially leading to order effects and impacting the validity of results. Furthermore, the study's controlled laboratory environment may limit its generalizability to real-world educational contexts. Factors such as individual infrastructure, access to technology, and classroom dynamics, which play significant roles in authentic educational settings, are not fully captured in this controlled environment.

Therefore, we suggest future research to address these limitations. Specifically, conducting larger-scale studies with diverse participant groups and considering various educational scenarios will help validate and generalize the findings. Additionally, future research should include a thorough power analysis to determine the necessary sample size to achieve statistical significance in regression models. This will help ensure that studies are adequately powered to detect meaningful effects and draw more robust conclusions. Researchers could also explore potential interaction effects and nonlinear relationships among variables and employ more advanced statistical techniques to account for these complexities. To enhance the external validity of the findings, future research should also consider contextual factors (e.g., content domains, task complexities, and study durations). Comparative studies across different educational settings and learning contexts will provide a more comprehensive understanding of how forms of interaction vary. Moreover, researchers should adopt a mixed-methods approach, combining quantitative data with qualitative insights to gain a deeper understanding of the dynamics of interaction in VR-based collaborative learning. This approach can capture nuances that quantitative measures may overlook. Lastly, future studies should aim to replicate collaborative technologies in more authentic educational environments. This will ensure that the results are applicable to real-world educational scenarios, providing valuable insights for educators and instructional designers.

Endnotes

(1) In this work, “VR videos” specifically refer to monoscopic 360-degree videos experienced through a VR headset.
(2) “One day in” 360° travel videos: https://www.youtube.com/playlist?list=PLHiCdB8YTO76Gv843e8rdcza-FJbYCYSj

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives. Longman.*

Fang, J., Wang, Y., Yang, C. L., Liu, C., & Wang, H. C. (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), 1-21.

Goldman, R. (2014). ORION™, An Online Digital Video Data Analysis Tool: Changing Our Perspectives as an Interpretive Community. In *Video research in the learning sciences (pp. 507-520)*. Routledge.

Jackson, R. L., & Fagan, E. (2000, September). Collaboration and learning within immersive virtual reality. In *Proceedings of the third international conference on Collaborative virtual environments (pp. 83-92)*

Jin, Q., Liu, Y., Yarosh, S., Han, B., & Qian, F. (2022). How will VR enter university classrooms? multi-stakeholders investigation of vr in higher education. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Jin, Q., Liu, Y., Sun, R., Chen, C., Zhou, P., Han, B., ... & Yarosh, S. (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 (pp. 1-18).*

Kallioniemi, P., Keskinen, T., Hakulinen, J., Turunen, M., Karhu, J., & Ronkainen, K. (2017, November). Effect of gender on immersion in collaborative iodv applications. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (pp. 199-207).*

Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior, 28*(3), 820-831

Lebie, L., Rhoades, J. A., McGrath, J. E. (1996). Interaction process in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work, 4*(2–3), 127–152.

Li, N., Verma, H., Skevi, A., Zufferey, G., & Dillenbourg, P. (2014). MOOC learning in spontaneous study groups: Does synchronously watching videos make a difference? (No. CONF, pp. 88-94). *PAU Education.*

Mason, C. H., & Perreault, W. D. (1991). Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research, 28*(3), 268-280.

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

Pirker, J., & Dengel, A. (2021). The Potential of 360-Degree Virtual Reality Videos and Real VR for Education-A Literature.

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

Saffo, D., Bartolomeo, S. D., Yildirim, C., & Dunne, C. 2021. Remote and Collaborative Virtual Reality Experiments via Social VR Platforms. Association for Computing Machinery, New York, NY, USA.

So, H.-J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education, 51*(1), 318–336.

Torff, B., & Tirotta, R. (2010). Interactive whiteboards produce small gains in elementary students’ self-reported motivation in mathematics. *Computers & Education, 54*(2), 379-383.

Vuopala, E., Hyvönen, P., & Järvelä, S. (2016). Interaction forms in successful collaborative learning in virtual learning environments. *Active Learning in Higher Education, 17*(1), 25-38.