



Collaborative Tasks in Immersive Virtual Reality Increase Learning

Anna C. M. Queiroz, Stanford University, acmq@stanford.edu
Eileen McGivney, Harvard University, eileen_mcgivney@g.harvard.edu
Sunny X. Liu, Stanford University, sunnyxliu@stanford.edu
Courtney Anderson, Stanford University, court22@stanford.edu
Brian Beams, Stanford University, beams@stanford.edu
Cyan DeVeaux, Stanford University, cyanjd@stanford.edu
Kai Frazier, Kai XR, kai@kaixr.com
Eugy Han, Stanford University, eugyoung@stanford.edu
Mark Roman Miller, Stanford University, mrmillr@stanford.edu
Xander S. Peterson, The Hydrous, xwpeterson@gmail.com
Erika S. Woolsey, The Hydrous / Stanford University, erika@thehydro.us
Jeffrey T. Hancock, Stanford University, hancockj@stanford.edu
Jeremy N. Bailenson, Stanford University, bailenso@stanford.edu

Abstract: Advances in immersive virtual reality (IVR) are creating more computer-supported collaborative learning environments, but there is little research explicating how collaboration in IVR impacts learning. We ran a quasi-experimental study with 80 participants targeting ocean literacy learning, varying the manner in which participants interacted in IVR to investigate how the design of collaborative IVR experiences influences learning. Results are discussed through the lens of collaborative cognitive learning theory. Participants that collaborated to actively build a new environment in IVR scored higher for learning than participants who only watched an instructional guide’s avatar, or participants who watched the guide’s avatar and subsequently discussed what they learned while in IVR. Moreover, feeling negative emotions, feeling active in the environment, and feeling bonded to the group members negatively correlated with learning. Results shed light on the mechanisms behind how collaborative tasks in IVR can support learning.

Introduction

As technology evolves and becomes more accessible and affordable, it enables collaboration in different spheres. Advances in virtual reality (VR) are creating more immersive computer-supported collaborative learning environments. Desktop-based VR (i.e., a virtual environment is displayed on a computer monitor and interacted with through a keyboard, mouse, or joystick) has leveraged multi-user virtual environments to allow learners to collaborate and join simulations simultaneously for many years (Clarke & Dede, 2005). However, the emergence and increasing adoption of networked headset-based immersive virtual reality (IVR) in educational contexts raise questions about how this media can properly foster collaborative learning when surrounded by a virtual, rich, and interactive environment with peers.

In IVR, the sight of the surrounding physical world is replaced by digital content refreshed according to the user’s movements (Rebelo et al., 2012). Several IVR features make it more immersive than other media, such as motion tracking level, stereoscopic vision, field of view, and user perspective (Cummings & Bailenson, 2015). Those features contribute to the users’ sense of *presence* inside the virtual environment (i.e., the feeling of “being there,” Sanchez-Vives & Slater, 2005), provide opportunities for embodied interaction with objects and people, and trigger emotions (Markowitz & Bailenson, 2021). These affordances of IVR make it suitable to simulate learning experiences that would be dangerous, impossible, counter-productive, or expensive in the physical world (Bailenson, 2018) and in which physical or emotional responses are essential. One particularly promising application is in ocean education, as first-hand experiences with the ocean are impossible for some and expensive for most learners. Marine education is well suited to the affordances of IVR, as visiting underwater environments to observe the environment and animals significantly enhances learning. But doing so is limited by barriers including distance to the coast, physical limitations, inability to swim, fear, and costs associated with dive training and equipment. Previously, IVR applications have been shown to effectively build authentic learning environments and promote ocean literacy (Carrol & Tambe, 2022; Fauville et al., 2020).

Much research on learning with IVR used solo IVR experiences (e.g., Mayer et al., 2022; Queiroz et al., 2022). Overall, the evidence of IVR’s impact on learning is mixed, with some studies finding higher gains (Alhalabi, 2016; Webster, 2015) and others finding lower gains for learners using IVR (Mayer et al., 2022). Building on cognitive learning theory, one explanation for the mixed results is that IVR adds extraneous cognitive load and overwhelms learners’ working memory (Mayer et al., 2022; Van Merriënboer & Sweller, 2005),



especially when it includes multiple channels of information, emotional arousal, or graphics in the environment that are extraneous to the information learners must process (Mayer, 2014).

Cognitive load in collaborative learning environments is even less understood. According to the collaborative cognitive load theory (Kirschner et al., 2018), there is a positive effect on learning when learners require the input and effort of the group members to complete a task (namely, *positive interdependence*). In addition, when learners are aware of others and their expertise and combine this to share and process information more efficiently, the group develops a *transactive memory*, requiring less individual cognitive effort during the learning process compared to learning individually (Janssen & Kirschner, 2020). On the other hand, investing mental effort in interacting with others can increase cognitive load and negatively affect learning (namely, *transaction costs*). The trade-off between the transactive memory and cognitive load in interacting with others is what is known as the *collective working memory effect*.

The collaborative learning process is also affected by the extent to which group members perceive being bonded or their *entitativity* (Campbell, 1958), which results from the effort the group puts into working together. In a theoretical explanation of the relationship between antecedents (e.g., group or task characteristics), processes (e.g., transaction activities, invested mental effort), and consequences of collaboration (individual achievements), Janssen and Kirschner (2020) propose that the group- or task- characteristics would indirectly influence individual achievements through collaborative processes. However, it is not yet well understood how these processes relate and the mechanisms underlying learning, particularly in collaborative tasks in IVR. In this context, we predicted that participants collaboratively building the IVR environment would score higher for learning (H1) and entitativity (H2) than participants watching an instructional recording or watching and discussing it in IVR.

Drawing on prior research of individual learning in VR, additional stimuli in the form of others' avatars and additional 3D models would add extraneous cognitive load to individual learners and make it difficult for students to reap the benefits of IVR and learn within it without learning how to use IVR first (Han et al., 2022). Collaborative IVR learning will also be impacted by affective dimensions of learning, including whether learners feel frustrated or overwhelmed by interaction demands, influencing their *self-efficacy*, or whether they believe they can learn (Bandura, 1977). Studies focused on IVR single-user experiences have shown a positive effect of immersion and agency on self-efficacy (Makransky et al., 2020; Makransky & Petersen, 2021; Queiroz et al., 2022). However, how increased technological demands and cognitive load in collaborative IVR environments affect cognitive and affective aspects of learning has not been well studied. Hence, this study investigated the effects of the level of interaction in collaborative learning tasks in IVR on self-efficacy (RQ1) and the relations between positive and negative affect on learning, self-efficacy, and entitativity (RQ2). We predicted that participants building the IVR environment together would score higher for agency than participants viewing an instructional recording or viewing and discussing it together in IVR (H3).

While experiencing IVR solo usually can take only a few minutes, experiencing it in a group adds time to the experience. In addition, collaboratively building a virtual environment requires the users to familiarize themselves with the creation tools, increasing the time and cognitive load of the experience. These extended times in IVR can influence how much participants feel physiological discomforts usually associated with IVR, or *cybersickness* (such as nausea and eye strain), and their sense of presence (Han et al., forthcoming). Hence, we predicted that participants would report higher feelings of presence when building the virtual environment together, compared to just viewing an instructional recording or viewing and discussing it together in IVR (H4).

This study explored the design and effectiveness of a collaborative IVR ocean experience by comparing three conditions with different levels of interactivity: a highly interactive and collaborative condition in which participants watched an IVR-guided dive, interacted with each other, and collaboratively built a virtual reef habitat using 3D models, a medium interactive condition in which participants watched the guided dive and discussed it, and a low interactive condition in which participants just watched the guided dive but were not encouraged to interact with each other. We also assessed mechanisms through which the collaborative activity may enhance learners' experience, including heightened presence, entitativity, sense of agency, self-efficacy, and emotions.

Methods & Materials

We conducted a design-based research (Barab & Squire, 2004) experiment with the aim of both designing an intervention leveraging IVR's affordances for collaborative learning that is feasible to use in authentic education environments. The experiment utilized rapid cycles of iteration over three days.

Participants

Participants were attendees from diverse backgrounds who enrolled in a session at an institutional program in March 2022. A total of 107 participants enrolled in the study, and 80 whose pre- and post-data were complete and could be matched are included in this analysis. Participants' ages varied from 19 to 69 years old ($M = 26.34$, SD



= 10.09). 41 participants identified as female (51%), 38 as male (48%), and one as something else (1%). Participants identified their race as: 34% White, 30% Asian, 12% Latinx, 9% Black, 3% Middle Eastern, 6% more than one, and 6% declined to answer. They were from North America (46%), Asia (30%), Latin America (12%), and Europe (10%).

Procedure

Procedures and materials were approved by the institutional review board. Participants were welcomed in a conference room, where two researchers explained the study procedures. Participants completed a consent form and a pre-survey on their mobile phones. Afterward, participants created and customized their avatars using the Engage mobile app (ENGAGE, 2022). Participants then joined the Engage VR platform using Meta Quest 2 VR headsets and received an approximately five-minute-long tutorial on how to use the headset, controllers, and the IVR application. Participants were seated while in IVR. They were then randomly assigned to a small group of six to eight learners, and one facilitator guided them to a separate room according to their condition assignment. Participants could participate in the virtual dive only once and only in one of the three conditions. A maximum of six groups were formed per day of data collection, as we had six facilitators (four women and two men). Three assistants roamed across rooms to help if needed. To reduce the influence of facilitators' expertise on the results, each day each facilitator was assigned to a different condition. Each facilitator could moderate one session per day and a maximum of three sessions in the study. After completing the IVR dive, participants took the headset off and answered a post-survey in the same room they had the dive or in the conference room, on their mobile phones.

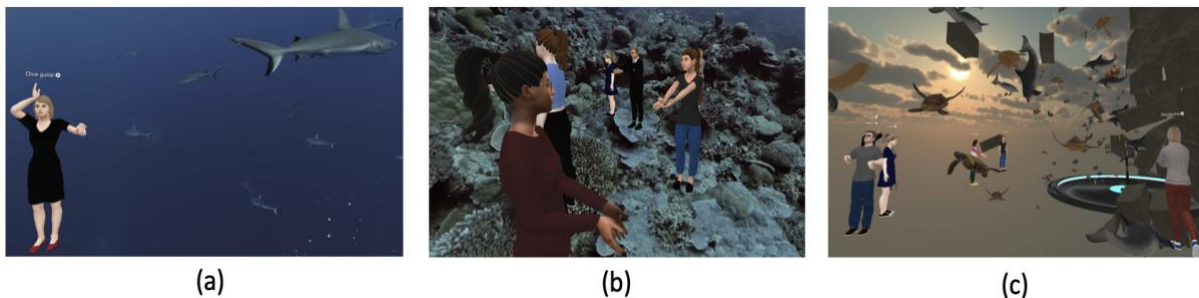
Designing a Dive Guide and collaborative learning VR experience

We developed an IVR experience of a guided ocean dive and designed a collaborative experience for learners to work together and with 3D models. The Dive Guide VR experience was created in Engage (ENGAGE, 2022), a collaborative VR platform. A virtual underwater environment was created using several 360-degree stereoscopic still photographs (also called "photospheres") of underwater scenes collected by the 501(c)3 nonprofit The Hydrous on Palauan coral reefs. We recorded a marine expert avatar (using the motion capture system embedded within the Engage platform) interacting with the photospheres while explaining about coral reefs and human activities' impacts on marine life. This recording lasted approximately ten minutes. In all conditions, participants watched the Dive Guide experience, joining the platform as an avatar, and the real-time movements of other learners rendered on avatars visible in the underwater environment. They all could move, talk, and see each other.

The recording was paused two times. In the collaborative tasks, learners were given 3D models of coral and marine animals and asked to work together to represent what they had learned. As data were collected over the three days, we employed an iteration cycle to respond to the needs of learners, for example, by removing the virtual stools to which learners were fixed on the first day to allow them to move more freely during the guide. While we had extensively pilot-tested the conditions with small groups, new challenges emerged when using large groups of novice users. The resulting experience met the technical constraints of widely available IVR hardware and software and maximized learners' interactivity and collaboration around a shared task. While the design of the learning experience varied by day, across all days participants were given 3D models to manipulate and use to apply their learning collaboratively. Screenshots of the Dive Guide and participants in the IVR environment watching the Dive Guide and collaborating are shown in Figure 1.

Figure 1

The Dive Guide and collaborative VR experience Screenshots



Note. (a) Dive guide screenshot; (b) Participants watching the Dive Guide (Viewing condition); (c) Participants in the collaborative task (Viewing/Discussion/Graphics condition).



Design

The collaborative learning experience described above is termed the “Viewing/Discussion/Graphics.” To assess the impact of different levels of interaction and collaboration on learning, we compared the Viewing/Discussion/Graphics condition to one condition with a lower cognitive load by omitting the 3D models but keeping the pauses and discussions (“Viewing/Discussion” condition) and one condition where no interaction was encouraged, but participants still had the experience in groups (“Viewing” condition), as detailed in Table 1. To ensure an adequate sample in the Viewing/Discussion/Graphics condition, the Viewing/Discussion condition was only used on day 1 of data collection, after debriefing about participants’ feedback on the first day of the study. To balance the experience pedagogy with participants’ familiarity with the technology, following Han and colleagues’ recommendations about learning collaboratively in IVR (Han et al., 2022), the facilitator taught participants how to use the IVR controllers and manipulate the 3D models before starting the Dive Guide recording.

Table 1
Conditions description and length

Conditions (day)	Description	n	Mean (SD) Length (minutes)
Viewing/Discussion/ Graphics (days 1, 2, and 3)	Participants watched the Dive Guide VR experience. The facilitator paused the recording two times to share 3D models with participants and asked them to collaborate to apply what they were learning with models of coral, marine animals, and building blocks for building.	45	31.35 (6.64)
Viewing/Discussion (day 1)	Participants watched the Dive Guide VR experience. The facilitator paused at the same times as the Viewing/Discussion/Graphics condition and asked participants to discuss what they were learning with each other.	11	28.06 (2.86)
Viewing (days 1, 2, and 3)	Participants watched the Dive Guide VR experience without pausing.	23	12.41 (0.93)

Note. Each participant joined only one of the three conditions.

Measures

Demographics and VR use. Participants were asked at the pre-test about their demographics (age, country of residence, gender, race, and English proficiency) and previous experience with VR (“How many times have you used VR before?” with answers ranging from “never” to “more than 10 times”).

Learning. Six multiple-choice questions were developed by a marine expert to assess participants’ knowledge on the topic of the experience and the effects of the intervention. The same six questions were asked in the pre-and post-test. Questions were coded as correct (score = 1) or incorrect (score = 0). An average of the six questions’ scores was used to create the learning composite. An example of a question is: “Why are corals important for tropical marine life?”

The following subjective measures had answers on a 5-point Likert scale:

Self-efficacy. To investigate the effects of the intervention on participants’ self-efficacy towards science, we asked five questions adapted from Tuan and colleagues (2005) in pre-and post-test. A composite of the self-efficacy questions was created by averaging the five questions’ scores (pre-test Cronbach’s alpha $\alpha = .79$, post-test $\alpha = .79$). An example of the questions asked is: “I am confident about understanding difficult science concepts on this virtual reality activity.”

Positive and negative affect. Seven items from the Positive and Negative Affect Schedule (PANAS, Watson et al., 1988) were used to assess participants’ moods and emotions before and after the intervention. The question had the prompt “please tell us how you feel right now,” followed by the items “interested,” “distressed,” “attentive,” “ashamed,” “active,” “jittery,” and “overwhelmed.” Positive affect scores were averaged to create a positive composite (pre-test $\alpha = .84$., post-test $\alpha = .81$), while negative ones were averaged to create the negative affect composite (pre-test $\alpha = .69$, post-test $\alpha = .72$).

Presence. Nine questions were adapted from the Presence Scale (Han et al., forthcoming) and asked at the post-test, with three questions investigating each type of presence: social, self, and environmental. A composite of each type of presence was created by averaging the corresponding questions (social $\alpha = .78$; self $\alpha = .79$; environmental $\alpha = .78$). A general presence composite including all nine questions was also created ($\alpha = .88$). A question example is “I felt as if I was inside the virtual world.”

Agency. Five questions adapted from McGivney (forthcoming) were asked to investigate how participants felt agency over the learning activity at the post-test. An example of a question is “The VR activity felt hands-on,” and a composite was created by averaging the five questions ($\alpha = .86$)

Entitativity. Six questions adapted from Han and colleagues (forthcoming) were asked at the post-test to assess how participants felt as part of the group. The average of the questions’ scores was used to create the entitativity composite ($\alpha = .85$). An example of a question is “The members of my group in VR are strongly bonded.”

Cyber-sickness. Four questions were asked at the post-test to measure participants’ physiological discomfort. The question prompt was “please rate the extent to which you are feeling right now, after today’s experience in VR,” and items included “I have a headache,” “I feel nauseated,” “I feel my eyes are strained,” and “I feel aware of my stomach.” A composite was created by averaging the four items’ scores ($\alpha = .73$).

Results

We used linear regression models to estimate the effects of conditions on the variables of interest. Models were controlled for individual characteristics (age, VR experience, gender, English proficiency), implementation day (to control for design iterations), and reported cyber-sickness. Models predicting self-efficacy or learning were controlled for their baseline levels. Post-hoc hypothesis testing was conducted with pairwise comparisons of the least squares means between the conditions, with significance levels adjusted by the Bonferroni method. All the data analyses were carried out in the software R Studio 2022.02.2 (Posit, 2022), build 485. The means and standard deviations of the variables measured per condition at pre- and post-test are shown in Table 2.

Table 2
Means and Standard Deviations at Pre and Post-test

Variables	Viewing (n = 23)				Viewing/Discussing/Graphics (n = 45)				Viewing/Discussing (n = 11)			
	Pre		Post		Pre		Post		Pre		Post	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Learning	0.33	0.2	0.47	0.25	0.36	0.2	0.61	0.22	0.29	0.15	0.56	0.29
Self-efficacy	3.63	0.88	3.78	0.77	3.88	0.66	3.79	0.81	3.8	0.64	3.53	0.78
Positive affect	3.72	0.84	3.88	0.94	3.87	0.89	3.64	0.91	3.84	0.66	3.30	0.97
Negative affect	2.00	0.81	2.08	0.92	1.94	0.64	2.15	0.85	1.95	0.54	1.86	0.67
Agency	-	-	3.82	0.98	-	-	3.82	0.87	-	-	2.78	1.00
Entitativity	-	-	2.6	1.13	-	-	2.91	0.79	-	-	2.92	0.86
Presence	-	-	3.29	0.91	-	-	3.25	0.74	-	-	2.75	0.71
Cyber-sickness	-	-	1.96	0.88	-	-	2.75	0.8	-	-	2.32	1.12

Note. Learning scores ranged from 0 to 1, and all the other variables’ scores ranged from 1 to 5.

Collectively building an environment using graphics in IVR enhanced learning compared to just viewing content (H1). The regressions showed a significant effect of condition on learning ($F(2,66) = 4.54, p = 0.014$, Cohen’s $f^2 = 0.09$, 95% CI [0.00, 0.28]). Pairwise comparisons showed that participants in the viewing/discussion/graphics condition scored significantly higher for learning at post-test than participants in the viewing condition ($t(66) = 2.65, p = 0.027, d = 0.77$, 95% CI [0.17, 1.36]). There were no significant differences between viewing/discussion and viewing conditions ($t(66) = -0.21, p = 0.977, d = -0.09$, 95% CI [-0.96, 0.79]) or between viewing/discussion and viewing/discussion/graphics conditions ($t(66) = -2.08, p = 0.102, d = -0.85$, 95% CI [-1.69, -0.02]). Given the difference in time spent on the learning task between conditions, and the role of time in learning (Bloom, 1974; van Merriënboer & Sweller, 2005), we investigated if time influenced the result found. Adding the time spent on the learning task to the model showed multicollinearity (variation inflation factor > 5). Hence, it was not included. A model predicting learning from the time spent in the learning task and controlling for age, gender, VR use, cyber-sickness, and English proficiency showed no significant effects of time on learning ($F(1,69) = 2.62, p = 0.110$, Cohen’s $f^2 = 0.02$, 95% CI [0.00, 0.15]).

The intervention had no effects on entitativity (H2: $t(67) = 1.27, p = 0.416, d = .36$, 95% CI [-0.21, 0.94]), agency (H3: $t(67) = 0.34, p = 0.938, d = 0.09$, 95% CI [-0.48, 0.67]) or presence (H4: $t(68) = -0.71, p = 0.757, d = -0.19$, 95% CI [-0.74, 0.35]).

The level of interaction between peers in the learning task did not influence self-efficacy (RQ1). The results of the regression model predicting self-efficacy showed no significant effects of the three conditions on how much participants felt able to learn ($F(2,66) = 0.46, p = 0.633$, Cohen’s $f^2 < 0.01$, 95% CI [0.00, 0.08]).

Self-efficacy and entitativity positively correlated with positive affect, while learning negatively correlated with entitativity and negative and positive affect (RQ2). Table 3 shows the correlations between the variables. Analyzing each question included in the affective composites, we found that learning showed negative correlations with feeling active ($r = -.35, p < .01$) and ashamed ($r = -.26, p < .05$). A mediation analysis was run including condition as a predictor, learning as the outcome and entitativity as a mediator, controlling for learning scores at pre-test, age, VR experience, gender, English proficiency, and cyber-sickness. Results showed a partial negative mediation of entitativity between condition and learning ($ACME = -0.01, p = 0.036$; $ADE = 0.19, p = 0.004$; $Total\ effect = 0.18, p = 0.008$).

Table 3
Correlations Between Positive and Negative Affect, Learning, Self-efficacy and Entitativity

Variable	1	2	3	4
1. Learning				
2. Self-efficacy	-.13			
3. Entitativity	-.27*	.25*		
4. Positive affect	-.28*	.41**	.35**	
5. Negative affect	-.28*	-.08	.09	-.06

Note. * indicates $p < .05$. ** indicates $p < .01$.

Discussion

This study investigated how collaboratively building an IVR environment during a learning activity influenced cognitive and affective aspects of learning and group formation. Because of the novelty of the experimental intervention, we followed a design-based research approach in developing the collaborative knowledge-building activity in IVR. Our main finding is that collaboratively building a virtual representation of knowledge in IVR showed positive effects on learning compared to having an IVR experience in groups without collaboration. This finding is aligned with the collaborative cognitive load theory assumption of positive interdependence effects on learning (Janssen & Kirschner, 2020). It indicates that although building a virtual environment in IVR added extraneous load to the participants, the transactive memory resulting from the group interaction overcame the transaction costs of this interaction, resulting in positive effects on learning.

Our data showed that while learning in groups in IVR, feeling part of the group and feeling able to learn positively correlated with each other. These findings are aligned with studies investigating the correlations between self-efficacy and learning through the lens of social cognitive theory (Bandura, 1977) and in computer-supported collaborative learning environments (Gegenfurtner, 2013). Interestingly, feeling part of the group negatively correlated with learning and mediated the relationship between conditions and learning, bringing empirical evidence of collaborative processes mediation between task characteristics and learning proposed by Janssen and Kirschner (2020). Although we do not have data to explain this finding, given that learners were in a sensory-rich, immersive virtual environment that demanded attention and interaction with not only the stimuli but also other members around them, they probably could not allocate as much of their cognitive resources to the learning aspect of the experience, and learning was secondary to the social aspect of the experience. It may also be because most of the participants did not know each other before the study, requiring them to spend more effort to connect with each other and adding to the transaction costs of the collaboration (Janssen & Kirschner, 2020). Future studies should investigate how this relationship occurs in groups where participants already know each other.

We found that positive and negative affect were negatively correlated with learning. Analyzing the correlations between each positive and negative affect in the composites, we found that correlation to be driven by feeling active and ashamed. The negative correlation between action and learning in VR is aligned with the assumption that the cognitive load related to physical activity in the IVR environment could add extraneous load to the learning process and impair learning (Queiroz et al., 2022). Moreover, dealing with the VR equipment could have been difficult for some participants, contributing to feelings of shame and negatively impacting learning (Perkun, 2006). On the other hand, the positive connection between positive affect and self-efficacy supports the theoretical link between self-efficacy and emotion (Bandura, 1977). The connections among affects, learning, and self-efficacy provide nuanced dynamics of learning in IVR contexts.

Running six simultaneous collaborative online-based IVR sessions with dozens of participants required a solid logistic plan, from wi-fi bandwidth to facilitators assignment. Adding customized 3D models to the environment worked for the pilot sessions, but it made the experience crash when dozens of participants were in the experiment. Also, although facilitators trained exhaustively before the study, they needed assistants roaming the rooms to help with system crashes, participants' needs, and hardware malfunction.



Although this study was designed in consultation with experts in education, psychology, communication, computer science, and marine biology to properly investigate the research questions and hypotheses, it has some limitations. Inherent to the design-based research are the changes in the learning activity after rapid iterations. These modifications to the learning environment throughout the study and the time spent in each condition could have impacted the results. Albeit we ran regression models controlling for the day of the data collection to reduce this possible effect, there could be some confounding effects between day and condition and influences on the results. Future studies should investigate the effects of the levels of interactivity in IVR without using a design-based research method to identify if the results we found are replicated.

Conclusion and Future Directions

This study investigated the effects of different levels of interaction in IVR on cognitive and affective aspects of learning. Results showed that participants who collaboratively built an environment while learning about the ocean scored higher for learning than participants in less interactive conditions. Subjective measures assessing affective aspects of learning and group formation showed that feeling active in the environment and bonding with peers decreased learning. These results align with collaborative cognitive learning theory assumptions that interacting with others in collaborative tasks has a transaction cost that impedes learning. In contrast, the possibility of sharing information effectively in a group contributes to increasing learning.

IVR is a promising tool for learning at scale, particularly as it shapes the development of a social learner ecosystem. It is an incredible medium for building things and for experiences that are impossible or too expensive to do in the real world. In our study, groups performed impossible tasks - they built biodiverse reef habitats as though they themselves were corals. Our results provide preliminary data that doing the impossible collaboratively in IVR facilitates learning. As many people will first experience collaborative virtual learning environments in the metaverse, a decentralized network of virtual spaces where users can socialize, learn, and play (Pimentel et al., 2022), more research is needed to better understand the effects of IVR within the context of social interaction and peer learning.

References

- Alhalabi, W. S. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behavior and Information Technology*, 35, 919–925. <https://doi.org/10.1080/0144929X.2016.1212931>.
- Bailenson, J. N. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. (1st ed.). W.W. Norton & Company.
- Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84(2), 191–215.
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The journal of the learning sciences*, 13(1), 1-14.
- Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29(9), 682–688. <https://doi.org/10.1037/h0037632>
- Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, 3, 14-25.
- Carrol, B., & Tambe, P. T. (2022). Advancing Ocean Science Literacy Through Immersive Virtual Reality. Summative Evaluation Report by Inverness Research for the National Science Foundation's Advancing Informal STEM Learning program.
- Clarke, J., & Dede, C. (2005). Making learning meaningful: An exploratory study of using multi-user environments (MUVES) in middle school science. *American Educational Research Association Conference*, Montreal, Canada.
- Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media psychology*, 19(2), 272-309.
- ENGAGE. (2022). [software] <https://engagevr.io/>
- Fauville, G., Queiroz, A. C. M., & Bailenson, J. N. (2020). Virtual reality as a promising tool to promote climate change awareness. *Technology and Health*, p. 91–108.
- Gegenfurtner, A., Veermans, K., & Vauras, M. (2013). Effects of computer support, collaboration, and time lag on performance self-efficacy and transfer of training: A longitudinal meta-analysis. *Educational Research Review*, 8, 75-89.
- Han, E., Miller, M.R., DeVeaux, C., Jun, H., Nowak, K.L., Hancock, J.T., Ram, N., Bailenson, J.N. (forthcoming). People, Places, and Time: A Large-scale, Longitudinal Study of Transformed Avatars and Environmental Context in Group Interaction in the Metaverse. *Journal of Computer-Mediated Communication*, <https://doi.org/10.1093/jcmc/zmac031>



- Han, E., Nowak, K. L., & Bailenson, J. N. (2022). Prerequisites for Learning in Networked Immersive Virtual Reality. *Technology, Mind, and Behavior*, 3(4: Winter). <https://doi.org/10.1037/tmb000094>
- Janssen, J., & Kirschner, P. A. (2020). Applying collaborative cognitive load theory to computer-supported collaborative learning: towards a research agenda. *Educational Technology Research and Development*, 68(2), 783-805.
- Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R, J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning*, 13(2), 213-233.
- Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, 113(4), 719-735. <https://doi.org/10.1037/edu0000473>
- Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, 33, 937-958. <https://doi.org/10.1007/s10648-020-09586-2>
- Markowitz, D. M., & Bailenson, J. (2021). Virtual Reality and Emotion: A 5-Year Systematic Review of Empirical Research (2015-2019). *PsyArXiv*. <https://doi.org/10.31234/osf.io/tpsmr>
- Mayer, R. E., Makransky, G., & Parong, J. (2022). The Promise and Pitfalls of Learning in Immersive Virtual Reality. *International Journal of Human-Computer Interaction*, 0(0), 1–10. <https://doi.org/10.1080/10447318.2022.2108563>
- McGivney, E. (Forthcoming). Interactivity, Presence, and Agency in Virtual Reality Learning Environments 2023 *AERA Annual Meeting*. April 12-16, 2023: Chicago, IL.
- Pekrun, R. The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educ Psychol Rev*, 18, 315–341 (2006). <https://doi.org/10.1007/s10648-006-9029-9>
- Pimentel, D., & Kalyanaraman, S. (2022). The effects of embodying wildlife in virtual reality on conservation behaviors. *Scientific Reports*, 12(1), 1-18.
- Posit. (2022). [computer software] <https://posit.co/>
- Queiroz, A. C. M., Fauville, G., Herrera, F., Leme, M. I. da S., & Bailenson, J. N. (2022). Do Students Learn Better With Immersive Virtual Reality Videos Than Conventional Videos? A Comparison of Media Effects With Middle School Girls. *Technology, Mind, and Behavior*, 3(3). <https://doi.org/10.1037/tmb0000082>
- Rebelo, F., Noriega, P., Duarte, E., & Soares, M. (2012). Using virtual reality to assess user experience. *Human Factors*, 54(6), 964–982.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332.
- Sundar, S. S., Kang, J., & Oprean, D. (2017). Being There in the Midst of the Story: How Immersive Journalism Affects Our Perceptions and Cognitions. *Cyberpsychology, Behavior, and Social Networking*, 20(11), 672–682. <https://doi.org/10.1089/cyber.2017.0271>
- Tuan, H. L., Chin, C. C., & Shieh, S. H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International journal of science education*, 27(6), 639-654. <https://doi.org/10.1080/0950069042000323737>
- Van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*, 17(2), 147-177.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Webster, R. (2015). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, 24, 1319–1333. doi: 10.1080/10494820.2014.994533

Acknowledgments

We thank Jason McGuigan, Nicole Minor, Leticia B. Cavagnaro, the University Innovation Fellowship Program, and The Hasso Plattner Institute of Design (the d.school) for their assistance in this study. This work was supported by the Stanford Accelerator for Learning and by The National Science Foundation (AISL Program) Award # 1907050.