

**Do Students Learn Better with Immersive Virtual Reality Videos than Conventional Videos? A  
Comparison of Media Effects with Middle School Girls**

Anna Carolina M. Queiroz<sup>1,2</sup>, Géraldine Fauville<sup>1</sup>, Fernanda Herrera<sup>1</sup>, Maria Isabel da S. Leme<sup>2</sup>, Jeremy N.  
Bailenson<sup>1</sup>

<sup>1</sup> Department of Communication, Stanford University

<sup>2</sup> Instituto de Psicologia, Universidade de São Paulo

Anna Carolina M Queiroz – ORCID 0000-0003-1032-1628

Géraldine Fauville – ORCID 0000-0001-5462-2591

Maria Isabel da Silva Leme – ORCID 0000-0003-0844-3554

Jeremy Bailenson – ORCID 0000-0003-2813-3297

Correspondence concerning this article should be addressed to Anna Carolina M Queiroz, 450 Jane Stanford Way,  
room 411, Stanford, CA, 94305, United States. Email: acmq@stanford.edu

**Authors Note**

*Keywords:* immersive virtual reality, learning, self-efficacy, education.

*Acknowledgements:* Thanks to Google Labs, Inc for donating the VR headsets used in these studies.

Queiroz, A.C.M., Fauville, G., Herrera, F., Leme, M.I.S., Bailenson, J. N. (in press). 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*. <https://doi.org/10.1037/tmb0000082>

### **Abstract**

This paper presents two studies comparing the effects of educational immersive virtual reality (IVR) versus traditional videos on conceptual knowledge and self-efficacy. Learning was measured through multiple-choice questions assessing conceptual knowledge, and open-ended questions assessing knowledge understanding, knowledge application, and knowledge creation, based on Bloom's Taxonomy of learning objectives. In Experiment 1, 53 eighth grade students from an all-girls school learned about humans' impact on the ocean through either 360-degree videos, using a virtual reality headset, or through traditional videos, using a computer monitor. Measures were taken before and right after treatment. In Experiment 2, 139 sixth-to-eighth grade students from the same school used the same instructional material and equipment from Experiment 1, and measures were taken four times (before treatment, right after two treatment sessions, and five weeks after treatment). Also, we measured learning agency and investigated its mediation role between condition and self-efficacy. The groups did not differ on the multiple-choice scores assessing conceptual knowledge in Experiment 1. In Experiment 2, participants in the IVR group scored higher for knowledge creation than participants in the desktop group, but not knowledge understanding or knowledge application. The IVR group scored higher on self-efficacy than the desktop group in Experiment 1, but not in Experiment 2. Finally, learning agency mediated the relationship between condition and self-efficacy in Experiment 2, indicating a possible mechanism underlying immersion effects on self-efficacy. Results are discussed in light of cognitive sciences and their implication for learning in immersive virtual reality.

*Keywords:* immersive virtual reality, learning, self-efficacy, education.

Since the first virtual reality head-mounted displays (HMDs) were built in the 1960s (Sutherland, 1968), there has been an interest in using immersive virtual reality (IVR) technology for learning (Dede, 2010). Yet over the last decade, IVR's commercial availability has increased considerably (Parong & Mayer, 2018) and is now becoming affordable to the general public. As a result, more attention is now being drawn to possible IVR applications in schools and research on the impact of IVR within the context of learning (Blascovich & Bailenson, 2011; Mado et al., 2022; Parong & Mayer, 2018). IVR is a genre of technology that takes many forms, and there is great variance in how scholars define the medium. In this paper, *IVR videos* are defined as 360-degree videos created from real images captured through 360-degree cameras and watched through a VR headset. *Traditional videos* refer to 2D format videos that are watched through a computer monitor. When in IVR, the headset blocks stimuli other than the digital content and allows the user to visually explore through head movements. In contrast, traditional videos limit the viewer's point of view. Hence, we ran two experiments comparing the effects of watching videos either in IVR or on a computer monitor on conceptual knowledge and *self-efficacy* (i.e., the feeling of being able to learn). We expect participants in IVR to perceive that they have higher control of the learning process (namely, *learning agency*) than participants in the desktop condition. Thus, we hypothesized that participants in the IVR condition will score higher for self-efficacy than participants in the desktop condition (H<sub>1</sub>). We also hypothesized that this relationship between condition and self-efficacy would be mediated by learning agency (H<sub>2</sub>). This study also aimed to answer the following research questions: How do IVR videos, compared to traditional videos, impact conceptual learning? (RQ<sub>1</sub>) and; Compared to traditional videos, how are the long-term effects of IVR videos on conceptual learning (RQ<sub>2</sub>) and self-efficacy (RQ<sub>3</sub>)? In the following sections, we discuss previous findings which support these hypotheses and guide the research questions, and also define the type of technology, content, and learning that past research has focused on.

### **Immersive virtual reality videos and computer-generated graphics content**

IVR can vary depending on the equipment, set up, and digital content used. The two main categories related to content are either computer-generated graphics (CG) or 360-degree videos. In this paper, a 360-degree video displayed in a VR headset will be called *IVR video*, the same term used by Li and colleagues (2017). This term is also used to distinguish it from other 360-degree videos displayed on other screens such as a computer monitor. The main difference between IVR videos and CG-based IVR is that CG content typically has lower photographic realism and allows greater interactivity than IVR videos. In CG environments, users can usually choose avatars,

select objects, and translate their head and hand position in the virtual environment while it is refreshed according to their body movement. On the other hand, IVR videos are created using real images captured with a 360-degree camera, which are then mapped to the inside of a sphere and surround the user seamlessly, bringing higher visual fidelity than CG content. Typically, IVR videos only update as a function of users' head movements, as opposed to hand tracking or room-scale IVR, where people walk around physically. Both categories are immersive but vary in terms of visual fidelity and interactivity.

In academia, VR has traditionally been implemented via computer graphics that enable head translations. However, over the past few years, one can make the argument that using VR headsets to watch 360 videos is common, perhaps as much so as computer graphics. For example, *Youtube 360* is one of the most utilized educational applications by parents (Mado et al., 2022). Moreover, in the corporate training space, 360 videos are the most common use of VR given the need to scale up to millions of users (Bailenson, 2020). Finally, one of the venues where VR reaches the public most readily is at events such as film festivals and museums, and 360 videos are quite common in these contexts (Jun et al., 2020).

Compared to a traditional video (2D) streamed on a computer monitor, the IVR videos are more immersive and interactive. Because IVR videos place the user in the middle of a digital sphere and refresh the digital content in real-time according to the user's head movements, IVR videos allow the user to feel immersed within the scene and as one who controls where to examine the scene over time (Sundar et al., 2017).

### **Immersion and Presence**

When comparing the psychological effects of VR headsets to computer monitors, it is important to highlight the difference between two critical concepts in the IVR research field: *immersion* and *presence*. *Immersion* is “the objective level of sensory fidelity produced by a virtual reality system” (Laha et al., 2013, p. 529). In other words, “the more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more it is immersive” (Slater, 2003, p. 1). The degree of immersion varies based on hardware aspects and the number of senses activated by the technology (Makransky & Petersen, 2021). Two important attributes used to quantify immersion are visual fidelity (e.g., field of view and resolution) and tracking coverage (e.g., head movement tracking) (Slater, 2003).

IVR systems usually rely on HMDs that block visual stimuli around the user, replacing it with a digital visual projection inside the HMD, creating a virtual environment. HMDs are considered fully immersive, as the virtual environment surrounds the user view (Rebelo et al., 2012). The system tracks the user's head movement and updates the virtual environment accordingly, which provides a 360-degree area of viewable digital content. Therefore, HMDs provide higher levels of immersion than computer monitors, along with a wider field of view and user's head movement tracking - two critical features of IVR systems (Slater, 2003). Consequently, the unique affordance of IVR systems of immersing users in virtual environments elicits feelings of *presence*. *Presence* is a subjective experience and "refers to the phenomenon of behaving and feeling as if we are in the virtual world created by computer displays" (Sanchez-Vives & Slater, 2005, p. 332). More specifically, when experiencing presence, "your perceptual, vestibular, proprioceptive, and autonomic nervous systems are activated in a way similar to that of real-life in similar situations" (Slater, 2003, p. 2). Previous research in this area has found that increased immersion increases feelings of presence (Sanchez-Vives & Slater, 2005).

### **Types of Learning**

There are generally three domains of learning: cognitive, affective, and psychomotor (Krathwohl & Anderson, 2009). For the cognitive domain, learning can be categorized as conceptual or procedural (Anderson et al., 2001). *Conceptual learning* connects the most essential elements that have already been organized and explained (Anderson et al., 2001; Krathwohl, 2002). It involves knowledge of classification, categorization, theories, models, structures, principles, and generalizations. This type of knowledge relates to the explanation of what is known and has been learned. *Procedural learning* refers to the use of criteria and methods in solving problems. It encompasses knowledge of specific techniques, skills, and methods, as well as the perception of how and when to use a particular procedure (Anderson et al., 2001; Krathwohl, 2002).

Most of the studies investigating the effects of VR headsets on learning used CG-based IVR (Jensen & Konradsen, 2018; Queiroz et al., 2018). Reviews from those studies suggest that IVR, when compared to other learning methods, positively impacts the affective (e.g., feelings and emotions) and psychomotor (e.g., motor skills) domains of learning. Results from Queiroz and colleagues (2018) indicated that students using IVR reported higher affective outcomes, such as greater confidence and more satisfaction in learning compared to other settings. All studies reviewed by Jensen and Konradsen (2018) investigating learners' attitudes towards IVR reported that

students perceived IVR as useful, exciting, and engaging. This pattern holds for studies examining IVR video as well as CG-based IVR.

Among studies comparing the effects on cognitive learning of CG-based IVR to other media, literature shows varied results. While some studies showed higher learning gains for IVR than other media (Alhalabi, 2016; Webster, 2015), some showed the opposite (Dede et al., 2000; Parong & Mayer, 2018). In addition, some studies showed no differences between conditions (Allcoat et al., 2021; Makransky, Borre-Gude, and Mayer, 2019). Among the studies comparing IVR videos to other media, there are studies that show that IVR was more effective for learning (Walshe & Driver, 2019; Rupp et al., 2019) or report no difference between conditions (Lee et al., 2017; Harrington et al., 2018); and to the best of our knowledge, none of the studies reported more significant learning gains when using other media compared to IVR.

One of the reasons indicated by scholars for lower conceptual learning when using CG-based IVR compared to other media is that the rich and interactive CG environment may cause cognitive overload and exceeds the learner's capabilities to process the conceptual information (Mayer, 2017; Moreno & Mayer, 2002; Parong & Mayer, 2018). Roussou and Slater (2017) compared the impacts of interactive versus passive CG-based IVR in evoking and sustaining conceptual change and found that passive IVR favored sustained conceptual change. These findings point out the need for further research about IVR's role for learning, particularly when considering different levels of interactions (e.g., IVR videos versus CG-based IVR) and its use at scale.

In summary, research comparing IVR videos to other media shows that IVR videos were more effective for learning than other media (Walshe & Driver, 2019; Rupp et al., 2019) or shows no difference between conditions (Lee et al., 2017; Harrington et al., 2018). Because there is a lot to learn about how immersion influences learning, our first research question (RQ) is how watching IVR videos on a headset or computer monitor will impact conceptual learning (RQ<sub>1</sub>).

**Cognitive objectives of learning.** Bloom (1956) proposed a hierarchical taxonomy emphasizing the cognitive objectives of learning. Mayer and colleagues (2001) reviewed Bloom's taxonomy and proposed six levels of cognitive dimensions, in crescent order: *remember*, *understand*, *apply*, *analyze*, *evaluate* and *create*. These levels are described as follows: *remember* refers to information recall, while *understand* refers to comprehension and the ability to translate what was learned to students' own words; *apply* relates to the ability to use the knowledge in new

situations that were not taught directly; *analyze* relates to the ability to break down information and discuss its elements; *evaluate* refers to the ability to make judgments about something, using some criteria; and, *create* constitutes the highest cognitive level and refers to inductive thinking and the ability to create based on what the student already knows. This taxonomy is particularly useful when planning and evaluating learning activities, helping to assess lower and higher cognitive processes involved in IVR. According to Jensen and Konradsen's (2018) review on IVR and learning, no study has investigated the effects of IVR on each higher level proposed by Bloom's taxonomy so far (e.g., apply and create), indicating the need for this kind of investigation and the novelty of the present study.

### **Self-efficacy**

Self-efficacy, a concept coined by Bandura (1977) in light of the Social Learning Theory, is defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391). In other words, self-efficacy refers to people's beliefs that they can succeed in a specific task, topic, or environment.

Perceived self-efficacy determines how much effort they will put into a particular activity and how long they will persist in succeeding at that activity (Bandura, 1977). Usher and Pajares (2008) suggested that people who perceive themselves as being able to learn and master some activity will face challenges more positively and be more persistent at mastering that activity than people who perceive themselves as incapable of attempting or mastering the same activity.

Self-efficacy has been considered a predictor of students' academic performance and career choices (Maddux, 2016; Zimmerman, 1995). Also, it has been shown to affect students' career choices by influencing the type of environments and activities they engage with because people tend to avoid environments they believe they will not be able to cope with. Thus, several research studies have focused on understanding self-efficacy in educational settings (Graham et al., 2005; Schunk, 1995; Schunk & DiBenedetto, 2016).

According to Bandura (1997), the four main sources of students' self-efficacy are 1) enactive mastery experiences, 2) vicarious experiences (observation), 3) social persuasions, and 4) physiological states. The author states that enactive mastery experiences (i.e., performance accomplishments) produce the highest and most generalized increase in perceived self-efficacy, as they provide genuine and immediate evidence that students can

succeed on the task at hand (Bandura, 1997). After receiving that evidence and interpreting the results, students tend to develop beliefs about their performance, impacting students' self-efficacy. Although middle and high school girls tend to have higher grades in STEM subjects (Science, Technology, Engineering and Math; Voyer & Voyer, 2014), stereotypes and self-confidence may prevent them from pursuing STEM careers (O'Dea et al., 2018). In this context, Gibbons and Borders (2010) suggest that students make their career decisions between 6th and 8th grade, and intervention programs targeting girls' STEM self-efficacy have shown to be more effective before 8th grade (Cho et al., 2009).

Many studies highlighted the role of teachers, parents, and schoolmates on students' self-efficacy (Schunk, 1995; Usher & Pajares, 2008). However, because of the recent increased technological adoption by students both inside and outside of schools, research has started to focus on the bidirectional relationship between technology and self-efficacy (Huang & Mayer, 2019; Kuo et al., 2014; Meyer et al., 2019; Makransky et al., 2020b; Petersen et al., 2020; Sun & Rueda, 2011). More specifically, given IVR's potential to tap into students' affective dimensions of learning, more attention has been drawn towards understanding the effects of IVR on self-efficacy (Bandura 1986; Plass & Kaplan, 2016; Pekrun, 2006; Zimmerman, 2000).

Findings from studies investigating the effects of CG-based IVR and IVR videos on self-efficacy indicate a positive effect (Makransky, Borre-Gude & Mayer, 2019; Makransky et al., 2020a, 2020b; Meyer et al., 2019). Makransky, Borre-Gude, and Mayer (2019) compared the motivational and cognitive effects of an IVR simulation, a desktop simulation (both using CG environments), and a conventional text-based safety manual for laboratory safety training. Although results indicated no significant difference on immediate retention tests between conditions, they showed a significant difference in motivation, enjoyment, and self-efficacy between IVR and the text condition. Participants in the IVR condition reported higher scores on these measures.

Francis and colleagues (2020) compared the effects of IVR videos to traditional lectures in self-efficacy among preclinical physician assistant students when learning about the operating room. Their results showed that participants in both groups reported higher self-efficacy after treatment than at pre-test. Still, participants in the IVR video condition reported significantly higher self-efficacy than participants that attended the lecture (Francis et al., 2020).

Although studies comparing VR to computer monitors on learning show varied results, ranging from positive to negative effects of VR on learning, results of studies targeting self-efficacy usually converge, indicating a positive effect of immersion. Considering these previous studies, our first hypothesis is that participants in the IVR condition will score higher for self-efficacy than participants in the desktop condition immediately after treatment ( $H_1$ ).

### **Sense of Agency**

Although the literature shows consistent results with positive effects of immersion on self-efficacy, the mechanisms underlying these effects are still unclear. We hypothesize that this effect is due to the *sense of agency* (the feeling of making things happen; Tapal et al., 2017) provided by the headset use. When watching IVR video, the user is virtually placed in the center of a sphere surrounded by digital content. The users can visually explore the virtual content moving their heads, and the content is refreshed accordingly. Movement and abstract actions are preceded by the intention to act (David et al., 2008), which is based on people's impression of the world and how the world responds to their actions (Triberti & Riva, 2016). Moreover, the brain is constantly analyzing body and environmental information to create predictive models about events and the actions' impact on these events. These predictions can be confirmed or not, since the greater the congruence between the action and the predicted consequence, the greater the sense of agency becomes (David et al., 2008).

This shows that watching IVR videos is different from watching a video on a computer monitor. People watching IVR videos can engage in motor actions of exploration and perceive themselves as able to cause changes to the environment, while people using a computer monitor can explore the video images on the camera angle shown in the monitor. When focusing on learning, the more someone feels that they can make things happen, control their learning, and experience positive emotions about their ability to learn, the greater their feeling of self-efficacy (Bandura, 1997).

A qualitative study comprising two case studies investigating VR use for high-schoolers science learning reported that students considered agency an essential aspect of the learning experience with this technology (McGivney, 2021). In one of the case studies, participants watched IVR videos at home using cardboard VR viewers. Data from the interviews and focus groups indicated that participants felt engaged because they could

explore and control where to look in the virtual environment. We speculate that watching IVR videos positively affects self-efficacy as it allows action on the environment.

Since self-efficacy has been found to affect academic performance and influence professional choices, Multon, Brown, and Lent (1991) argued that finding new ways to support self-efficacy for educational, professional, and personal growth is crucial. Thus, understanding how teaching methods and in-classroom tools impact self-efficacy plays an important role in improving students' performance. Focusing on IVR in particular, given its recent and increasing adoption in formal and informal learning environments, it is necessary to understand its role in fostering self-efficacy. Moreover, it is important to note that most studies assessing the relationship between IVR and self-efficacy have only used CG virtual environments. There is still a lot to learn about the relationship between IVR videos and self-efficacy.

Finally, although studies have reported positive effects of immersion on self-efficacy, the mechanism underlying this effect is still not clear. Makransky and Petersen (2021) proposed a theoretical framework describing the process of learning in IVR. In their framework, control factors, such as degree, immediacy, and mode of control are the most important predictors of agency. In turn, a sense of agency seems to influence users' engagement and sense of autonomy in VR experiences (McGivney, 2021). Thus, we predict that learning agency (i.e., the feeling that one can control how and what to learn) will mediate the relationship between the conditions and self-efficacy immediately after treatment ( $H_2$ ). In other words, the increased immersion and visual exploration enabled by the VR headset will positively influence the participants' feeling of being able to control what they learn (learning agency), which will in turn have positive effects on how they feel that they are able to learn (self-efficacy).

### **Long-term effects of IVR for learning**

Before implementing IVR in formal education at scale, it is essential to understand its long-term effects. In a study by Stepan et al. (2017), participants learned about the ventricular system and brain using IVR or online textbooks. Learning and motivation measures were taken before, right after, and eight weeks after the treatment sessions, and there was no significant difference in the learning assessment between conditions over time. To the best of our knowledge, this is the first study to measure the long-term effects of IVR videos on learning, and only a few have investigated the long-term effects of CG-based IVR. This lack of evidence limits the predictions about IVR videos' long-term effects on learning, and open avenues to exploratory investigations. Experiment 2

investigates IVR and traditional videos' long-term effects on conceptual learning and self-efficacy through the following research questions: Compared to traditional videos, how are the long-term effects of IVR videos on conceptual learning (RQ<sub>2</sub>) and self-efficacy (RQ<sub>3</sub>)?

We conducted two studies to compare the effects of educational IVR and traditional videos on conceptual knowledge, measured through multiple-choice questions and open-ended questions (the open-ended questions were built following the Bloom's Taxonomy of learning objectives and assessed knowledge understanding, knowledge application, and knowledge creation), and self-efficacy. In Experiment 1, 53 eighth grade students from an all-girls school learned about humans' impact on the ocean through either IVR videos, using a virtual reality headset, or through traditional videos, using a computer monitor. In Experiment 2, 139 sixth-to-eighth grade students from the same school used the same instructional material and equipment from Experiment 1, and measures were taken four times (before treatment, right after two treatment sessions, and five weeks after treatment).

## **Experiment 1**

### **Method**

The Stanford University Institutional Review Board approved procedures and materials in this study. Participation was voluntary and written parental consent forms and assent forms were obtained from all participants and their parents/legal guardians. Researchers worked with the participants' science teacher to define the logistics of the experiment and the wording of the questionnaires used. Power analysis with a small effect size (Cohen's  $f = 0.20$ ) and a significance level of 0.05, powered at 0.80 (Cohen, 1992), suggested a total of 52 participants ( $N = 52$ ) would be required. The entirety of the study occurred before participants covered any learning material in their coursework related to topics investigated in this study – namely, ocean acidification and coral reefs.

### **Participants**

The study was run in an all-girls school in the United States, given an existing relationship between the researchers and the science teachers, which was critical given the logistics of multiple sessions over time. Also, we chose an all-girls school because we were specifically interested in boosting self-efficacy in science learning and working with girls was important both theoretically (Cho et al, 2009) and for improving society (Gibbons & Borders, 2010). An initial sample of 55 female 8th-grade students answered the pre-test using a MacBook Air

computer, and two students failed to complete some part of the study and were excluded from analyses. The final sample consisted of 53 female participants.

### **Materials and Apparatus**

The instructional material consisted of two 6-minute long videos: *The Crystal Reef* and *Coral Compass*, developed by the Virtual Human Interaction Lab at Stanford University. We used immersive 360-degrees video in the IVR condition to keep the experimental and control condition as similar as possible. To investigate the influence of immersion on the learning outcomes, we decided that IVR videos would allow the two conditions to be similar in terms of time, pace, and content, while varying the level of immersion. The development of each video took years and was the result of interdisciplinary collaborations with marine scientists as domain experts and learning scientists. Moreover, compared to many 360-degree videos, these were designed to minimize distraction (i.e., avoiding voiceover and textual cues during important visual events) and motion sickness (i.e., no movement of the camera to disturb the vestibular system). Both videos were Official Selections at the Tribeca Film Festival, and each has been used in hundreds of classrooms, museums, and other informal learning institutions across the planet. They are available in the following links: Coral Compass (<https://stanfordivr.com/coralcompass/>); The Crystal Reef (<https://stanfordivr.com/the-crystal-reef/>).

Both videos depict how human actions have been negatively affecting the ocean. *The Crystal Reef* focuses on ocean acidification and how carbon dioxide (CO<sub>2</sub>) from human emissions negatively impacts the ocean. The video presents a female scientist diving into the Mediterranean Sea, in a site where natural underwater vents spew CO<sub>2</sub> into the water. The high concentration of CO<sub>2</sub> in the water decreases the water pH, resulting in more acidic water and reduced biodiversity. *Coral Compass* focuses on how human activities have impacted the coral reefs in Palau, a small island in the Western Pacific. A male scientist narrates the video. It presents how tourism and land practices have affected the coral reefs' health and shows how the country's government has been acting to reduce the negative impact on the ocean. Screenshots of these videos can be seen in Figures 1 and 2.

## Figure 1

### *Coral Compass video images*



(a)

(b)

(c)

**Fig. 1** (a) Tourists diving in the coral reef in Palau. (b) Researchers and Palauan policymakers discussing conservation measures. (c) Coral reefs affected by sediments from bad land practices

## Figure 2

### *The Crystal Reef video images*



(a)

(b)

(c)

*Note.* (a) Healthy coral reef. (b) Volcanic vents releasing CO<sub>2</sub> into the water. (c) Unhealthy coral reef with reduced biodiversity

Estimated field of view, resolution, visual exploration and occlusion of the surrounding environment are presented in Table 1. The HMD used in this study is considered more immersive than the computer used as the HMD has a wider field of view (110 degrees, against approximately 30 degrees on a computer monitor) and a larger area of visual exploration (360 degrees area of exploration, against 13.3” on the computer monitor; Cummings &

Bailenson, 2016). Also, the HMD offers complete occlusion of the physical environment and tracks participants' head movements, placing them in the center of a virtual sphere that projects the video in 360-degree, contributing to increased immersion. Depending on which condition participants were randomly assigned to, they either watched both videos on the screen of a 13.3" Mac-Book Air (desktop condition) or on a Lenovo Mirage Solo VR headset (IVR condition).

Participants in both conditions listened to the audio using headphones and were exposed to the same content (i.e., words and images). The desktop versions of each IVR video were created using a screen capture software (Fraps™) from a person going through the IVR experience. The length and sequence of the experience were the same for both conditions. While participants in the IVR condition could turn their heads and have the digital content refreshed accordingly, participants in the desktop condition could only see what was on the computer screen, and no movement on the camera angle was possible. We chose this condition over the interactive desktop video condition to simulate what people typically do in classrooms as a control condition.

**Table 1**

*Devices features*

Feature	HMD (Lenovo Mirage Solo)	Computer monitor (Mac-Book Air)	Relevant differences influencing immersion
Field of view	110 degrees	~30 degrees	A wider field of view increases immersion (Cummings & Bailenson, 2016)
Resolution	2560 x 1440	2560 x 1600	-
Area for visual exploration (Field of Regard)	360 degrees	13" screen	The 360 degrees area of exploration refreshed according to HMD users' head movements allows them to act on the digital content (Sundar et al., 2017)
Occlusion of Physical Classroom	Complete	11.97" x 8.36"	The physical environment occlusion HMD provides increases the immersion and the feeling of being part of the scene (Kim, 2017)

**Design and Procedure**

Participants were invited by their science teacher to participate in this study. Participants who accepted to participate and received consent from their parents answered an online questionnaire investigating their knowledge,

self-efficacy, concerns and opinions about ocean acidification, during their science class. Three weeks later, the researchers visited the school to conduct the treatment sessions, during which participants were randomly assigned to the IVR condition ( $n = 28$ ) or the desktop condition ( $n = 25$ ).

In the IVR condition, participants watched the IVR video versions of *The Coral Reef* and *Coral Compass* using a Lenovo Mirage VR headset and headphones. In contrast, participants in the desktop condition watched the traditional versions of the videos on a MacBook Air laptop and headphones in a separate room. The treatment session and post-test took place during a 90-minute science class period. To account for order effects, half of the participants in each condition watched *The Crystal Reef* first while the other half started with *Coral Compass*.

To mitigate fatigue, after participants watched the first video, the researcher asked them to remove the headset and headphones (IVR condition) or headphones only (desktop condition) and answer three questions assessing presence, three open-ended questions, and three multiple-choice questions assessing learning on a laptop computer. When participants finished that survey, the researchers helped participants in the IVR condition put the headset on again and started playing the second video. After the second video, participants again answered the three questions assessing presence, three open-ended questions, three multiple-choice questions assessing learning, three additional questions about the environmental concern, knowledge about ocean acidification, and twelve questions to measure self-efficacy. In this paper, we will present and discuss the multiple-choice and open-ended learning assessment measures and presence and self-efficacy measures to focus on the effects of IVR videos on learning and self-efficacy.

Finally, after all students completed the experiment, they were debriefed about the study, and researchers answered students' questions.

## Measures

**Presence.** In order to confirm the difference in immersion between the two conditions, presence was used as a manipulation check. Three items assessing feelings of presence using a 5-point Likert scale ( $1 = \textit{not at all}$ ,  $5 = \textit{extremely}$ ) were adapted from Nowak and Biocca (2003). These items were: "To what extent did you feel like you were inside the virtual experience?"; "To what extent did you feel immersed in the virtual experience?" and "How much did it feel as if you visited another place?" Participants answered these three questions during the post-

questionnaire immediately after watching each of the videos. Reliability for the presence questionnaire was  $\omega = 0.96$ .

**Open-ended learning assessment questions.** Six open-ended questions were created to assess conceptual learning based on the critical thinking theory and revision of Bloom's taxonomy. They focused on knowledge understanding, knowledge application, and knowledge creation (Krathwohl & Anderson, 2009; Mayer et al., 2001). These questions are intended to evaluate how immersion would affect learning according to each cognitive level. The questions assessing understanding measured how they remembered the information they received in the videos. The questions focusing on knowledge application required them to apply the information they received in new contexts or purposes, and the questions assessing knowledge creation prompted participants to propose solutions for environmental problems. For example, one of the questions assessing knowledge creation was: "Propose strategies to increase engagement of the general public with ocean acidification." These questions can be found in Appendix A.

Two researchers, blind to condition, used a rubric developed according to Saxton, Belanger, and Becker (2012) to score them in a range between 0 and 5 points. The average rate of agreement between researchers was 89.27% in pre-test and 88% in post-test. When researchers had an initial disagreement on the scoring, they looked for a consensus score, and a score that both agreed upon was used. The six questions were grouped according to their corresponding Bloom Taxonomy level - knowledge understanding, knowledge application, and knowledge creation - with two questions compounding each composite.

**Multiple-choice learning assessment.** Four questions about ocean acidification and coral reefs were adapted from the International Ocean Literacy Survey (Fauville et al., 2019), and a marine biologist expert created two questions about Palauan government actions regarding coral reefs. These questions aimed to evaluate participants' conceptual knowledge about the topic. They are not aligned with Bloom's taxonomy of learning objectives. Participants were given a point for every correct answer and received zero points for incorrect answers. Participants' final score was the sum of all of their correct answers. For example, one of the questions was: "Select the land practice that can greatly help protect the coral reefs in Palau," and the participant had to select the correct answer among four options.

**Self-efficacy and motivation to learn science.** Twelve questions from Tuan et al. (2005), Pekrun et al. (2005 and 2011) were adapted, using a 5-point Likert scale (1= *strongly disagree*, 5 = *strongly agree*) to investigate the effects of immersion on self-efficacy. For example, one of the questions was “Whether the science content is difficult or easy, I am sure that I can understand it.” Negatively framed questions were reverse coded. Reliability for self-efficacy and motivation to learn was  $\omega_{pre} = 0.79$  and  $\omega_{post} = 0.85$ . The model fit was marginal (RMSEA<sup>1</sup> = 0.10, 90% Confidence Interval (CI) [0.06, 0.15], RMSR<sup>2</sup> = 0.13,  $X^2(53) = 85.33$ ,  $p < 0.01$ ). The exact wording of the questions in the self-efficacy questionnaire can be found in Appendix A.

The data and analytic methods are publicly available at <https://osf.io/mpk2j/> (Queiroz et al., 2021).

### Results and Discussion

All the analyses were carried out using RStudio Version 1.1.463, except for power analyses, for which we used the G\*power application. Analysis of covariance (ANCOVA) models predicting each dependent variable of interest with condition as predictor were run. The models were controlled for the pre-test scores of the learning (RQ<sub>1</sub>) and self-efficacy (H<sub>1</sub> and RQ<sub>1</sub>) measures, to address the significant difference found between conditions at pre-test for these variables (multiple-choice learning assessment:  $t(50.69) = 3.16$ ,  $p = 0.003$ ,  $d = 0.86$ , 95% CI [0.29, 1.42]; self-efficacy:  $t(44.49) = 2.27$ ,  $p = 0.028$ ,  $d = 0.63$ , 95% CI [0.08, 1.19]). Means and SDs for each variable are shown in Table 2.

For the models investigating learning (RQ<sub>1</sub>), self-efficacy pre-test scores were included in the model because previous studies indicate that self-efficacy scores have been shown to impact future task performance (Bandura, 1986). For the model investigating self-efficacy (H<sub>1</sub>), given that self-efficacy shows to be impacted both from previous self-efficacy beliefs and current task performance (Bandura, 1986; Bernacki et al., 2015), the learning scores at post-test were included in the model. Finally, testing for multicollinearity with variance inflation factors was run for each model. While there is no consensus on the variance inflation factors value that indicates a concern about multicollinearity (Vatcheva et al., 2016), values below 5 usually suggest the absence of multicollinearity. Predictors from all models testing H<sub>1</sub> and investigating RQ<sub>1</sub> showed variance inflation factors values lower than 2,

---

<sup>1</sup> RMSEA = root mean square error of approximation

<sup>2</sup> RMSR = root mean square of residuals

suggesting no multicollinearity. Tables with the summary of the models we ran are available in the Supplemental Material.

**Table 2**

*Means and Standard Deviations at Pre-test and Post-test*

Measures	Desktop (n = 25)				IVR (n = 28)			
	Pre-test		Post-test		Pre-test		Post-test	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Multiple-choice	0.37	0.17	0.75	0.21	0.54	0.20	0.87	0.15
Understanding	0.96	0.58	2.46	1.03	0.88	0.44	2.57	0.74
Application	0.80	0.74	2.12	0.94	0.86	0.56	2.20	0.81
Creation	1.54	0.64	3.04	0.98	1.45	0.44	3.10	0.83
Self-efficacy	3.20	0.57	3.12	0.71	3.52	0.43	3.65	0.50
Presence	-	-	2.21	0.79	-	-	4.05	0.58

*Note.* *M* = means; *SD* = standard deviations

A significant difference between conditions for presence was found at post-test ( $t(43.55) = -9.59, p < 0.001, d = -2.69, 95\% \text{ CI} [-3.43, -1.93]$ ), with participants in the IVR condition scoring higher than participants in the desktop condition after treatment. This indicates that the experimental manipulation was successful.

After controlling for the self-efficacy pre-test differences, a medium significant effect of the condition was found on self-efficacy post-test scores ( $F(1,46) = 6.62, p = 0.013, \eta^2_p = 0.13, 90\% \text{ CI}^3 [0.02, 0.28]$ ), with participants in the IVR condition reporting higher self-efficacy than in the desktop condition after treatment. Post Hoc power analysis indicated that this test achieved 0.79 power. These results support  $H_1$ .

One possible explanation for this finding is that feelings of greater control of the situation can contribute to the higher self-efficacy perception in IVR (Glenberg, 2018; Hite et al., 2019). Also, a sense of agency and the intention to act have been shown to involve causal efficacy (David et al., 2008; Schlosser, 2012) and may have

<sup>3</sup> For  $\eta^2_p$ , 90% CI is reported instead of 95% in accordance to Lakens (2013)

contributed to increased self-efficacy in the IVR condition. While we have no formal way to test this mechanism with the current data, we hypothesize that acting on the environment while using a headset to watch the IVR video could have increased participants' predictions about their actions and the environment responses. As these predictions were confirmed throughout the experience, they increased the sense of agency, which was positively associated with self-efficacy. However, more studies are needed to investigate this association between a sense of agency and self-efficacy when using IVR.

In addition, this positive impact of IVR on self-efficacy may also be due to the positive effects that IVR has shown in affective aspects of learning such as engagement and enjoyment (Bailenson et al., 2008; Makransky & Lilleholt, 2018; Parong & Mayer, 2018). These aspects are known to influence self-efficacy positively (Bandura, 1995) and therefore may mediate the positive impact of IVR on self-efficacy. However, this study has not focused on these variables, and future studies should consider their relationship with self-efficacy and learning in IVR video contexts.

Given that the students have never experienced IVR videos in the classroom, the novelty effect of IVR may be positively associated with self-efficacy. In addition, more attention was given to the students in the IVR condition. The researchers spent a few minutes before each session explaining to participants how to use the VR headset and help them put the headset on. This procedure could have impacted their feelings of appreciation. As physiological states are one of the sources of self-efficacy (Bandura, 1997), those factors may have positively influenced students' physiological states, which could have been associated with their self-efficacy perception. These possibilities should be considered in future research.

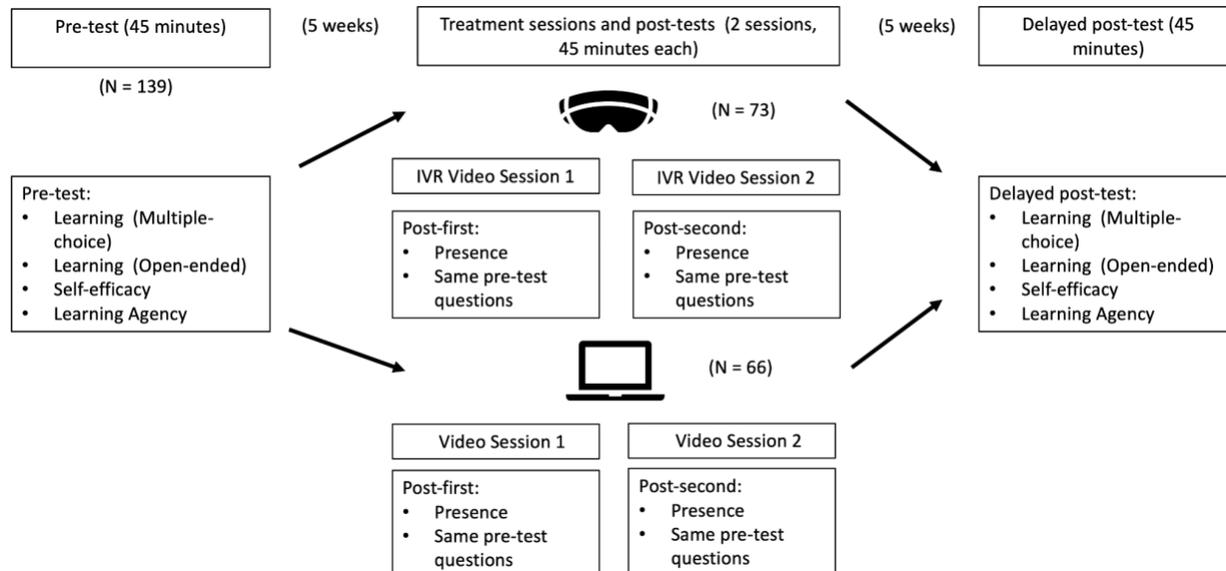
The models investigating the effects of treatment on learning (RQ<sub>1</sub>) showed no significant effect of condition on any variable assessing learning (multiple-choice assessment:  $F(1,49) = 0.03, p = 0.855, \eta^2_p < 0.01, 90\% \text{ CI } [0, 0.04]$ ); knowledge understanding:  $F(1,49) = 0.06, p = 0.814, \eta^2_p = 0.001, 90\% \text{ CI } [0, 0.05]$ ; knowledge application:  $F(1,49) = 0.01, p = 0.915, \eta^2_p < 0.001, 90\% \text{ CI } [0, 0.02]$  or knowledge creation:  $F(1,49) = 0.07, p = 0.797, \eta^2_p = 0.001, 90\% \text{ CI } [0, 0.05]$ ). This finding suggests that more immersion did not lead to better learning in this experiment and corroborates previous studies that compared conceptual learning outcomes between IVR videos and desktops and found no significant difference in learning between these media (Harrington et al., 2018).

## Experiment 2

In the second experiment, the sample was enlarged, with 139 participants from 6<sup>th</sup> to 8<sup>th</sup> grades in the same school in Experiment 1. A matched random assignment (Cozby, 2009) was used to avoid significant differences between conditions at pre-test. Questionnaires measuring learning agency were added to this study to investigate possible mechanisms involved in the immersion effects on self-efficacy. The self-efficacy questionnaire used the complete version from Tuan et al. (2005). Also, the treatment sessions were increased to two, in which participants watched one video per session, and measures were taken after each session. Finally, participants took a delayed post-test, five weeks after the treatment sessions, to evaluate the long-term effects of the intervention. Consequently, in this experiment, measures were taken at four different times: pre-test, post-first (five weeks after pre-test), post-second (one or two days after post-first), and delayed post-test (five weeks after post-second). The Experiment 2 procedure is shown in Figure 3.

**Figure 3**

### Experiment 2 Procedure



## Method

### Participants

Participants were recruited from the same middle school as in Experiment 1. Participation was voluntary. An initial sample of 164 female students from 6<sup>th</sup> to 8<sup>th</sup> grade ( $M = 7$ ,  $SD = 0.83$ ), aged from 11 to 14 years old ( $M$

= 12.27,  $SD = 0.91$ ). Twenty-five students failed to complete some part of the study and were excluded from the analysis. The final sample consisted of 139 participants.

### Materials and Apparatus

The instructional materials used were the same in Experiment 1, the videos *The Crystal Reef* and *Coral Compass*. Participants in the IVR condition used the same VR headsets as in Experiment 1. Participants in desktop condition watched the same traditional versions of the videos and used the same computer model from Experiment 1.

### Design and Procedure

Similar to Experiment 1, all participants answered a pre-test questionnaire on a laptop computer during their science class. This questionnaire had the same questions used in Experiment 1 to assess learning (multiple-choice and open-ended). The questionnaire also included 18 questions assessing self-efficacy (from Tuan et al., 2005) and five questions assessing learning agency (developed by the authors). We used the complete version of the self-efficacy questionnaire from Tuan et al. (2005) in Experiment 2, aiming to increase the questionnaire reliability and model fit. Reliability was good ( $\omega_{\text{pre}} = 0.85$ ,  $\omega_{\text{post-first}} = 0.87$ ,  $\omega_{\text{post-second}} = 0.90$ ,  $\omega_{\text{delayed}} = 0.89$ ), as well as the model fit (RMSEA = 0.06, 90% CI [0.036, 0.077], RMSR = 0.05,  $X^2(139) = 148.66$ ,  $p < 0.01$ )

Five weeks later, the researchers visited the school to conduct the treatment sessions. Participants were assigned to either IVR ( $n = 73$ ) or desktop condition ( $n = 66$ ), after completing the pre-test, through random matched assignment (Cozby, 2009).

Each treatment session took place during a 45-minute science class. Half of the participants in each condition watched the video *The Crystal Reef* in the first treatment session, while the other half watched *Coral Compass*. Researchers helped participants in the IVR condition put the headsets on, start the videos and remove the headsets after participants had finished watching the video. The two treatment sessions were one to two days apart and followed the same procedure, in which participants watched one of the videos in the first session (post-first) and the other video in the second session (post-second). After watching each video, participants used a computer to answer the same questions used in Experiment 1, assessing presence and learning (multiple-choice and open-ended), and also answered questions assessing self-efficacy (from Tuan, Chin & Shieh, 2005) and learning agency.

Five weeks after the treatment sessions, all students answered the delayed post-test questionnaire during a 45-minute science class. The same questions used in the pre-test were used at the delayed post-test. After participants completed this questionnaire, they were debriefed about the study, and a researcher answered students' questions.

## Measures

As mentioned before, items assessing presence (manipulation check;  $\omega_{\text{post-first}} = 0.92$ ,  $\omega_{\text{post-second}} = 0.94$ ) and learning (multiple-choice and open-ended) were the same as Experiment 1.

**Learning agency.** To investigate participants' subjective perception of agency related to learning and visual exploration of the digital content, questions were built on theories of agency, efficacy, and embodiment (Baker et al., 2003; David et al., 2008; Gonzalez-Franco, & Peck, 2018; Johnson-Glenberg, 2018; Judge et al., 2002; Schlosser, 2012; Spengler et al., 2009). Five questions using a 5-point Likert scale were created ( $\omega_{\text{pre}} = 0.43$ ,  $\omega_{\text{post-first}} = 0.60$ ,  $\omega_{\text{post-second}} = 0.61$  and  $\omega_{\text{delayed}} = 0.63$ ). For example, one of the questions was: "I learn more when I am the one selecting what I am going to learn" (1 = *strongly disagree*, 5 = *strongly agree*) (for the complete questionnaire, see Appendix B).

**Self-efficacy and motivation to learn science.** In order to investigate the effects of conditions on self-efficacy, 18 questions from Tuan et al. (2005) were adapted to assess participants self-efficacy and motivation to learn ( $\omega_{\text{pre}} = 0.85$ ,  $\omega_{\text{post-first}} = 0.87$ ,  $\omega_{\text{post-second}} = 0.90$  and  $\omega_{\text{delayed}} = 0.89$ ) using a 5-point Likert scale. Negatively framed questions were reverse coded. The exact wording of the questions in the self-efficacy questionnaire can be found in Appendix B.

## Results and Discussion

Welch Two-Sample t-tests were run to compare the pre-test scores between conditions and the presence scores between conditions at post-first and post-second. ANCOVA models were run to test  $H_1$  and to investigate the research questions. Similar to Experiment 1, the models predicting treatment effects on self-efficacy ( $H_1$  and  $RQ_3$ ) included the condition as the predictor. They controlled for the previous self-efficacy score, the learning scores at the time being analyzed, and which grade participants were attending. These models also included learning agency at the time being analyzed. The models predicting learning after treatment ( $RQ_1$ ) included condition as the predictor, controlled for learning scores at pre-test, previous self-efficacy scores, and grade. The models investigating long-

term effects of treatment on learning (RQ<sub>3</sub>) included condition as the predictor. They controlled for grade and the learning scores and self-efficacy at the second treatment session. Testing for multicollinearity with variance inflation factor was run for each model. In all models, all variables showed variance inflation factor values below 2, indicating the absence of multicollinearity. Tables with the summaries of the models we ran are available in the Supplemental Material. Descriptive statistics for each dependent variable at each time measured are shown in Table 3.

**Table 3**

*Means and Standard Deviations at pre-test, post-first and post-second tests.*

Variable	Pre-test		Post-first		Post-second		Delayed	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	Desktop (n = 66)							
Multiple-choice	0.40	0.20	0.70	0.30	0.78	0.32	0.64	0.24
Understanding	0.93	0.64	2.07	1.24	2.32	0.99	1.80	0.87
Application	0.71	0.61	1.94	1.76	1.92	0.98	1.61	0.87
Creation	1.61	0.80	2.02	0.83	1.86	0.88	1.75	0.65
Self-efficacy	3.83	0.42	3.75	0.49	3.73	0.53	3.73	0.50
Learning Agency	3.56	0.50	3.33	0.41	3.32	0.44	3.45	0.56
Presence	-	-	2.20	0.90	2.17	0.92	-	-
	IVR (n = 73)							
Multiple-choice	0.41	0.20	0.73	0.30	0.77	0.29	0.63	0.22
Understanding	0.93	0.52	2.03	0.99	2.34	0.97	1.62	0.93
Application	0.91	0.70	2.08	1.14	2.18	1.00	1.75	0.86
Creation	1.55	0.80	2.27	0.93	2.26	0.93	1.84	0.67
Self-efficacy	3.92	0.46	3.92	0.44	3.90	0.47	3.81	0.52
Learning Agency	3.55	0.49	3.46	0.43	3.42	0.38	3.58	0.54

Presence	-	-	3.80	0.80	4.02	0.63	-	-
----------	---	---	------	------	------	------	---	---

Note. *M* = mean; *SD* = standard deviation

No significant differences between conditions were found at pre-test for the variables measured, as shown in Table 4. Also, there were no significant differences between conditions for age ( $t(134.83) = 0.301$ ;  $p = 0.764$ ,  $d = 0.05$ , 95% CI [-0.28, 0.38]) and grade they were in ( $t(134.66) = 0.203$ ;  $p = 0.840$ ;  $d = 0.03$ , 95% CI [-0.30, 0.37]).

**Table 4**

*Means and Standard Deviations at pre-test, t-test results between conditions, and corresponding effect size*

	IVR		Desktop		<i>t</i>	<i>p</i>	<i>Cohen's d</i>	95% CI	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				<i>LL</i>	<i>UL</i>
Multiple-choice	0.41	0.2	0.4	0.2	0.34	0.735	0.06	-0.28	0.39
Self-efficacy	3.92	0.46	3.83	0.42	1.19	0.235	0.2	-0.13	0.54
Learning Agency	3.55	0.49	3.56	0.5	-0.15	0.882	-0.03	-0.36	0.31

Note. IVR = immersive virtual reality; CI = confidence interval; SD = standard deviation; LL = lower limit; UL = upper limit.

A significant difference between conditions was found for presence at post-first ( $t(130.87) = -11.01$ ,  $p < 0.001$ ,  $d = -1.88$ , 95% CI [-2.28, -1.48]) and post-second ( $t(112.95) = -13.705$ ,  $p < 0.001$ ,  $d = -2.37$ , 95% CI [-2.80, -1.93]), with participants in the IVR condition reporting higher feelings of presence than participants in the desktop condition after treatment. These results indicate that the treatment manipulation was successful.

No significant effect of the condition was found on self-efficacy at post-first ( $F(1,129) = 2.37$ ,  $p = 0.126$ ,  $\eta^2_p = 0.02$ , 90% CI [0, 0.07]) and post-second ( $F(1,129) = 0.01$ ,  $p = 0.757$ ,  $\eta^2_p < 0.001$ , 90% CI [0, 0.3]), thus not supporting H<sub>1</sub>. Bernacki and colleagues (2015) investigated the stability and change of high school students' self-efficacy during learning math, measuring self-efficacy four times while students were solving math problems. They found that self-efficacy varied over the course of the learning task. Their results showed that compared to the first self-efficacy measure taken, participants reported higher self-efficacy at the second time measured but lower at the third and fourth time measured, which is aligned with our findings. Bandura (1986) stated that learners perceive their efficacy based on past evidence of their performance completing a task. We speculate that their overall self-efficacy perception influenced participants' self-efficacy scores at pre-test. However, after treatment sessions 1 and

2, self-efficacy could have been influenced by their performance in the learning assessments, given that they answered the self-efficacy questionnaire after answering the learning assessment. Also, they could have compared their answers at pre-test and the videos' information and realized possible mistakes made at pre-test, contributing to a decrease in self-efficacy after the treatment sessions.

The models investigating the immediate effects of treatment on learning (RQ<sub>1</sub>) showed no condition effects on the multiple-choice assessment at post-first or post-second (post-first:  $F(1,133) = 0.10, p = 0.758, \eta^2_p < 0.001$ , 90% CI [0, 0.03], post-second:  $F(1,133) = 0.25, p = 0.617, \eta^2_p < 0.001$ , 90% CI [0, 0.02]). The analyses on each cognitive process revealed a significant effect of condition on knowledge creation at post-second ( $F(1,133) = 17.15, p < 0.001, \eta^2_p = 0.05$ , 90% CI [0.01, 0.13]). Post Hoc power analysis showed that this test achieved 0.77 power. This is a novel finding in the VR research field and corroborates Bell and Fogler's (1995) theoretical statement that virtual reality can address higher cognitive levels of learning, such as analysis and evaluation. Also, because IVR videos are less distracting and involve less cognitive load to interact with than CG-based IVR experiences, it may set cognitive capabilities free that enhance higher cognitive levels and support conceptual learning.

Although there were no effects of the condition on self-efficacy, we followed Rucker and colleagues' (2011) recommendations to run the mediation analysis if a relationship is of research interest. Causal mediation analysis using nonparametric bootstrap confidence intervals with the percentile method was run over post-first and post-second scores, with the condition as the independent variable, learning agency as the mediator, and self-efficacy as the outcome, and learning as a control. Results showed that learning agency mediated the relationship between condition and self-efficacy (ACME<sup>4</sup> = 0.031,  $p = 0.014$ , 95% CI [0, 0.06]; ADE<sup>5</sup> = 0.11,  $p = 0.050$ , 95% CI [0, 0.22]), supporting H<sub>2</sub>. Although we investigated agency as it relates to learning specifically, this result is aligned with the Cognitive Affective Model of Immersive Learning theoretical framework proposed by Makransky and Petersen (2021), in which agency would be an IVR affordance between immersion and self-efficacy.

The analyses investigating the long-term effects of treatment on learning (RQ<sub>2</sub>) revealed no effect of condition at delayed post-test on the multiple-choice learning assessments ( $F(1,133) = 2.04, p = 0.155, \eta^2_p = 0.12$ , 90% CI [-0.04, 0.28]) or the Bloom's taxonomy levels (knowledge understanding:  $F(1,133) = 1.38, p = 0.242, \eta^2_p < 0.01$ , 90% CI [0, 0.06]; knowledge application:  $F(1,133) = 0.89, p = 0.346, \eta^2_p < 0.01$ , 90% CI [0, 0.05]; knowledge

---

<sup>4</sup> ACME = average causal mediation effects

<sup>5</sup> ADE = average direct effects

creation:  $F(1,133) = 1.15, p = 0.284, \eta^2_p < 0.01, 90\% \text{ CI } [0, 0.01]$ ). Moreover, the analyses to investigate long-term effects of intervention on self-efficacy (RQ<sub>3</sub>) showed no significant effect of condition at delayed post-test ( $F(1,133) = 0.06, p = 0.800, \eta^2_p < 0.001, 90\% \text{ CI } [0, 0.02]$ ). While no long-term effects of IVR videos on self-efficacy are reported in the literature, this finding is aligned with Stepan and colleagues' (2017) study that compared CG-based IVR and computer-based materials for neuroanatomy learning. In their study, no significant difference for the learning assessment was found between conditions eight weeks after treatment.

### General Discussion

The present research investigated the effects of watching IVR videos or traditional videos on self-efficacy and conceptual learning. Two studies were carried out, and measures for learning, self-efficacy, learning agency, and presence were taken at multiple time points.

Experiment 1 showed a significant effect of condition on self-efficacy. Although this effect was not found in Experiment 2, a mediation effect of learning agency between condition and self-efficacy was found. The possibility of visually exploring the video in the IVR condition may have contributed to increasing participants' self-efficacy in the IVR condition in Experiment 1. Aligned with this finding, previous studies investigating the effects of visual exploration and active learning on self-efficacy have shown positive impacts of exploratory behavior on self-efficacy (Hardy III et al., 2014 and 2019). Self-efficacy towards STEM among middle-school students has been shown to be an important predictor of career choices (Gibbons & Borders, 2010). Although more research is needed on the relationship between immersion and self-efficacy, the positive effect found in Experiment 1, indicates that IVR can potentially enhance self-efficacy in science learning among middle-school girls. The mediation effect of learning agency between condition and self-efficacy found in Experiment 2 indicates that this may be a plausible mechanism underlying the positive effects of immersion on self-efficacy. However, more research is needed to understand this effect better.

When investigating learning, there were no significant differences between conditions on the multiple-choice question scores in Experiments 1 and 2, indicating that more immersion did not lead to better conceptual learning. IVR videos were not better than 2D videos in three out of the four measures of learning (multiple-choice questions, knowledge understanding, and knowledge application). Hence there is no strong evidence for the instructional effectiveness of IVR videos.

Notably, in Experiment 2, watching IVR videos showed a positive effect on a higher cognitive level of learning, knowledge creation, compared to using a computer monitor. Although the effect size was small, it was interesting to find a difference in knowledge creation just by varying the immersion level, given that the narrative and visuals were the same between conditions. This novel finding brings experimental evidence for the theoretical statement from Bell and Fogler (1995) that virtual reality would be capable of addressing higher cognitive levels of learning. This novelty may also be explained by the fact that IVR videos were used instead of CG-based IVR. The IVR videos usually add less extraneous load when compared to CG-based IVR, and they may have free cognitive resources that enable the enhancement of higher cognitive levels of learning. Given the small effect size of this finding, there is a need for future studies replicating these results of IVR videos' effects on Bloom's Taxonomy of learning objectives.

In light of the cognitive load theory, high levels of interactivity and constant animations, usually present in CG-based IVR, can add extraneous load to the learning material and hinder learning (Fiorella & Mayer, 2018; Parong & Mayer, 2018). That can be inferred by the fact that no study reporting lower learning gains when using IVR videos compared to traditional videos was found in the literature, while some studies using CG-based IVR have reported lower learning gains than traditional videos (Dede et al., 2000; Parong & Mayer, 2018). In this sense, IVR videos may maximize agency while minimizing cognitive load and distraction.

In addition, the medium itself cannot solely account for the learning outcomes (Clark, 1994), and it is essential to highlight the role the design and content of the instructional material play in learning. In CG, it is possible to develop almost any type of experience. For example, it is possible to create an experience in which objects do not fall when dropped. Also, CG allows users to select, move and manipulate objects in the virtual environment. This affordance of CG is often explored to develop entertainment content, such as in games. In 360 videos, the experience is produced using real images captured using a 360-degree camera, and the user interacts with the content through head movements. It makes the development cheaper and faster than CG, given that no 3D models or interaction feedback other than camera field of view need to be developed (Violante et al., 2019). Thus, 360-degrees videos seem useful in creating educational materials or documentaries quickly and at scale. Mayer (2008) points out the importance of developing learning multimedia experiences that reduces extraneous cognitive processing (processing information that is not related to the learning goal) and favors essential and generative cognitive processing (selecting relevant information and integrating it with prior knowledge). Fiorella and Mayer

(2018) provided guidelines on what is effective in terms of content, design, and learner attributes when developing instructional videos. For example, they reported the positive effects of segmenting the video and using different camera viewpoints on learning complex content.

*Coral Compass* and *The Crystal Reef* were chosen for the present studies because they were developed by experts in the VR, education, and marine fields to ensure that they followed scientific instructional material standards. Both videos were proven effective for learning for IVR and desktop conditions, as a significant increase in learning was shown in both conditions after treatment in both studies. Also, flat versions of the IVR videos were used in the desktop condition to avoid a confounding effect usually found in studies comparing videos to lectures or video to animated multimedia materials (Meyer et al., 2019).

Finally, our findings suggest that the effects found on self-efficacy and conceptual learning were short-termed, as they were not found five weeks after the treatment sessions. These findings align with previous studies investigating the effects of immersion on affective domains of learning. They are important in light of a changing society in which technology use is increasing fast, and devices continue to augment the level of digital immersion.

### **Limitations and future directions**

Although results from this study yielded relevant findings, there are some limitations. First, even though participants were randomly assigned to each condition, there was a significant difference between conditions for the variables measured at pre-test in Experiment 1. A statistical method that considers pre-test differences was selected in order to analyze the data properly. However, it does not rule out the possibility of interference at post-test results. A matched random assignment and a larger sample were successfully used in Experiment 2 to address this shortcoming from Experiment 1. Second, we didn't measure their prior experience with VR, hence novelty effects could have influenced the results. Third, participants were female students from a school situated in an affluent neighborhood. No race, ethnicity, or socio-economic status data were collected, limiting the generalization of the results. These findings should be taken with care when considering other populations.

Assessing learning at multiple times through multiple-choice and open-ended questions has some inherent limitations. Although multiple-choice questions are considered standard practice in learning assessments, they may cause some inaccuracy due to students' excessive training in answering these kinds of questions. Some students' scores can be impacted by their ability to identify the most probable correct answer. To reduce this impact, we also

used open-ended questions to assess learning. However, even open-ended questions have limitations, mainly because they rely on subjective scoring and prompt fatigue. An objective rubric was developed, and blind score ratings were used in both studies to reduce the subjectivity of this assessment. Also, the treatment was split into two 45-minute sessions in Experiment 2 as an attempt to reduce fatigue. Although the focus of this investigation was the comparison between conditions, and both conditions were subject to similar testing effects driven by the multiple tests study design, a testing effect could have impacted the learning results over time.

In this research, participants answered the questions assessing learning before the questions assessing self-efficacy. Future studies should investigate if the sequence that the learning and self-efficacy questions are presented at post-test makes a difference in participants' self-efficacy, given that the learning performance has been shown to impact self-efficacy.

To alleviate these limitations, future studies should consider a diverse population, multiple IVR exposures and use multiple methods to assess learning to improve the generalizability of these findings. Also, future studies must investigate the associations between interactivity, conceptual learning, and self-efficacy. A future study could compare watching a 360-degree video through a headset with watching it on a desktop computer in which the participant's eye movements are tracked and the point of view is refreshed accordingly while the image is centralized to the gaze direction. This way, participants in the desktop condition would have a similar interaction as participants in the IVR condition. Moreover, future studies should consider the unique aspects of both IVR videos and CG-based IVR environments, such as interactivity, realism, and their effects on cognitive processes, and investigate the associations between interactivity on learning and self-efficacy that goes beyond head movements. Finally, these studies should investigate the long-term effects of IVR videos exposure, such as measuring variables months after the treatment sessions, to contribute to the practical implications of IVR videos in education.

### References

- Allcoat, D., Hatchard, T., Azmat, F., Stansfield, K., Watson, D., & von Mühlennen, A. (2021). Education in the Digital Age: Learning Experience in Virtual and Mixed Realities. *Journal of Educational Computing Research*, 59(5), 795-816. doi: 10.1177/0735633120985120.
- 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>

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Abridged edition. White Plains, NY: Longman.
- Bailenson, J. N. (2020, September 18). Is VR the future of corporate training? *Harvard Business Review*. Retrieved March 11, 2022, from <https://hbr.org/2020/09/is-vr-the-future-of-corporate-training>
- Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences, 17*(1), 102-141. <https://doi.org/10.1080/10508400701793141>
- Baker, C. W., Little, T. D., & Brownell, K. D. (2003). Predicting adolescent eating and activity behaviors: the role of social norms and personal agency. *Health Psychology, 22*(2), 189-198. <https://doi.org/10.1037/0278-6133.22.2.189>
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ, USA.
- Bandura, A. (1995). *Self-efficacy in Changing Societies*. Cambridge University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan. London, UK.
- Bell, J. T., & Fogler, H. S. (1995, June). The investigation and application of virtual reality as an educational tool. In: *Proceedings of the American society for engineering education annual conference*. Anaheim, CA. 1718-1728.
- Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2015). Examining self-efficacy during learning: Variability and relations to behavior, performance, and learning. *Metacognition and Learning, 10*(1), 99-117. doi: [10.1007/s11409-014-9127-x](https://doi.org/10.1007/s11409-014-9127-x)
- Blascovich, J., & Bailenson, J. (2011). *Infinite reality*. New York, NY: HarperCollins.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: Handbook I: Cognitive Domain*. New York: David McKay Co Inc.
- Cho, S., Goodman, M., Oppenheimer, B., Codling, J., & Robinson, T. (2009). Images of women in STEM fields. *Journal of Science Communication, 8* (3) A03. DOI: <https://doi.org/10.22323/2.08030203>

- Cozby, P. C., Bates, S., Krageloh, C., Lacherez, P., & Van Rooy, D. (1977). *Methods in behavioral research*. Houston, TX: Mayfield publishing company.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42, 21-29. <https://doi.org/10.1007/BF02299088>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- 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. <https://doi.org/10.1080/15213269.2015.1015740>
- David, N., Newen, A., & Vogeley, K. (2008). The “sense of agency” and its underlying cognitive and neural mechanisms. *Consciousness and cognition*, 17(2), 523-534. <https://doi.org/10.1016/j.concog.2008.03.004>
- Dede, C. (2010). Comparing frameworks for 21st Century Skills. In J. Bellanca & R. Brandt (Eds.), *21st Century Skills: Rethinking how students learn* (p. 51–75). Bloomington, IN: Solution Tree Press.
- Dede, C., Salzman, M., Loftin, R. B., & Ash, K. (2000). Using virtual reality technology to convey abstract scientific concepts. In M. J. Jacobson & R.B. Kosma (Eds.), *Innovations in science and mathematics education: advanced designs for technologies of learning* (pp.361-414). Lawrence Erlbaum.
- Fauville, G., Strang, C., Cannady, M. A., & Chen, Y. F. (2019). Development of the International Ocean Literacy Survey: measuring knowledge across the world. *Environmental Education Research*, 25(2), 238-263. <https://doi.org/10.1080/13504622.2018.1440381>
- Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, 89, 465-470. <https://doi.org/10.1016/j.chb.2018.07.015>.
- Francis, E. R., Bernard, S., Nowak, M. L., Daniel, S., & Bernard, J. A. (2020). Operating Room Virtual Reality Immersion Improves Self-Efficacy Amongst Preclinical Physician Assistant Students. *Journal of Surgical Education*, 77(4), 947-952 <https://doi.org/10.1016/j.jsurg.2020.02.013>.
- Gibbons, M. M., & Borders, L. D. (2010). A measure of college-going self-efficacy for middle school students. *Professional School Counseling*, 13(4). <https://doi.org/10.1177/2156759X1001300404>
- Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*. 5(81). <https://doi.org/10.3389/frobt.2018.00081>

- Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5(74). <https://doi.org/10.3389/frobt.2018.00074>
- Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, 30(2), 207-241. <https://doi.org/10.1016/j.cedpsych.2004.08.001>
- Hardy III, J. H., Day, E. A., Hughes, M. G., Wang, X., & Schuelke, M. J. (2014). Exploratory behavior in active learning: A between-and within-person examination. *Organizational Behavior and Human Decision Processes*, 125(2), 98-112. <https://doi.org/10.1016/j.obhdp.2014.06.005>
- Hardy III, J. H., Day, E. A., & Arthur Jr, W. (2019). Exploration-exploitation tradeoffs and information-knowledge gaps in self-regulated learning: Implications for learner-controlled training and development. *Human Resource Management Review*, 29(2), 196-217. <https://doi.org/10.1016/j.hrmr.2018.07.004>
- Harrington, C. M., Kavanagh, D. O., Ballester, G. W., Ballester, A. W., Dicker, P., Traynor, O., ... & Tierney, S. (2018). 360 Operative videos: A randomized cross-over study evaluating attentiveness and information retention. *Journal of surgical education*, 75(4), 993-1000. <http://dx.doi.org/10.1016/j.jsurg.2017.10.010>
- Hite, R. L., Jones, M. G., Childers, G. M., Ennes, M., Chesnutt, K., Pereyra, M., & Cayton, E. (2019). Investigating Potential Relationships Between Adolescents' Cognitive Development and Perceptions of Presence in 3-D, Haptic-Enabled, Virtual Reality Science Instruction. *Journal of Science Education and Technology*, 28(3), 265-284. <https://doi.org/10.1007/s10956-018-9764-y>
- Huang, X., & Mayer, R. E. (2019). Adding Self-Efficacy Features to an Online Statistics Lesson. *Journal of Educational Computing Research*, 57(4), 1003-1037. <https://doi.org/10.1177%2F0735633118771085>
- Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), 1515-1529. <https://doi.org/10.1007/s10639-017-9676-0>
- Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, 5(81). <https://doi.org/10.3389/frobt.2018.00081>
- Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct?. *Journal of personality and social psychology*, 83(3), 693-710. <https://doi.org/10.1037/0022-3514.83.3.693>

- Jun, H., Miller M., Herrera, F., Reeves, B., & Bailenson, J. N. (2020). Stimulus Sampling with 360-Videos: Examining Head Movements, Arousal, Presence, Simulator Sickness, and Preference on a Large Sample of Participants and Videos. *IEEE Transactions on Affective Computing*, 5 (2), 112-125. doi: 10.1109/TAFFC.2020.3004617.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218. [https://doi.org/10.1207/s15430421tip4104\\_2](https://doi.org/10.1207/s15430421tip4104_2)
- Krathwohl, D. R., & Anderson, L. W. (2009). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Kim, K., Maloney, D., Bruder, G., Bailenson, J. N., & Welch, G. F. (2017). The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR. *Computer Animation and Virtual Worlds*, 28(3-4), e1771.
- Kuo, Y. C., Walker, A. E., Schroder, K. E., & Belland, B. R. (2014). Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *The internet and higher education*, 20, 35-50. <https://doi.org/10.1016/j.iheduc.2013.10.001>
- Laha, B., Bowman, D. A., & Schiffbauer, J. D. (2013). Validation of the MR simulation approach for evaluating the effects of immersion on visual analysis of volume data. *IEEE transactions on visualization and computer graphics*, 19(4), 529-538. DOI: 10.1109/TVCG.2013.43.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863, 1–12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lee, S. H., Sergueeva, K., Catangui, M., & Kandaurova, M. (2017). Assessing Google cardboard virtual reality as a content delivery system in business classrooms. *Journal of Education for Business*, 92, 153–160. <https://doi.org/10.1080/08832323.2017.1308308>.
- Li, B. J., Bailenson, J. N., Pines, A., Greenleaf, W. J., & Williams, L. M. (2017). A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self-report measures. *Frontiers in psychology*, 8, 2116, 1-10. <https://doi.org/10.3389/fpsyg.2017.02116>
- McGivney, E. (2021). Virtual Reality as a Tool for Increasing Student Motivation in Remote Learning: Student Experiences of Agency in Immersive Learning Environments. *APA Convention Division 15*.

- Maddux, J. E. (2016). Self-efficacy. In Trusz, S. (Ed.), Babel, P. (Ed.). *Interpersonal and Intrapersonal Expectancies*. (p. 41-46). London: Routledge, <https://doi.org/10.4324/9781315652535>
- Mado, M., Fauville, G., Jun, H., Most, E., Strang, C. & Bailenson, J. N. (2022). Accessibility of Educational Virtual Reality for Children during the COVID-19 Pandemic. *Technology, Mind and Behavior*, 3 (1) Spring 2022. DOI: 10.1037/tmb0000066
- Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020a). 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., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 35(6), 691-707. <https://doi.org/10.1111/jcal.12375>
- Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, 66(5), 1141–1164. <https://doi.org/10.1007/s11423-018-9581-2>.
- Makransky, G., Mayer, R., Nøremølle, A., Cordoba, A. L., Wandall, J., & Bonde, M. (2020b). Investigating the feasibility of using assessment and explanatory feedback in desktop virtual reality simulations. *Educational Technology Research and Development*, 68(1), 293-317. <https://doi.org/10.1007/s11423-019-09690-3>
- 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>
- Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *American psychologist*, 63(8), 760. DOI: 10.1037/0003-066X.63.8.760
- Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, 33(5), 403–423. <https://doi.org/10.1111/jcal.12197>.
- Mayer, R. E., Paul, R., Raths, J., & Wittrock, M. C. (2001). A Taxonomy for Learning Teaching and Assessing. In: Anderson, L. & Krathwohl, D. (Eds.), *A revision of Bloom's Taxonomy of Educational Objectives*. Abridged Edition, Longman.

- Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education, 140*, 103603. <https://doi.org/10.1016/j.compedu.2019.103603>
- Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94*(3), 598-610. <https://doi.org/10.1037/0022-0663.94.3.598>
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38.
- Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 12*(5), 481-494. <https://doi.org/10.1162/105474603322761289>
- O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications, 9*, Article 3777. <https://doi.org/10.1038/s41467-018-06292-0>
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785-797. <https://doi.org/10.1037/edu0000241>.
- Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In *Emotions, technology, design, and learning* (pp. 131-161). Academic Press. <https://doi.org/10.1016/B978-0-12-801856-9.00007-4>
- Pekrun, R., Goetz, T., & Perry, R. P. (2005). Achievement emotions questionnaire (AEQ). User's manual. Unpublished Manuscript, University of Munich, Munich.
- Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315-341. <https://doi.org/10.1007/s10648-006-9029-9>.
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36-48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology, 51*(6), 2099-2115. <https://doi.org/10.1111/bjet.12991>

- Queiroz, A.C.M., Nascimento, A.M., Tori, R., Leme, M.I.S. (2018) Using HMD-Based Immersive Virtual Environments in Primary/K-12 Education. In: Beck D. et al. (eds) Immersive Learning Research Network. iLRN 2018. *Communications in Computer and Information Science*, 840. Springer, Cham.  
[https://doi.org/10.1007/978-3-319-93596-6\\_11](https://doi.org/10.1007/978-3-319-93596-6_11)
- Queiroz, A. C. M., Fauville, G., Herrera, F., da Silva Leme, M. I., & Bailenson, J. (2021, October 9). Effects of IVR videos on conceptual knowledge and self-efficacy. DOI 10.17605/OSF.IO/MPK2J
- Rebelo, F., Noriega, P., Duarte, E., and Soares, M. (2012). Using virtual reality to assess user experience. *Hum. Factors*, 54, 964–982. doi: 10.1177/0018720812465006
- Roussou, M., & Slater, M. (2017). Comparison of the effect of interactive versus passive virtual reality learning activities in evoking and sustaining conceptual change. *IEEE Transactions on Emerging Topics in Computing*, 8(1), 233-244. <https://doi.org/10.1109/TETC.2017.2737983>
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5(6), 359-371.  
<https://doi.org/10.1111/j.1751-9004.2011.00355.x>
- Rupp, M. A., Odette, K. L., Kozachuk, J., Michaelis, J. R., Smither, J. A., & McConnell, D. S. (2019). Investigating learning outcomes and subjective experiences in 360-degree videos. *Computers and Education*, 128, 256–268.  
<https://doi.org/10.1016/j.compedu.2018.09.015>.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332. <https://doi.org/10.1038/nrn1651>
- Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*, 17(4), 251-270. <https://doi.org/10.1016/j.asw.2012.07.002>
- Schlosser, M. E. (2012). Causally efficacious intentions and the sense of agency: In defense of real mental causation. *Journal of Theoretical and Philosophical Psychology*, 32(3), 135-160.  
<https://psycnet.apa.org/doi/10.1037/a0027618>
- Schunk, D. H. (1995). Self-efficacy and education and instruction. In *Self-efficacy, adaptation, and adjustment* (pp. 281-303). Springer, Boston, MA. [https://doi.org/10.1007/978-1-4419-6868-5\\_10](https://doi.org/10.1007/978-1-4419-6868-5_10)

- Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In: Wentzel, K. & Miele, D. (2016) *Handbook of motivation at school*, 2, 34-54. New York: Routledge.
- Slater, M. (2003). A note on presence terminology. *Presence connect*, 3(3), 1-5.
- Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *Neuroimage*, 46(1), 290-298.  
<https://doi.org/10.1016/j.neuroimage.2009.01.047>
- Stanley, L. M., & Edwards, M. C. (2016). Reliability and Model Fit. *Educational and psychological measurement*, 76(6), 976–985. <https://doi.org/10.1177/0013164416638900>
- Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., & Iloreta, A. (2017). Immersive virtual reality as a teaching tool for neuroanatomy. *International forum of allergy & rhinology*, 7(10), 1006-1013. <https://doi.org/10.1002/alr.21986>
- Sun, J. C. Y., & Rueda, R. (2011). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British journal of educational technology*, 43(2), 191-204.  
<https://doi.org/10.1111/j.1467-8535.2010.01157.x>
- 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>
- Sutherland, I. E. (1968, December). A head-mounted three-dimensional display. In *Proceedings of ACM fall joint computer conference, part I* (pp. 757-764). <https://doi.org/10.1145/1476589.1476686>
- Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in psychology*, 8, 1552.  
<https://doi.org/10.3389/fpsyg.2017.01552>
- Triberti, S., and Riva, G. (2016). Being present in action: a theoretical model about the “interlocking” between intentions and environmental affordances. *Frontiers in Psychology*, 6, 2052.  
<https://doi.org/10.3389/fpsyg.2015.02052>
- 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>

- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of educational research*, 78(4), 751-796. <https://doi.org/10.3102%2F0034654308321456>
- Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)*, 6(2), 227. doi:10.4172/2161-1165.1000227
- Violante, M. G., Vezzetti, E., & Piazzolla, P. (2019). Interactive virtual technologies in engineering education: Why not 360° videos?. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 13(2), 729-742. <https://doi.org/10.1007/s12008-019-00553-y>
- Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140 (4), 1174-1204. <https://doi.org/10.1037/a0036620>
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In: Bandura, A. *Self-efficacy in changing societies*, 1, (pp. 202-231). Cambridge: Cambridge University Press.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-91. <https://doi.org/10.1006/ceps.1999.1016>
- Walshe, N., & Driver, P. (2019). Developing reflective trainee teacher practice with 360-degree video. *Teaching and Teacher Education*, 78, 97–105. <https://doi.org/10.1016/j.tate.2018.11.009>.
- Webster, R. (2015). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, 24, 1319–1333. doi: 10.1080/10494820.2014.994533

## Appendix A

### Experiment 1 and 2 - Open-ended questions:

1. Explain what you know about how human activities change ocean chemistry.
2. Describe ways that could help reduce ocean acidification.
3. Propose strategies to increase engagement of the general public with ocean acidification.
4. Palau's coral reefs are considered to be one of the “Seven Underwater Wonders of the World.” What actions is the Palau government taking to reduce damages to the coral reefs?
5. How could the Palauan government’s actions protect the coral reefs be applied to other environmental issues, for example, the fast erosion of South California beaches?
6. Considering the Palauan government’s efforts to protect the coral reefs, think about one issue from your community that needs immediate attention. Propose an action plan that could help address this issue.

### Experiment 1 and 2 - Multiple-choice questions:

1. What is the main cause of ocean acidification?
  - Chemical spills in the ocean
  - Acid rain
  - Absorption of carbon dioxide by the ocean
  - Warmer ocean temperatures
2. Corals, shellfish, and other marine organisms use the carbon dissolved in the ocean to:
  - build shells.
  - breathe underwater.
  - regulate body temperature.
  - assist in reproduction.
3. Which of the following is a result of human-caused carbon dioxide emissions?
  - Ocean salinity and the frequency of oil spills are increasing.
  - Coral reefs are degrading, and the diversity of ocean life is decreasing.
  - The frequency of oil spills is increasing, and coral reefs are degrading.
  - The diversity of ocean life is decreasing, and ocean salinity is increasing.
4. How are carbon dioxide emissions affecting the ocean?
  - Carbon and salt are increasing in the ocean.
  - Phytoplankton and salt are increasing in the ocean.
  - Phytoplankton and temperatures are increasing in the ocean.
  - Carbon and temperatures are increasing in the ocean.
5. Which sentence best describes how tourism in Palau is damaging the coral reefs?

- Tourism in Palau is responsible for a great part of carbon dioxide emissions in the world.
  - Tourists in Palau throw more trash into the ocean than most tourists in other places.
  - Tourists are damaging the corals with their fins while diving.
  - Acid rain is increasing in Palau due to tourism's carbon dioxide emissions.
6. Select the land practice that can greatly help to protect the coral reefs in Palau.
- Place taro farms in strategic areas.
  - Eliminate the mangrove forests that run along the coastline.
  - Replace the mangrove forests with farms.
  - Avoid sediment traps along coastline.

### **Experiment 1 - Self-efficacy questionnaire:**

**Participants were asked to express their agreement to each statement below, in a 5-point Likert scale (1= *strongly disagree*, 5 = *strongly agree*):**

1. I think that I can be proud of what I know about this subject.
2. I think that learning science is important because I can use it in my daily life.
3. No matter how much effort I put in, I cannot learn science. (Reverse coded)
4. It is important to have the opportunity to satisfy my own curiosity when learning science.
5. Whether the science content is difficult or easy, I am sure that I can understand it.
6. I am not confident about understanding difficult concepts. (Reverse coded)
7. I study more than required because I enjoy it so much.
8. When science activities are too difficult, I give up or only do the easy parts. (Reverse coded)
9. I am willing to participate in this science course because the content is exciting.
10. The subject scares me since I don't fully understand it. (Reverse coded)
11. I am so happy about the progress I made that I am motivated to continue studying.
12. This subject is so enjoyable that I am motivated to do extra readings about it.

## Appendix B

### Experiment 2 – Self-efficacy questionnaire:

Participants were asked to express their agreement to each statement below, in a 5-point Likert scale (1= *strongly disagree*, 5 = *strongly agree*):

1. Whether the science content is difficult or easy, I am sure that I can understand it.
2. I am not confident about understanding difficult science concepts. (Reverse coded)
3. I am sure that I can do well on science tests.
4. No matter how much effort I put in, I cannot learn science. (Reverse coded)
5. When science activities are too difficult, I give up or only do the easy parts. (Reverse coded)
6. During science activities, I prefer to ask other people for the answer rather than think for myself. (Reverse coded)
7. When I find the science content difficult, I do not try to learn it. (Reverse coded)
8. I think that learning science is important because I can use it in my daily life.
9. I think that learning science is important because it stimulates my thinking.
10. In science, I think that it is important to learn to solve problems.
11. In science, I think it is important to participate in inquiry activities.
12. It is important to have the opportunity to satisfy my own curiosity when learning science.

### Experiment 2 - Learning agency questionnaire:

Participants were asked to express their agreement to each statement below, in a 5-point Likert scale (1= *strongly disagree*, 5 = *strongly agree*):

1. I can control my academic performance.
2. I learn more when I am the one selecting what I am going to learn.
3. My choices about what to study influence my learning outcomes.
4. I prefer other people to tell me what I have to study rather than deciding on my own.
5. Being able to choose where to focus my attention when watching a video is important to me.