International Society of the Learning Sciences

An Exploratory Study of Using Interactive Volumetric Video in VR for Embodied Learning

Qiao Jin, Carnegie Mellon University, georgiej@andrew.cmu.edu
Yu Liu, University of Southern California, yliu2170@usc.edu
Yuxuan Huang, University of Minnesota, huan2076@umn.edu
Bo Han, George Mason University, bohan@gmu.edu
Feng Qian, University of Southern California, fengqian@usc.edu
Svetlana Yarosh, University of Minnesota, lana@umn.edu

Abstract: Volumetric video (VV) has the potential to revolutionize traditional video-based learning (VBL) by offering immersive, 3D content that enhances student engagement and comprehension. However, the limited interactivity of current pre-recorded VV restricts its educational applications. To address this, we developed and evaluated an interactive VV viewing system that enables learners to engage with spatial interactive areas. These areas trigger multimodal outcomes and support both static and continuous interactions tailored to specific educational objectives. In this work, we utilized the Lego construction learning task as an exploratory caser. Compared to the basic condition (VV without interaction), the interactive VV condition demonstrated higher germane cognitive load and interest in learning but showed no significant differences in short-term memory retention.

Introduction

While traditional video-based learning (VBL) has been a key component of online education for years, incorporating VBL into virtual reality (VR) creates new educational possibilities to transform how students engage with video content, reducing environmental distractions and fostering social presence. One of the most widely utilized VR video formats in education is 360-degree video, which has been the focus of substantial research exploring its applications across a range of educational domains presence (Jin et al., 2024). A key limitation of 360-degree (and other 2D) videos in VR is the lack of depth information and limited camera viewpoint flexibility, which confines students to the original perspective and hinders spatial exploration (Y. Jin et al., 2023; Q. Jin et al., 2023a). Volumetric video (VV), also known as free-viewpoint video, is a technique that captures real-life objects or people directly using multi-view photography or depth sensing, offering new opportunities in VBL by presenting more immersive and spatially rich content. Compared with other computer-generated VR environments, VV is more cost-effective for content creation, as it does not require advanced skills in 3D modeling, animation, or computer graphics to build realistic environments, which are often resource intensive. VV has been utilized to support educational goals by creating more authentic and engaging learning experiences for students such as virtual tours, healthcare training, and creative storytelling (Liu et al., 2024).

Prior studies have shown that interactivity in videos leads to better learning performance and higher learner satisfaction compared to non-interactive videos, while also increasing learners' attention spans (Zhang, Zhou, Briggs, & Nunamaker, 2006; Ploetzner et al., 2024). Viewing VV in VR offers an inherently interactive experience by providing six degrees of freedom (DoF) for embodied learning (Jin et al., 2024). Although there are multiple forms of interaction in traditional videos, interaction with pre-recorded VV remains largely restricted to basic functions, such as play, pause, and adjusting the viewing angle or position. Additionally, pre-recorded VV has limited support for real-time interaction or manipulation of digital objects by the user. This articulates a research gap in enhancing the interactivity of pre-recorded VV and validating its effectiveness in learning. To address this, we developed and evaluated an interactive VV (IVV) viewing system, using a Lego construction task as an exploratory example for embodied learning. The system enables students to interact with VV through spatial interactive areas (IAs), receiving multimodal feedback (e.g., text, images, video branches), thereby expanding the potential and utility of VV in education. Our contributions consist of an evaluation of an IVV viewing system, comparing its impact on short-term memory retention, cognitive load, confidence and interest in learning with conventional VV in VR. We also discussed the implications of using IVV for educational purposes.

Methods

Participants and settings

We conducted a within-subject experiment with 20 participants (ages 22–41, $M = 27.9$, $SD = 5.08$), recruited from a college student mailing list. The sample comprised ten males, nine females, and one individual identifying as

non-binary. In terms of VR expertise, four participants were self-identified as experts, nine as possessing passing knowledge, three as knowledgeable, and four as having no knowledge. Each participant experienced an IVV condition, and a Basic VV condition without any interactions in counterbalanced order using Oculus Quest 1, and received $15 gift cards as compensation.

Learning content and volumetric video-based viewing systems

We used Lego construction tasks (Richardson, Hunt, & Richardson, 2014) as the learning content in the volumetric videos for three reasons: (1) They leverage the 3D affordances of volumetric video, requiring spatial understanding and visuospatial memory (McDougal et al., 2023); (2) From the education perspective, construction play develops spatial sense which is critical for math and science learning (Zosh, Hassinger-Das, & Laurie, 2022); and (3) From a practical standpoint, VR is valuable for remote manufacturing training and assembly instructions (Doolani et al., 2020). Lego construction mirrors these tasks in spatial and procedural aspects but with reduced complexity, making it suitable for diverse participants and time-limited exploratory lab studies. We followed the guidelines from Richardson and Hunt's work (Richardson, Hunt, & Richardson, 2014) to design the Lego construction tasks and tested their difficulty through a pilot with five volunteers. Finally, formal study used two tasks built from six Lego pieces selected from a pool of eleven (see Figure 1), while a simpler tutorial task used five pieces from a pool of nine to familiarize participants with the system.

Figure 1

One Example of a Lego Construction Task

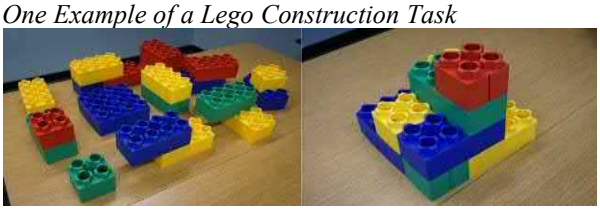
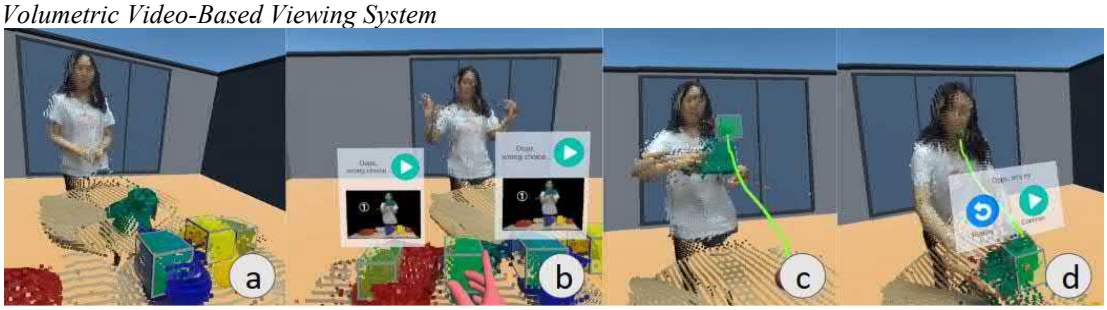


Figure 2

Volumetric Video-Based Viewing System



For each task, we recorded a VV where an instructor demonstrates the construction procedure. The videos were approximately two minutes long, consisting of around 1800 frames at a frame rate of 15 FPS. Each video had two presentation formats corresponding to the two conditions of the study, i.e. Basic VV, and IVV. For the Basic condition, the video could only be played and paused by the learners. For the IVV, learners can interact with the virtual instructor through the interaction areas (IAs), which are highlight-specific areas in 3D space. Two types of IA are described below; more technical details are described in our prior work (Q. Jin et al., 2023b).

- *Static IA* (Figure 2.a): When the instructor is selecting a component, the video pauses. Learners must choose the correct component by triggering the corresponding *static IA*. If correct, the video continues; if not, a pop-up message (Figure 2.b) displays the correct component. 20 *static IAs* were integrated.
- *Continuous IA* (Figure 2.c): When the instructor is placing a component, the video pauses. Learners must follow the path that specifies the placement of the component with their hand. If successful, the video continues; if not, users can replay the path and try again (Figure 2.d). 6 *continuous IA* were integrated.

Measures

The study was designed to assess whether adding interaction into a standard VV viewing system influenced the participants' experiences on the following measures: *Short-term memory retention* is the ability to remember information in the short term after its presentation (Norman, 1966). It can be tested in two ways: recall and recognition. For this study, researchers recorded the number of correctly chosen components for recognition

memory, and the number of correctly placed components for recall memory. Both values were between 0-6. *Cognitive load*, *confidence* and *interest in learning* were measured using a total of 7 questions. These variables are essential to academic success, including future course-taking and performance (Harackiewicz, Smith, & Priniski, 2016; Jin et al., 2024b). *Cognitive load* was measured as intrinsic (ICL), extraneous (ECL), and germane (GCL) cognitive load (Cierniak, Scheiter, & Gerjets, 2009). The *confidence* and *interest in learning* were measured by four questions adapted from Blair's work (Blair & Frezza, 2020). These questions assess an emotional response ("I feel confident that my Lego construction is correct and well-made"; "This course was engaging and interesting.") and behavioral planning ("I am confident that I can independently study and learn this topic in the future"; "This course makes me interested in learning more in the future.").

Procedure

The study took approximately 50 minutes to complete and included the following steps: Participants were first introduced to the study's goals, procedures, and systems (5 minutes), then completed consent forms and a training session (10 minutes) where they donned the VR headset and familiarized themselves with both conditions and the learning task. The main phase (25 minutes) involved two learning units—one per condition—each consisting of: (1) a 5-minute video viewing session, which participants with the ability to pause freely and typically enough time to watch it twice; (2) a hands-on construction session, where participants assembled physical Lego models based on the video content. A researcher observed the construction session and recorded the correctness of the selection and placement of each component; and (3) a questionnaire assessing cognitive load, confidence, and interest in learning. Finally, a 10-minute interview captured their experiences, challenges, and suggestions for improvement.

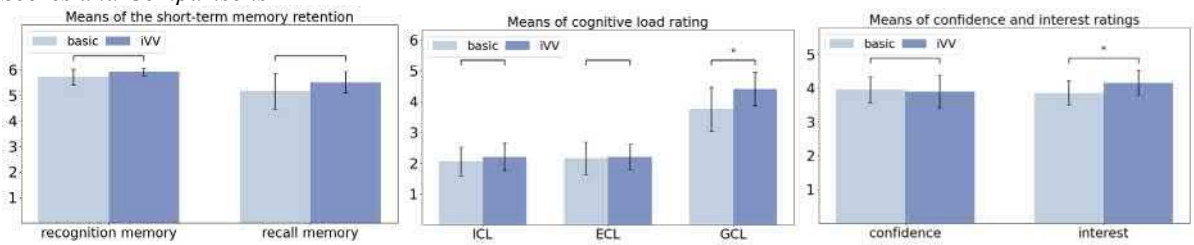
Results

Qualitative results

We used the Anderson-Darling test to assess normality for quantitative data, which suggested that none of them was normally distributed. Thus, we used the Wilcoxon Signed-Rank Test to perform pairwise comparisons. We reported our results at the 0.05 significance level, and calculated Cohen's *d* to measure the effect size.

Figure 3

Scores and Comparisons



Asterisk (*) indicates a statistically significant difference between conditions: $p < .05$ (*).

We measured ***short-term memory retention*** through both actions (recall memory) and components selection (recognition memory). Although we see both these two measures in IVV condition ($M_{recall} = 5.9$, $SD_{recall} = .31$; $M_{recognition} = 5.5$, $SD_{recognition} = .89$) has higher correctness than the Basic condition ($M_{recall} = 5.7$, $SD_{recall} = .66$; $M_{recognition} = 5.15$, $SD_{recognition} = 1.5$), we did not observe any significant differences for recall and recognition memory (Figure 3). The ***cognitive load*** questionnaire measured intrinsic cognitive load (ICL), extraneous cognitive load (ECL) and germane cognitive load (GCL). Our results did not reveal statistically significant differences between conditions on means of ICL and ECL. However, we found a significant difference in the means for the GCL between two conditions ($p = .012$, $d = .48$). That suggests participants invested more resources and focused more intently on the learning process in the IVV condition compared to the Basic condition. This increased concentration in the IVV condition is advantageous for learning, as it helps learners to focus on relevant processes. The analysis of ***confidence in learning*** scores did not reveal a significant difference between two conditions. The mean confidence score for the Basic condition was 3.95 ($SD = .83$), while the mean confidence score for the IVV was 3.90 ($SD = 1.03$). The analysis revealed a significant difference in ***interest in learning*** between two conditions ($p = .02$, $d = .64$). The mean of interest in learning score was higher for the IVV condition ($M = 4.15$, $SD = .78$) compared to the Basic condition ($M = 3.85$, $SD = .76$), suggesting that participants found the IVV condition more engaging.

Qualitative Results

For the qualitative data, we processed the data using data-driven thematic analysis, inspired by Grounded Theory. Two authors generated and reviewed open codes, clustering them with an affinity map and refining clusters iteratively until themes emerged. Disagreements were resolved through discussion or with a third author. All authors then identified the most relevant themes, which are presented below.

Theme 1: Realistic representation and interaction influence engagement and interest in learning

Almost all participants agreed that the realism of representation and interaction in IVV contributed significantly to their engagement and interest in learning. In both conditions, participants appreciated the realism of VV's one-to-one scale and high fidelity, distinguishing it from other mediums like 360-degree video (P8). It provided rich spatial, shading, and detail information, allowing for exploration from multiple angles (P10, P15). Participants particularly valued IVV for its intuitive interaction and active engagement compared to Basic VV. P6 emphasized the ability to trigger the continuous IA, stating, "I liked that I could see the trajectory in the air—it felt like a natural way to guide my actions, similar to real-world tasks". Similarly, P8 emphasized the value of static IA as it created a sense of engaging with "actual objects." Interaction with objects exhibits higher levels of realism compared to interactions with human (P16). This discrepancy arises due to the absence of eye contact and other spontaneous reactions in the video, which is important for creating authentic interpersonal connections.

Theme 2: Interactions in volumetric videos might introduce distractions

While IVV are effective in promoting engagement, participants noted that when the interactive elements were not closely aligned with the learning objectives (in our case, correctly choosing and placing the components), they could introduce unnecessary distractions. For instance, some participants found that the requirement to precisely follow a trajectory distracted them from understanding the primary objective—placing the Lego components in the correct location. As one participant noted, "I felt like I was focusing too much on the movement, and that distracted me from remembering the shapes of the Lego model" (P12). Another added, "The trajectory didn't match how I would actually perform the task, so it felt less useful for remembering" (P16). This mismatch between the design of the interactive element and the intended learning goal created moments of frustration and confusion, with one participant noting, "When the system said I was doing it wrong, I kept re-trying to figure it out, but it broke my concentration on the learning task" (P2). In addition, P6 noted that pauses to interact in IVV sometimes disrupted his own learning pace, stating, "I felt like all the pauses…weren't really needed to remember everything."

Discussion

The study offers key implications for using IVV in education, especially for embodied learning. Although no significant gains in memory or confidence were observed from questionnaires, the increased GCL and interest in learning with IVV suggests positive effects on learner engagement and increased concentration on the learning content, a finding further validated by our qualitative results (Theme 1). This suggests that educators can use IVV to foster active participation, particularly in lessons that require hands-on tasks, enhancing the motivation. This increased interest and engagement also have the potential to support long-term retention of learning materials (Nemati, 2009). Future research could explore whether the interactive features in VV influence long-term memory retention. Additionally, since working memory is influenced by age (Light & Anderson, 1985) and learning task, future research could investigate the effects of interactive features on memory retention or other learning outcomes across age groups and training activities. To maximize its benefit, future technologies should offer additional interaction modalities that replicate or enrich real-world interactions, such as eye contact. For example, one work (Son et al., 2020) introduced an MR system designed to simulate lifelike eye contact in a human avatar.

Our results did not show a statistically significant improvement in short-term memory retention with IVV. Qualitative data (Theme 2) suggest that one potential reason is that pauses for interaction in IVV may disrupt learners' own learning pace, causing distractions. Furthermore, misaligned interactions—such as emphasizing precise movements that are not central to the learning task—can divert attention and hinder learners' ability to process and retain information effectively. These highlight the need to design interactive features that align closely with the intended learning objectives. Future work should explore more personalized IVV systems that adapt interaction types, timing (e.g., pause vs. play during interactions), and frequency of interactions based on learners' prior knowledge and individual learning preferences. In addition, our results showed no significant improvement in learning confidence with IVV, possibly due to factors like VR novelty (e.g., discomfort with interactive features), or individual differences (e.g., prior familiarity with Lego construction). Future work should explore strategies to reduce the impact of technological novelty and support diverse learner profiles, such as offering more comprehensive training or adjusting interaction levels to match user comfort.

References

Blair, M., & Frezza, S. (2020). Assessing interest and confidence as components of student motivation in informal STEM learning. In *2020 IEEE Frontiers in Education Conference (FIE)* (pp. 1–5). IEEE.

Cierniak, G., Scheiter, K., & Gerjets, P. (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), 315-324.

Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., & Makedon, F. (2020). A review of extended reality (XR) technologies for manufacturing training. *Technologies, 8*(4), 77.

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 220–227.

Jackson, W. (2016). The composition of digital video: Timeline editing. *Digital Video Editing Fundamentals,* 101–113. Springer.

Jin, Q., Liu, Y., Yuan, Y., Han, B., Qian, F., & Yarosh, S. (2024a). Virtual reality, real pedagogy: A contextual inquiry of instructor practices with VR video. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Article No. 665, pp. 1–21). Association for Computing Machinery.

Jin, Q., Sodhi, I., Chen, A., & Yarosh, S. (2024b). Interaction Forms of Collaborative VR Video Learning: An Exploratory Study. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024*, pp. 91-98. International Society of the Learning Sciences.

Jin, Q., Liu, Y., Sun, R., Chen, C., Zhou, P., Han, B., ... & Yarosh, S. (2023a, April). 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).

Jin, Q., Liu, Y., Zhou, P., Han, B., Yarosh, S., & Qian, F. (2023b, March). Volumivive: An authoring system for adding interactivity to volumetric video. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 569-570). IEEE.

Jin, Y., Hu, K., Liu, J., Wang, F., & Liu, X. (2023). From capture to display: A survey on volumetric video. *arXiv preprint arXiv:2309.05658*.

Liu, Y., Jin, Q., Zhang, Z., Han, B., Yarosh, S., & Qian, F. (2024, October). HoloClass: Enhancing VR Classroom with Live Volumetric Video Streaming. In *Adjunct Proceedings of the 37th Annual ACM* Symposium *on User Interface Software and Technology* (pp. 1-3).

McDougal, E., Silverstein, P., Treleaven, O., Jerrom, L., Gilligan-Lee, K. A., Gilmore, C., & Farran, E. K. (2023). Associations and indirect effects between Lego® construction and mathematics performance. *Child Development, 94*(5), 1381–1397.

Navarro, D., & Sundstedt, V. (2017). Simplifying game mechanics: Gaze as an implicit interaction method. In *SIGGRAPH Asia 2017 Technical Briefs* (pp. 1–4).

Nemati, A. (2009). Memory vocabulary learning strategies and long-term retention. *International Journal of Vocational and Technical Education, 1*(2), 14–24.

Newbury, R., Satriadi, K. A., Bolton, J., Liu, J., Cordeil, M., Prouzeau, A., & Jenny, B. (2021). Embodied gesture interaction for immersive maps. *Cartography and Geographic Information Science, 48*(5), 417–431.

Norman, D. A. (1966). Acquisition and retention in short-term memory. *Journal of Experimental Psychology, 72*(3), 369.

Ploetzner, R. (2024). The effectiveness of enhanced interaction features in educational videos: A meta-analysis. *Interactive Learning Environments, 32*(5), 1597–1612.

Richardson, M., Hunt, T. E., & Richardson, C. (2014). Children's construction task performance and spatial ability: Controlling task complexity and predicting mathematics performance. *Perceptual and Motor Skills, 119*(3), 741–757.

Son, J., Gül, S., Bhullar, G. S., Hege, G., Morgenstern, W., Hilsmann, A., Ebner, T., Bliedung, S., Eisert, P., Schierl, T., Buchholz, T., & Hellge, C. (2020). Split rendering for mixed reality: Interactive volumetric video in action. *In Proceedings of SIGGRAPH Asia 2020 XR* (Article 8, pp. 1–3). Association for Computing Machinery.

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F., Jr. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management, 43*(1), 15–27.

Zosh, J. M., Hassinger-Das, B., & Laurie, M. (2022). Learning through play and the development of holistic skills across childhood. *Billund, Denmark: Lego Foundation*.