Long-term effectiveness of immersive vr simulations in undergraduate science learning

Lessons from a media-comparison study

Pande, Prajakt; Thit, Amalie; Sørensen, Anja Elaine; Mojsoska, Biljana; Møller, Morten Erik; Jepsen, Per Meyer

Published in:
Research in Learning Technology

DOI:
10.25304/rlt.v29.2482

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Pande, P., Thit, A., Sørensen, A. E., Mojsoska, B., Møller, M. E., & Jepsen, P. M. (2021). Long-term effectiveness of immersive vr simulations in undergraduate science learning: Lessons from a media-comparison study. *Research in Learning Technology*, *29*(29), 1-24. Article 2482. https://doi.org/10.25304/rlt.v29.2482

Research in Learning Technology
Vol. 29, 2021

ASSOCIATION FOR LEARNING TECHNOLOGY ALT

ORIGINAL RESEARCH ARTICLE

Long-term effectiveness of immersive VR simulations in undergraduate science learning: lessons from a media-comparison study

Prajakt Pande[a,b]*, Amalie Thit[c], Anja Elaine Sørensen[c], Biljana Mojsoska[b,c], Morten E. Moeller[b,c] and Per Meyer Jepsen[b,c]

[a]*Department of People and Technology, Roskilde University, Roskilde, Denmark;*
[b]*Centre for Virtual Learning Technologies, Roskilde University, Roskilde, Denmark;*
[c]*Department of Science and Environment, Roskilde University, Roskilde, Denmark*

(Received: 18 June 2020; Revised: 9 November 2020; Accepted: 16 November 2020; Published: 18 January 2021)

Our main goal was to investigate if and how using multiple immersive virtual reality (iVR) simulations and their video playback, in a science course, affects student learning over time. We conducted a longitudinal study, in ecological settings, at an undergraduate field-course on three topics in environmental biology. Twenty-eight undergraduates were randomly assigned to either an iVR-interaction group or a video-viewing group. During the field-course, the iVR group interacted with a head-mounted device-based iVR simulation related to each topic (i.e. total three interventions), while the video group watched a pre-recorded video of the respective simulation on a laptop. Cognitive and affective data were collected through the following checkpoints: a pre-test before the first intervention, one topic-specific post-test immediately after each intervention, a final post-test towards the end of the course, and a longitudinal post-test deployed approximately 2 months after the course. Through a descriptive analysis, it was found that student performance on the knowledge tests increased considerably over time for the iVR group but remained unchanged for the video group. While no within- or between-group differences were noted for intrinsic motivation and self-efficacy measures, students in the iVR group enjoyed all the simulations, and perceived themselves to benefit from those simulations.

Keywords: virtual reality; longitudinal; science education; higher education; educational technology

Introduction

Head-mounted device-based (i.e. HMD-based) immersive virtual reality (hereafter, iVR) simulation interfaces are powerful learning and teaching media in science, technology, engineering and mathematics (STEM). They afford learner's engagement and interaction with abstract or inaccessible scientific entities, phenomena and concepts in innovative ways (Pande 2020; Pande and Chandrasekharan, 2017). With the advent of diverse iVR interfaces and their potential applications in higher educational settings,

*Corresponding author. Email: pande@ruc.dk, prajakt123@gmail.com

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

educational technologists are increasingly focusing on how these interfaces could be utilised to support student learning of scientific content (e.g. laboratory procedures, concepts) and also affective aspects of student engagement with that content (Kafai and Dede 2014). However, despite its promises, and the enthusiasm among educational technology communities, iVR is still not considered a mainstream technology in higher education (Cochrane 2016; Gutierrez-Maldonado, Andres-Pueyo, and Talarn-Caparros 2015). The actual implementation of iVR as an educational technology involves logistical, technical and economic challenges (Khor *et al*. 2016; Liagkou, Salmas, and Stylios 2019). Importantly, there exist several gaps in current research on instructional effectiveness of iVR in STEM education: Firstly, research on the learning–teaching effectiveness of iVR, particularly in university-level STEM education, is limited in volume (Jensen and Konradsen 2018; Radianti *et al*. 2020). Secondly, the existing volume of research fails to provide a coherent and conclusive picture of whether using iVR in university STEM teaching improves students' learning of the content (Checa and Bustillo 2020; Concannon, Esmail, and Roberts 2019). Thirdly, and perhaps intricately related to the second, there are limited process analyses of how learning happens in/with iVR (Concannon, Esmail, and Roberts 2019; Schott and Marshall 2018), particularly in ecological settings (Hew and Cheung 2010). Finally, research fails to provide insights into the long-term effectiveness and usefulness of iVR-based instruction (Sánchez, Lumberas, and Silva 1997).

This paper attempts to fill the gaps in the recent and still limited literature on the usefulness and effectiveness of iVR-based instruction in higher STEM education.

Recent research on iVR in STEM education: Gaps and limitations

In a systematic review, Concannon, Esmail and Roberts (2019) found that most studies reporting implementation of iVR-based instruction were situated within STEM education or allied fields (e.g. health sciences, psychology education). Most iVR studies in STEM education involved controlled experiments (e.g. iVR instruction vs. no instruction) and media comparison experiments evaluating the instructional effectiveness of iVR in comparison to other modes of instruction (e.g. traditional lecture-based, video-based, desktop-based). These studies typically investigate the effects of iVR instruction on various cognitive (e.g. procedural learning, retention and recall, conceptual learning) and non-cognitive (e.g. motivation, enjoyment) aspects of learning (Hew and Cheung 2010; Kavanagh *et al*. 2017; Parong and Mayer 2018). For instance, Lamb *et al*. (2018) recently showed that iVR not only leads to better learning outcomes (i.e. gain in test scores) on molecular biology topics, but also triggers significantly higher cognitive engagement and processing (e.g. neural activation) as compared to video lectures. The authors also reported that learning science through iVR simulations is cognitively (e.g. in terms of activity in certain brain areas) equivalent to engaging in serious educational games and hands-on activities (Lamb *et al*. 2018). On the other hand, a study comparing the learning and presence effects of two levels of immersion demonstrated that iVR instruction overloaded the students cognitively (Makransky, Terkildsen, and Mayer 2019). In this study, iVR did lead to a higher feeling of presence and liking among participants as opposed to a desktop. However, it yielded lower knowledge-related learning outcomes compared to a desktop. This indicates that liking an intervention does not necessarily lead to better learning outcomes.

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

Another study (Makransky, Borre-Gude, and Mayer 2019) compared the effects of iVR simulation, with those of an identical desktop simulation, and text-based instruction, in training undergraduate students for general lab safety. They found that students from the iVR condition significantly outperformed the other two conditions on practical lab tests and feeling of self-efficacy. However, iVR did not add any advantage to knowledge retention, intrinsic motivation, and perceived enjoyment, compared to the other two conditions (Makransky, Borre-Gude, and Mayer 2019). Several other studies have demonstrated positive effects of iVR on most affective aspects of learning regardless of: iVR's success in supporting actual content or conceptual learning (Garcia-Bonete, Jensen, and Katona 2019; Madden *et al*. 2020; Parong and Mayer 2018; Tan and Waugh 2013), the implementation-related logistical limitations of iVR (Ba *et al*. 2019; Makransky, Terkildsen, and Mayer 2019) and the cyber-sickness that iVR may induce (e.g. Moro, Štromberga, and Stirling 2017). These effects have been demonstrated in isolation and also in comparison to other media of instruction (e.g. desktop simulations, video demonstrations, text-based instruction). While research on iVR instruction fails to provide conclusive results on its effectiveness, even in comparison with other traditional as well as new instructional modes, studies examining iVR simulations in combination with other interventions, such as generative learning strategies (Fiorella and Mayer 2016), report positive effects on various cognitive and affective aspects of learning. In a media-comparison study, asking students to summarise parts of a recently experienced iVR lesson increased their knowledge-related learning outcomes in contrast to students who received instruction through traditional slide show-based lessons (Parong and Mayer 2018). In a media (iVR vs. video) and methods comparison experiment (pre-training vs. no-pre-training), students who received pre-training about certain concepts in biology (e.g. the cell) before receiving iVR instruction were found to perform significantly better in post-tests related to knowledge retention, transfer and self-efficacy, as compared to students who did not receive any pre-training, or those who received pre-training in combination with video-based instruction (Meyer, Omdahl, and Makransky 2019). In a similar media (iVR vs. video) and methods (enactment vs. no-enactment) comparison experiment, Andreasen *et al*. (2019) showed that performing an enactment of scientific procedures (e.g. pipetting) immediately after interacting with an iVR simulation of those procedures (i.e. iVR + enactment) leads to significantly better learning outcomes in terms of retention and transfer as compared to only interacting with an iVR simulation, watching its video or even watching a video and thereafter performing enactment (Andreasen *et al*. 2019; Makransky *et al*. 2020). They also found positive effects of the iVR + enactment intervention on intrinsic motivation and self-efficacy. However, these positive results may not identify and/or justify the specific role iVR plays in the learning process.

Most research on the instructional effectiveness of iVR in STEM education involves participants experiencing only a single exposure to the iVR environment (Concannon, Esmail, and Roberts 2019; Makransky, Borre-Gude, and Mayer 2019; Makransky, Terkildsen, and Mayer 2019; Southgate 2020; Vergara *et al*. 2019). As a result, it could be argued that the positive affective learning outcomes associated with iVR interventions reported by many of these studies, could be a temporary consequence of the well-known 'novelty effect' (Moos and Marroquin 2010). The novelty effect is typically manifested as, or characterised by, higher ratings reported by participants on the self-report affective measurement scales subjected to them immediately after an intervention. Anecdotally, this effect may result from participants'

increased excitement around the new technology. iVR is often an overwhelming experience for students, particularly for those with little or no prior exposure to immersive virtual environments. Hence, such students are more likely to rate themselves highly positively across the various affective measurement items immediately after such an experience (Chen *et al*. 2016). This could also explain the lack of conclusive success of iVR interventions in comparison to interventions based on other media, particularly in relation to knowledge retention, conceptual understanding, and other cognitive-procedural aspects of learning (Moro, Štromberga, and Stirling 2017; Radianti *et al*. 2020). Moreover, the promising affective influences of iVR on students may not be long-term as it has been found that students' increased interest in technology-based activities may not result in an extended interest in the corresponding academic content (Torff and Tirotta 2010). It is thus critical to the progress of research and development in the educational technology domain to understand the long-term instructional effectiveness of iVR-based interventions (Madani *et al*. 2016; Southgate 2020).

Finally, and most importantly, there is a lack of longitudinal studies focusing on the long-term usefulness and/or effectiveness of iVR and iVR-based interventions in STEM education (Chittaro and Buttussi 2015; Wu, Yu, and Gu 2020).

Study objective and research question

To contribute to the current body of research on instructional effectiveness of iVR in higher STEM education, and to address the above-mentioned issues related to iVR's long-term as well as ecological validity, a cross-sectional media-comparison quasi-experiment was conceptualised at a 6-day undergraduate science course. This experiment sought answers to the following multi-faceted research questions:

- How does student interaction with multiple iVR simulations on different topics in environmental biology affect their (1) learning of the content related to those topics, (2) intrinsic motivation, (3) self-efficacy, and (4) perceived learning, and enjoyment over time? (Cross-sectional/longitudinal patterns)
- How does the effect of iVR simulations on the above-mentioned learning outcomes compare with video-viewing (where students watch videos of the respective simulations)? (Media comparison)

Students participating in this course learned about the following three topics in environmental biology – photosynthesis, biodiversity, and food webs. For each topic, a student interacted with either an iVR simulation (iVR condition) or a pre-recorded video of the respective simulation (video condition). In addition, all students attended common lectures, fieldtrips, and laboratory work related to each topic. Through a pre-test, we collected data on students' prior knowledge related to all the three topics (i.e. established a knowledge baseline), as well as self-report affective measures related to students' engagement with those topics (e.g. self-efficacy, enjoyment). To understand possible changes in knowledge-related and affective learning outcomes over time, post-intervention data were collected in the following manner: one post-test immediately after each intervention (i.e. total three topic-specific post-tests – 1, 2 and 3); a final post-test on all the three topics on the 5th day of the course (final post-test), and a follow-up post-test (same as the pre-test and the final post-test) approximately 2 months after the last intervention (longitudinal post-test).

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

Pedagogical rationale behind the study

One major pedagogical motivation behind our iVR intervention, as well as the study, was the urge to facilitate experiential learning among students, of the relationships between experimental procedures and the scientific concepts involved in those procedures (Kolb 1984; Schott and Marshall 2018). University STEM students have often been found to struggle with a mismatch between their theoretical knowledge about a topic (e.g. photosynthesis and subcellular locations of the different reactions occurring during photosynthesis), and the practical or procedural knowledge related to that topic (e.g. extraction of photosynthetic pigments from algae; personal observation; OECD 2018). Due to limited time and resources, it is difficult for most university teachers to support students, for instance, by providing them more than one opportunity to perform experiment(s), and experience the procedures and the theoretical constructs behind them. The pedagogical interventions based on multiple iVR simulations tested in this study were conceptualised to help undergraduate students develop a sense of the connections between the theoretical material presented in the environmental biology course and the practical and hands-on elements of that course (e.g. how the theory behind sampling, or analysis techniques, connects with the choice of a specific sampling technique in the field). We hoped that the three simulations, as well as their videos, could help students experience the experiments related to the respective topics in environmental biology before they proceeded to perform them in the real laboratory.

In the next section, we discuss the study in detail in terms of the pedagogical and research-related steps taken.

The study

Sample

Twenty-eight fourth semester undergraduate students (14 female) from a major university in Denmark participated in a 6-day residential environmental biology fieldcourse at a pre-planned off-campus site. Randomly, half the students were assigned to the iVR condition (eight female) while the other half were given the video condition (six female).

The study was conducted according to the Helsinki declaration, and written consent was gathered from all participants prior to the study, in accordance with the European Union General Data Protection Regulation (GDPR) guidelines, as well as the university's local regulations.

Materials

iVR simulations and videos

The iVR simulations and their respective videos for the three topics in environmental biology were provided by the company named Labster. The simulations used in the study were: Pigment Extraction (Labster 2019a), Food web Structure (Labster 2019b), and Biodiversity (Labster 2019c). Figure 1 shows snapshots of some important moments from each simulation. For a detailed description of each simulation, please refer to Appendix 1.

Each student in the iVR group was provided with a Lenovo Mirage Solo HMD, which ran the respective Labster simulation through the Labster applet supported by the Google Daydream platform. Each simulation covered different types of activities

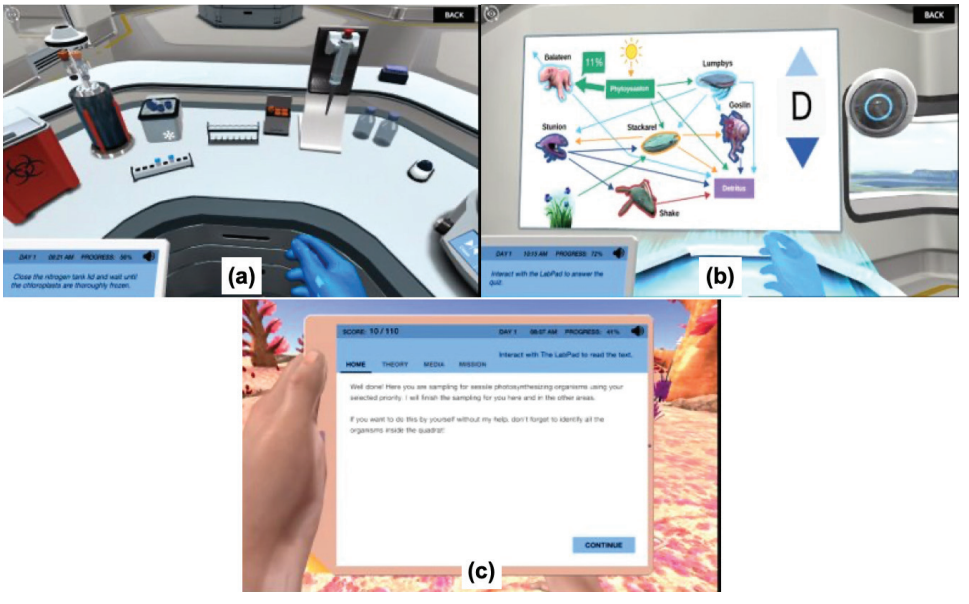

Figure 1. Snapshots of important moments from the three simulations used in the present study. (a) Pigment extraction – measurement of blank and pigment samples using a spectrophotometer, (b) Food webs – training scheme to observe the impact of different factors on the ecosystem (e.g. removal of primary food source) and (c) Biodiversity – sample collection for further identification.

related to the respective topic (e.g. laboratory work, fieldwork, reading of theory in text, problem-solving). Though reviewing the designs of interactive iVR learning environments is out of scope of this paper for practical reasons, it is important to comment briefly on the general design of the three simulations used in this study, primarily because this has implications for our experimental choices (e.g. comparison condition). Firstly, the learning activities in the simulations are linear and their sequence is predetermined – a student is first placed in a context (usually in the form of a story – e.g. encountering complex life forms while visiting an exoplanet), and is then posed a task, which requires the student to visit a laboratory (e.g. sampling and analysing biodiversity). Inside the simulation, the student is orally instructed by a drone about each step the student should take in order to proceed through the simulation (e.g. wear a lab-coat, start or run equipment by clicking a button, teleport to another desk, point-and-click on a device to use it, read theory on the pad if necessary). The instruction is also delivered in text through a virtual text-pad. Secondly, the interactivity in these simulation designs is particularly minimal, as all the student can do, is turn their head around to see elements in the 360-degree environment and point-and-click to activate an element (in comparison, many other iVR learning simulations use full-body or gesture-based interaction with or without haptic feedback). The point-and-click interaction happens through a laser pointer-like controller; it is not hugely different from the mouse-based interactions one may have in a desktop environment. This is primarily why, in the media-comparison experiment, we did not choose to compare the iVR condition with a desktop condition.

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

Comprehensive screen recording (video + audio) of each simulation is professionally made available by Labster. Each video is approximately 30 min long (roughly equivalent to the average time taken to complete the respective simulation). In terms of content coverage, the videos are identical to their respective simulation. As the sequence of events in most Labster simulations is fixed, the video recordings do not differ much from the simulations in content presentation.

Each student watched the respective video on a laptop screen and listened to the audio through headphones. The student could play, pause, review or forward the video at will.

Considering all the design aspects of the simulation (and the hardware facilitating it), the main difference between the two conditions was that the iVR students experienced immersion, and a higher sense of agency/control.

Data gathering tools (tests and questionnaires)

The knowledge-related pre- and post-tests for each topic consisted of nine multiple-choice questions (Appendix 2) for each of the three topics/interventions.

The standardised self-report affective measurement (Appendix 3) in the pre and post-tests included a 7-point Likert scale on intrinsic motivation (10 items; Monteiro, Mata, and Peixoto 2015), and self-efficacy (eight items; Makransky, Thisgaard, and Gadegaard 2016). The post-tests included 7-point Likert scales for two more affective aspects: Perceived learning (seven items; Lee, Wong, and Fung 2010), and iVR/video enjoyment (five items; Monteiro, Mata, and Peixoto 2015).

The tests were administered using SurveyXact (Rambøll 2014).

Experimental protocol

The quasi-experiment began on the 1st day with briefing the students about the details of the course, the teachers involved, and a general schedule of events. The students were also briefed about the schedule of the experimental elements of the course (e.g. pretest >>> division into treatment groups >>> interaction with the respective virtual learning tool and so on). Each student then read and signed a consent form. The consent form also randomly assigned each student a code that indicated their treatment condition (iVR or video). Each student would use this code throughout the study while responding to the data-collection tools.

After that, the students were administered a pre-test comprising knowledge-related questions on all the three topics (i.e. 27 multiple-choice questions in total), followed by the affective questionnaire on intrinsic motivation and self-efficacy. On completion of the pre-test, the iVR group was taken to another room equipped with ready-to-use HMDs (with earphones), one for each student. The iVR group received a brief pre-training on how to use the equipment (e.g. wearing the HMD, functions of the controller buttons). As the iVR group interacted with the 1st simulation (Pigment Extraction) in this other room, the students in the video group watched its respective video on a laptop. The video was made available via a link. Two researchers in each room monitored the students. Each student, on completion of the iVR or video intervention, answered post-test 1, containing questions on pigment extraction, and affective measures (intrinsic motivation, self-efficacy, perceived learning and enjoyment).

A similar protocol was followed on the 2nd and the 4th days in relation to the food webs and biodiversity topics/simulations (and the corresponding post-tests 2 and 3) respectively. Except, there was no pre-test on these days. In addition, the students also took part in common lectures, field trips and laboratory work related to the respective topic on that day. On the 5th day, students took the final post-test (all the 27 questions related to the three topics combined).

Finally, the students were administered the longitudinal post-test, approximately 2 months after the course, to measure if there were any long-term effects of the interventions.

A close collaboration with the course teacher (last author) ensured a smooth and cohesive integration of the experimental material and protocol with the regular course activities.

Analysis

Due to a technical failure during the final (biodiversity) iVR simulation, all the data related to the third intervention/topic were omitted from the analysis. One student from the iVR group did not participate in the study after the first intervention due to cyber-sickness, and three students from the video group failed to participate in at least one of the interventions. Hence, our effective sample for the data analysis included 24 participants (13 in the iVR group, and 11 in the video group). Further, the longitudinal post-test unfortunately recorded low response rate, with only eight iVR participants and nine video participants responding to all the questions.

We were interested to understand if and how interaction with multiple iVR simulations changes the cognitive and affective aspects of student learning over time, in comparison to watching videos of those simulations. In consideration of our limited effective sample size, our use of statistics to report the results is for descriptive purposes only. We do not present any significance-related claims.

The graphical data presentations discussed in the results section were made using GraphPad Prism 8.3.1 (San Diego, USA). Data on knowledge trends are presented as means ± standard deviation (SD). Data on affective endpoints (self-efficacy, intrinsic motivation, perceived learning and enjoyment) are reported on the 7-point Likert scale. Frequency is presented as a cumulative percentage of students reporting the individual score and calculated according to the equation:

Cumulative percent number of students giving score x = (frequency of students reporting the score x / total sum of all Likert scores) × 100%, where x is equal to the Likert score 1–7.

Theoretical expectations

Based on the literature, particularly in relation to the novelty effect (e.g. Chen *et al*. 2016; Moos and Marroquin 2010), the following trends were expected in knowledge-related and affective outcomes among our participants over time.

Knowledge trends

We expected both within- and between-group differences to emerge as the study progressed.

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

Within-group trends through the checkpoints

In comparison to their respective baseline scores, the scores for both groups were expected to improve in post-test 1, perhaps due to sensitisation, priming and/or learning (Lana 2009; Willson and Putnam 1982). The scores for the iVR group were then expected to improve further thereafter through to the final post-test, but drop for the video group after post-test 1 due to lack of interaction in the video-viewing process (Ryan and Deci 2000). Finally, the scores in the longitudinal post-test were expected to remain unchanged for the iVR group but to drop for the video group (this is clari-fied further in the 'between-group differences section').

Between-group differences

The gain after the first intervention was expected (possibly due to priming and/or learning) to be similar for both the groups. However, after the second intervention, the scores for the iVR group were expected to remain increasingly higher than the video group, as the novelty effect among iVR students, if any, would fade away due to increasing familiarity with the immersive environment (Chen *et al*. 2016). In parallel, the video group would gradually lose interest due to lack of sense of agency in learning (Ryan and Deci 2000).

Affective trends

The affective expectations are also discussed as within- as well as between-group comparisons.

Within-group trends through the checkpoints

A gain in the affective outcomes was expected for both groups, attaining a peak right after the first intervention (i.e. in post-test 1). We expected a drop in student ratings across all the affective measurement scales after the second intervention (i.e. in post-test 2) for both groups.

Hidi and Renninger's four-phase model of interest development seems particularly relevant in this context (Hidi and Renninger 2006; Renninger and Hidi 2015). It predicts that students' interest is initially triggered by extrinsic elements in learning environments (in this case video or iVR) through interaction. With continued exposure to the triggering environments, the students enter a second phase of maintained situational interest, which may lead to an emergence of individual interest (phase 3) and/or well-developed individual interest (phase 4). While examining student interest was not particularly intended in this study, we use Schiefele's definition of interest as content-specific intrinsic motivation (Schiefele 1991). Further, considering the strong links with intrinsic motivation, similar trends were expected for self-efficacy, perceived learning and enjoyment for both the groups (Bandura 1997; Wigfield and Eccles 2002).

Between-group comparison

Due to the novelty effect, gains in the affective measures for the iVR group were expected to be larger in magnitude than those for the video group after the first

intervention. We did not expect students to experience the novelty effect while watching videos of simulations.

Further, a bigger drop in student ratings across all the affective measures was expected for the video group, as these students would lack an experience or sense of agency (e.g. making choices and experiencing relatedness) as opposed to students in the iVR group (Ryan and Deci 2000).

Results

Knowledge trends

In terms of the cognitive aspects (i.e. knowledge performance), the pre-test served as a proxy for students' baseline understanding of the topics and ensured that pre-existing knowledge did not influence the results of the study. The video group's baseline knowledge performance (mean percent = 70.2% ± 16.7%) was higher than that of the iVR group (61.1% ± 16.3%).

Figure 2 shows performance trends in the knowledge-related tests captured at the different assessment checkpoints during the study. As expected for the knowledge-related within-group trend, the scores for both groups improved after the first intervention (as shown by difference between pre-test and post-test 1 scores). However, the knowledge gain immediately after the first intervention was higher for the iVR group (percent means: post-test1 = 77.8% ± 13.6%, post-test2 = 72.7% ± 19.0%, final post-test = 74.3% ± 13.9%) than for the video group (percent means: post-test1 = 76.8% ± 11.6%,

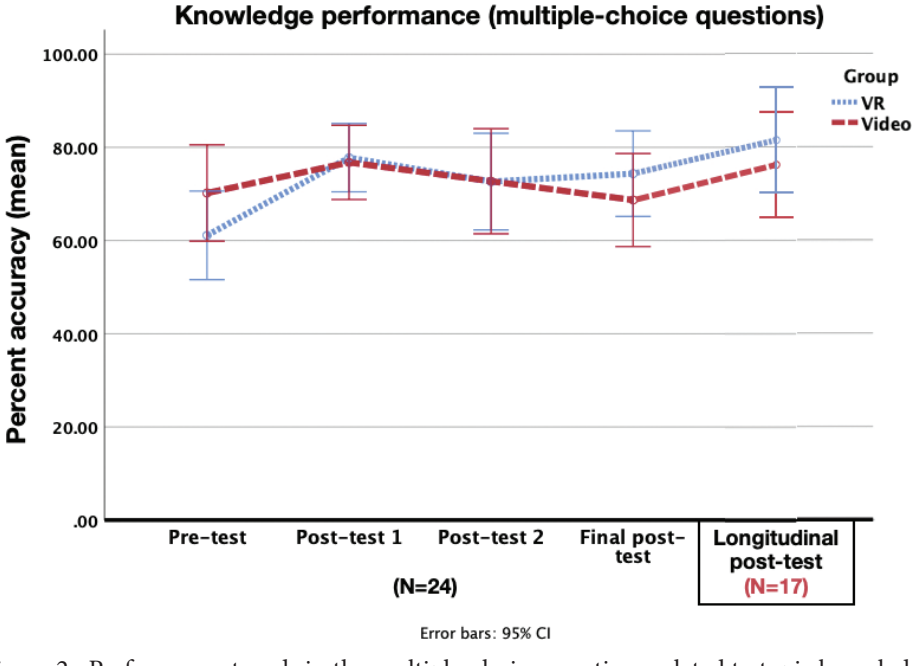

Figure 2. Performance trends in the multiple-choice questions related to topic knowledge (mean percent accuracy). There was an increase in knowledge retention for both groups. However, the increase for the iVR group (about 30% increase) was more than that for the Video group (less than 10%).

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

post-test2 = 72.7% ± 16.7%, final post-test = 68.7% ± 18.1%), which had negligible effect of the interventions on test performance. Scores for the iVR group, on the other hand, remained stable after the first intervention, maintaining the overall gain relative to the baseline performance. Further, performance in the longitudinal post-test (that was deployed 2-months after the intervention) suggested a long-term retention of this knowledge gain on average by participants in the iVR group (mean score = 81.2% ± 23.7%). In comparison, the longitudinal post-test scores for the video group revealed no conclusive pattern of performance (mean score = 76.4% ± 14.5%). In summary, there was an increase in knowledge performance for both groups, especially for the iVR group (nearly 30% increase for iVR group vs. less than 10% for the video group).

Affective trends

Intrinsic motivation

Student intrinsic motivation in both groups remained consistently high (considerably better than chance) through the entire course to the longitudinal post-test (means iVR: baseline = 5.9 ± 0.8, post-test1 = 5.8 ± 0.8, post-test2 = 5.6 ± 1.1, final post-test = 5.9 ± 1.1, longitudinal post-test = 6.0 ± 1.45; means video: baseline = 5.7 ± 1.3, post-test1 = 5.7 ± 0.9, post-test2 = 5.9 ± 0.9, final post-test = 5.8 ± 1.4, longitudinal post-test = 6.1 ± 1.69). Figure 3 shows the proportion of participants choosing a particular rating on the intrinsic motivation scale. The proportion of participants rating the highest possible score on the scale increased among both groups as the course progressed to the final post-test (iVR: pre-test = 28%, post-test1 = 36%, post-test2 = 39%, final post-test = 47%; video: pre-test = 32%, post-test1 = 34%, post-test2 = 34%, final post-test = 50%). In the longitudinal post-test, many participants chose the highest rating of 7 (iVR = 53%; video = 65%) while most reported their motivation as high (~ rating 5) or above (iVR = 91%; video = 82%). However, there were no

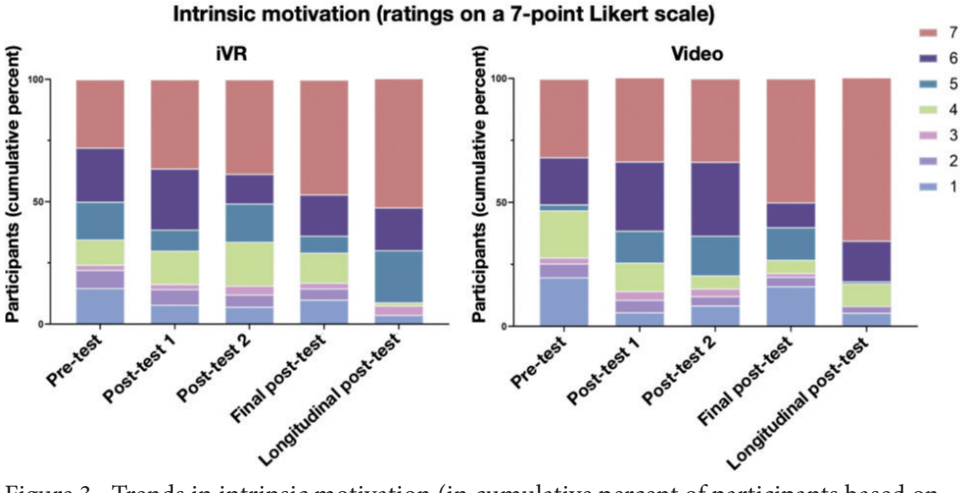

Figure 3. Trends in intrinsic motivation (in cumulative percent of participants based on their average ratings on the 7-point Likert scale). Intrinsic motivation remained high for both groups throughout the study, with a slight increase in the proportion of students scoring 7, and a decrease in the proportion of students scoring 1 on the Likert scale.

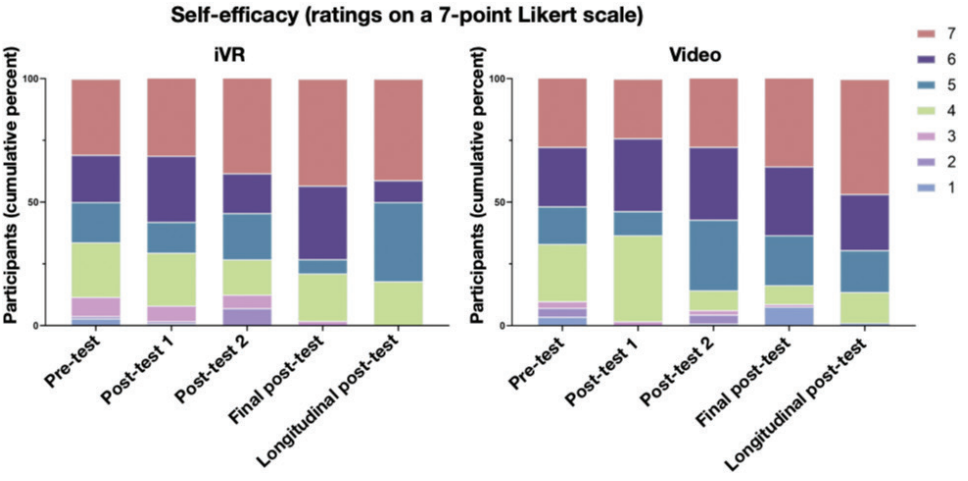

Figure 4. Trends in self-efficacy (in cumulative percent of participants choosing a particular rating on the 7-point Likert scale). Self-efficacy remained high for both groups throughout the study, with a slight increase in the proportion of students scoring 7 on the Likert scale.

between-group differences in the intrinsic motivation over time. In summary, intrinsic motivation remained high for both groups throughout the study, with a slight, but comparable increase for both groups in the proportion of students scoring 7, and a decrease in the proportion of students scoring 1, on the Likert scale.

Self-efficacy

Both groups reported high self-efficacy throughout the entire course (means iVR: baseline = 5.2 ± 1.3, post-test1 = 5.3 ± 1.2, post-test2 = 5.3 ± 1.6, final post-test = 5.8 ± 1.1, longitudinal post-test = 5.7 ± 1.2; means video: baseline = 5.2 ± 1.2, post-test1 = 5.3 ± 1.1, post-test2 = 5.5 ± 1.3, final post-test = 5.6 ± 1.6, longitudinal post-test = 6.0 ± 1.2). Figure 4 shows the proportion of participants with respect to the ratings they chose on the self-efficacy scale. Through the course, the proportion of participants rating their self-efficacy as high (~ rating 5 or more) steadily increased among both groups after the first intervention in comparison to their respective baseline/pre-test scores (iVR: pre-test = 66%, post-test1 = 71%, post-test2 = 73%, final post-test = 79%, longitudinal post-test = 82%; video: pre-test = 67%, post-test1 = 63%, post-test2 = 86%, final post-test = 84%, longitudinal post-test = 86%).

In summary, self-efficacy remained high for both groups throughout the study, with a slight, but comparable increase for both groups in the proportion of students scoring 7 on the Likert scale.

Perceived learning

The average perception of having learnt through the intervention increased considerably and steadily among students in both groups over time as the course progressed (means iVR: post-test1 = 4.7 ± 1.6, post-test2 = 5.2 ± 1.4, final post-test = 6.0 ± 1.0,

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

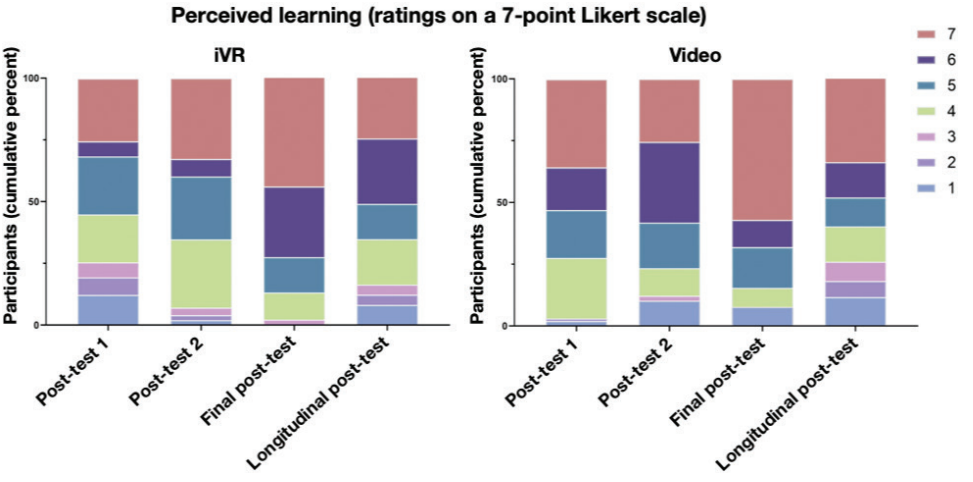

Figure 5. Trends in perceived learning (in cumulative percent of participants choosing a particular rating on the 7-point Likert scale). Perceived learning was similar between the beginning and the end of the study (longitudinally). However, there was an increase in perceived learning immediately after the interventions (final post-test) for both groups (about 50% increase in the number of students scoring 7 on the Likert scale).

longitudinal post-test = 5.0 ± 1.8; means video: post-test1 = 5.5 ± 1.2, post-test2 = 5.2 ± 1.8, final post-test = 5.9 ± 1.7, longitudinal post-test = 4.9 ± 2.1). Figure 5 shows the proportion of participants with respect to their choice of ratings on the perceived learning scale. It can be seen from the figure that the proportion of participants rating as high (~ rating 5 or more) on the scale gradually increased for both groups (iVR: post-test1 = 55%, post-test2 = 65%, final post-test = 87%; video: post-test1 = 72%, post-test2 = 77%, final post-test = 85%). This proportion, however, decreased for both groups longitudinally (iVR = 65%; video = 60%).

In summary, perceived learning was similar in the beginning and at the end of the study (longitudinally). However, there were changes during the study, with an increase in perceived learning immediately after the interventions (final post-test) for both groups (i.e. about 50% increase in number of students scoring 7 on the Likert scale).

Enjoyment

Finally, while the change in enjoyment of the respective intervention was negligible for both groups over time, students in the iVR group enjoyed the interventions more (means: post-test1 = 4.7 ± 1.9, post-test2 = 4.2 ± 1.5, final post-test = 4.1 ± 1.9, longitudinal post-test = 4.5 ± 2.1). In contrast, students in the video group largely reported boredom, as indicated by their overall below-4 ratings (means: post-test 1 = 3.8 ± 1.6, post-test2 = 3.4 ± 1.5, final post-test = 3.2 ± 1.5, longitudinal post-test = 4.6 ± 2.5). Interestingly, the number of students scoring 7 on the Likert scale increases dramatically for the Video group through to the final long-term test. However, the proportion of students reporting high enjoyment (ratings of 5 and above on a 7-point Liker scale) often remained 50% or less for the iVR group (Figure 6),

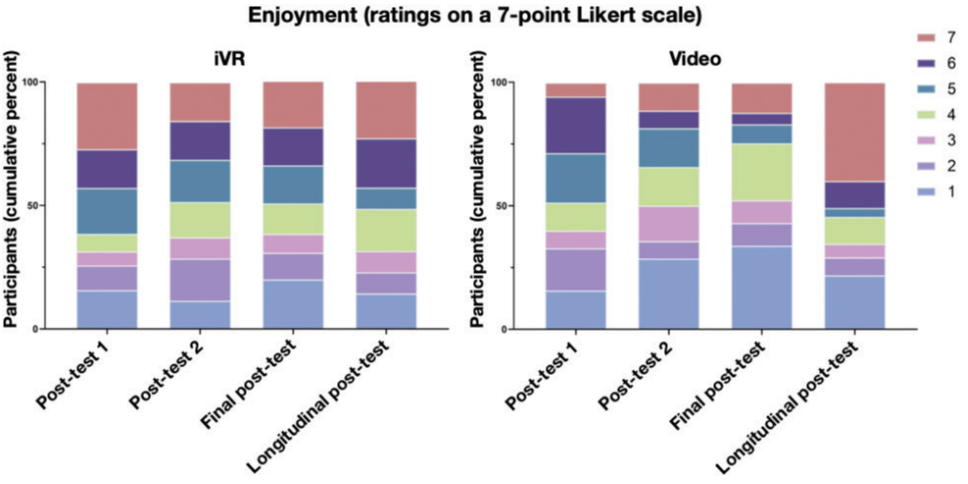

Figure 6. Trends in (intervention) enjoyment (in cumulative percent of participants choosing a particular rating on the 7-point Likert scale). There was a considerable increase in the number of students scoring 7 on the Likert scale for the Video group. The mean score for enjoyment remained the same throughout the study for both groups.

including in the longitudinal post-test (iVR: post-test1 = 61%, post-test2 = 49%, final post-test = 49%, longitudinal post-test = 52%). For the video group, the proportion of students reporting ratings of 5 and above decreased (post-test1 = 49%, post-test2 = 34%, final post-test = 25%, longitudinal post-test = 55%); inversely, the proportion of students reporting boredom (i.e. ratings of 4 or less) increased considerably through the different checkpoints during the course before increasing again after the 2-month delay (video: post-test1= 51%, post-test2 =6 6%, final post-test = 75%, longitudinal post-test = 45%). In summary, though there was a considerable increase in the number of students in the video group scoring 7 on the Likert scale, the mean score remained the same throughout the study for both groups.

Summary and discussion

The primary objective of this study was to investigate if and how interaction with multiple iVR field and laboratory simulations, and their video playbacks, affects student learning over time. Overall trends observed in student performance on multiple-choice knowledge tests indicate that multiple exposures of iVR simulations result in a considerable gain in topic-knowledge over time, as compared to watching videos of those simulations. Student performance in the longitudinal post-test indicates that interacting with iVR simulations is more effective in supporting long-term retention of knowledge (Figure 1). This is among the first evidences to indicate positive longitudinal knowledge-related learning outcomes of multiple exposures of iVR (Mikropoulos and Natsis 2011; Southgate 2020). These results are consistent with much of the cross-sectional and longitudinal work investigating how virtual interventions (e.g. virtual simulations, games, e-learning) affect learning in higher education settings over time (Gaupp, Drazic, and Körner 2019; Hanus and Fox 2015; Madani *et al*. 2014, 2016).

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

Although the affective results do not show any conclusive cross-sectional or longitudinal between-group differences for intrinsic motivation and self-efficacy, students in both groups were highly motivated and self-confident about working on the topics (larger proportions of overall ratings above 5 on a 7-point Likert scale; Figures 2 and 3). This indicates that both interventions were successful in triggering student interest (Makransky *et al.* 2020). The gradual increase in the proportion of students in both groups who choose higher ratings on the respective scales further indicates that multiple exposures of the respective interventions helped in maintaining this interest among students over time (Hidi and Renninger 2006; Renninger and Hidi 2015; Wigfield and Eccles 2002).

Concerning perceived learning, the proportion of students who expressed a positive feeling about learning the topics through the simulations increased during the course for both groups. However, there was a drop in this proportion longitudinally (Figure 4), suggesting that the perceived learning benefits of both interventions faded eventually. This result can be easily explained for the video-group, where students experienced little control over (and consequently, little active involvement in) the various aspects of the intervention (e.g. due to passive viewing). These students were more likely to reflect on the video-viewing process as a not so beneficial activity (Lee, Wong, and Fung 2010; Makransky and Lilleholt 2018). However, it is not clear why even iVR students did not perceive the intervention to be beneficial in the long term. It is possible that their responses to this scale were negatively influenced (knowingly or unknowingly) by perceived difficulty of the knowledge-related questions in the longitudinal test (e.g. cognitive benefit; Makransky and Lilleholt 2018). In other words, if students perceived the knowledge-related questions to be more difficult in the longitudinal test than during the course due to the delay in testing, they were more likely to choose low ratings on the perceived learning scale in the longitudinal test. In every test, the affective questionnaires were presented after the knowledge-related multiple-choice questions. Students, thus, did have some opportunity to reflect on their performance on the knowledge test before recording their affective responses. A decline in the feeling of having experienced a novel intervention (novelty effect) over time could also explain why there was a longitudinal drop in the perceived benefits of iVR instruction (Tokunaga 2013).

Interestingly, the data trends on the affective outcomes did not provide any direct evidence of the novelty effect among students in the iVR group (we neither expected nor found this effect for video intervention). This could be because students in our iVR group were relatively more familiar and/or used to immersive technologies (e.g. through gaming, museum visits or previous exposure at other courses; Renninger and Hidi 2015). However, given that immersive interfaces are becoming increasingly accessible due to their cheap cost and availability for individual use in different forms (Radianti *et al.* 2020), future studies may not need to consider the novelty factor depending on the population/sample demographics.

Further, iVR was relatively more enjoyable than the (non-interactive/passive) video-viewing intervention, which was rated as 'boring' by an increasing proportion of participants during the course (Bailenson *et al.* 2008). However, even for iVR, most students rated their enjoyment of the immersive experience to be below 5 on a 7-point Likert scale throughout the different checkpoints (Figure 5). This indicates that they only marginally enjoyed the iVR simulations and/or did not like some aspects of it.

Lessons from educational technology design studies have strongly suggested how various cognitive and affective learning outcomes are determined by, or (at the least are)

dependent largely upon, the nature of iVR interface design (Pekrun 200. There is a great diversity in the designs of iVR learning environments, as well as the design principles, theories inspiring them (Wu, Yu, and Gu 2020). While the iVR simulations used in our study are interactive, a user can be said to experience, for instance, limited control over the interface elements and the sequence of interaction with those elements. These simulations have a pre-determined and strict sequence of events (e.g. science laboratory/experimental protocol) on which a user has no control. Moreover, navigation in the simulations is based on a system utilising only three degrees-of-freedom. This is actualised as a mouse-like point-and-click interaction in a rather detailed and immersive virtual environment. These aspects could explain the observed affective outcome trends for the iVR group in our study. Several structural equation modeling approaches have suggested that the extent to which a learner experiences control over various elements of an interface predicts some important affective learning outcomes, such as perceived learning, and enjoyment (Pekrun and Stephens 2010; Plass and Kaplan 2016).

Conclusion and implications

Our study presents promising results on the cross-sectional and longitudinal effects of multiple exposures of different iVR simulations for the learning of different topics in environmental biology. The study also indicated positive effects of multiple iVR interventions on perceived learning and enjoyment. These learning effects of iVR were presented in comparison with the learning effects of desktop-based video-playbacks. The strengths of our study and overall methodological approach are: (1) the ecological implementation of the experiment involving a teacher-supported systematic integration of multiple iVR simulations and their videos in a regular field course without disrupting other course elements (Merchant *et al*. 2014), and (2) the longitudinal nature; comprising multiple, immediate as well as delayed data-collection checkpoints.

Incorporating these strong dimensions in our approach, however, acted as a double-edged sword. For instance, to maintain ecological validity and fidelity of implementation, it was necessary to sacrifice experimental control during the study (e.g. control for diffusion of treatments). Further, the use of multiple data-collection checkpoints and the inability to capture if and how this influenced the data also turned out to be a limitation of our study. Multiple testing checkpoints involved asking the same questions to students several times. While verbal-data related to student experiences were not collected in this study, exhaustion and boredom were observed among many students, as they also felt increasingly 'annoyed', particularly during the later tests (i.e. post-tests deployed towards the end of the course). In addition, some students also perceived the integration of technology and the accompanying multiple tests in the course pedagogy as 'additional work', corroboratively indicating a growing disengagement among students from the interventions over time.

Though such cross-sectional/longitudinal experimental designs have a clear advantage in capturing micro as well as long-term changes in student cognitive outcomes, future studies are strongly recommended to consider alternative ways of capturing/assessing these changes. It may help to advise students/participants, in such studies, that repeated testing might increase their understanding of the subject matter and eventually improve their overall performance in formal university assessments. Finally, for future ecological studies as well as iVR-integrated pedagogies to be more

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

successful, advanced orientation sessions for students may be needed to better prepare them for such technology-enriched experiences alongside other pedagogical elements.

Nevertheless, our study is one of the first and few works investigating the longitudinal effects of using iVR simulations on various cognitive and affective aspects of learning compared to watching pre-recorded videos of those simulations.

Despite our low sample size, and weak but indicative results on longitudinal learning outcome trends, our study strongly demonstrates the necessity of such investigative approaches to better understand if and how innovative iVR simulations for higher education could be designed and implemented in practice. To better test the novelty effect hypothesis, it may help to include the Focused Attention and Presence scale in future investigations.

Competing interests

All the authors declare that they have no competing interests with regards to this manuscript.

Notes on contributors

All the authors (alphabetically arranged initials – AES, AT, BM, MEM, PMJ and PP) collectively designed and conducted the experiment at the site. PMJ was the coordinating teacher for the environmental biology course, where the experiment was conducted. AES, AT, BM, MEM and PP carried out data analysis independently as well as collaboratively. PP wrote the manuscript. All the other authors critically reviewed and edited the manuscript. PP co-supervised the study.

Acknowledgment

We sincerely thank the anonymous reviewers for their constructive comments on the manuscript. We are indebted to Søren Larsen for his administrative support in planning and conducting this study. We also thank Simon Warren and Guido Makransky for their valuable feedback on the research design during the early stages of the study. This work was possible with the support of an extraordinary grant received from the Danish Ministry of Education and Science.

References

Andreasen, N. K., *et al*., (2019) 'Virtual reality instruction followed by enactment can increase procedural knowledge in a science lesson', *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),* IEEE, Osaka, Japan, pp. 840–841. doi: 10.1109/VR.2019.8797755

Ba, R., *et al*., (2019) 'Virtual reality enzymes: an interdisciplinary and international project towards an inquiry-based pedagogy', in *VR, Simulations and Serious Games for Education*, eds Y. Cai, W. V. Joolinger & Z. Walker, Springer, Singapore, pp. 45–54. doi: 10.1007/978-981-13-2844-2_5

Bailenson, J., *et al*., (2008) 'The effect of interactivity on learning physical actions in virtual reality', *Media Psychology*, vol. 11, no. 3, pp. 354–376. doi: 10.1080/15213260802285214

Bandura, A. (1997) *Self-Efficacy: The Exercise of Control*, Freeman, New York, NY. doi: 10.5860/choice.35-1826

Checa, D. & Bustillo, A. (2020) 'A review of immersive virtual reality serious games to enhance learning and training', *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 5501–5527. doi: 10.1007/s11042-019-08348-9

Chen, J. A., *et al*., (2016) 'A multi-user virtual environment to support students' self-efficacy and interest in science: a latent growth model analysis', *Learning and Instruction*, vol. 41, no. C, pp. 11–22. doi: 10.1016/j.learninstruc.2015.09.007

Chittaro, L. & Buttussi, F. (2015) 'Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety', *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 4, pp. 529–538. doi: 10.1109/TVCG.2015.2391853

Cochrane, T. (2016) 'Mobile VR in education: from the fringe to the mainstream', *International Journal of Mobile and Blended Learning (IJMBL)*, vol. 8, no. 4, pp. 44–60. doi: 10.4018/IJMBL.2016100104

Concannon, B. J., Esmail, S. & Roberts, M. R. (2019) 'Head-mounted display virtual reality in post-secondary education and skill training: a systematic review', *Frontiers in Education,* vol. 4, p. 80. doi: 10.3389/feduc.2019.00080

Fiorella, L. & Mayer, R. E. (2016) 'Eight ways to promote generative learning', *Educational Psychology Review,* vol. 28, no. 4, pp. 717–741. doi: 10.1007/s10648-015-9348-9

Garcia-Bonete, M. J., Jensen, M. & Katona, G. (2019) 'A practical guide to developing virtual and augmented reality exercises for teaching structural biology', *Biochemistry and Molecular Biology Education*, vol. 47, no. 1, pp. 16–24. doi: 10.1002/bmb.21188

Gaupp, R., Drazic, I. & Körner, M. (2019) 'Long-term effects of an e-learning course on patient safety: a controlled longitudinal study with medical students', *PLoS One*, vol. 14, no. 1, p. e0210947. doi: 10.1371/journal.pone.0210947

Gutierrez-Maldonado, J., Andres-Pueyo, A. & Talarn-Caparros, A. (2015) 'Virtual reality to train teachers in ADHD detection', *Society for Information Technology & Teacher Education International Conference*, Association for the Advancement of Computing in Education (AACE), Las Vegas, USA, pp. 769–772.

Hanus, M. D. & Fox, J. (2015) 'Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance', *Computers & Education*, vol. 80, pp. 152–161. doi: 10.1016/j.compedu.2014.08.019

Hew, K. F. & Cheung, W. S. (2010) 'Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: a review of the research', *British Journal of Educational Technology*, vol. 41, no. 1, pp. 33–55. doi: 10.1111/j.1467-8535.2008.00900

Hidi, S. & Renninger, K. A. (2006) 'The four-phase model of interest development', *Educational Psychologist*, vol. 41, no. 2, pp. 111–127. doi: 10.1207/s15326985ep4102_4

Jensen, L. & Konradsen, F. (2018) 'A review of the use of virtual reality head-mounted displays in education and training', *Education and Information Technologies*, vol. 23, no. 4, pp. 1515–1529. doi: 10.1007/s10639-017-9676-0

Kafai, Y. B. & Dede, C. (2014) 'Learning in virtual worlds', in *The Cambridge Handbook of the Learning Sciences*, ed R. K. Sawyer, Cambridge University Press, Chapel Hill, USA, pp. 522–544.

Kavanagh, S., *et al*., (2017) 'A systematic review of Virtual Reality in education', *Themes in Science and Technology Education*, vol. 10, no. 2, pp. 85–119.

Khor, W. S., *et al*., (2016) 'Augmented and virtual reality in surgery – the digital surgical environment: applications, limitations and legal pitfalls', *Annals of Translational Medicine*, vol. 4, no. 23, p. 423. doi: 10.21037/atm.2016.12.23

Kolb, D. A. (1984) *Experiential Learning: Experience as the Source of Learning and Development*, FT Press.

Labster. (2019a) *Pigment Extraction: Use Photosynthesis to Produce Biofuel and Reduce Pollution*, [online] Available at: https://www.labster.com/simulations/pigment-extraction/

Labster. (2019b) *Food Webs: Learn about Interactions between Trophic Levels*, [online] Available at: https://www.labster.com/simulations/food-webs/

Labster. (2019c) *Biodiversity: Assess and Compare Biodiversity on an Exoplanet Virtual Lab Simulation*, [online] Available at: https://www.labster.com/simulations/biodiversity/

Lamb, R., *et al*., (2018) 'Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy', *Computers & Education*, vol. 124, pp. 14–26. doi: 10.1016/j.compedu.2018.05.014

Lana, R. E. (2009) 'Pretest Sensitization', in *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*, eds R. Rosenthal & R. L. Rosnow, Oxford University Press, New York, USA, p. 93.

Lee, E. A. L., Wong, K. W. & Fung, C. C. (2010) 'How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach', *Computers & Education*, vol. 55, no. 4, pp. 1424–1442. doi: 10.1016/j.compedu.2010.06.006

Liagkou, V., Salmas, D. & Stylios, C. (2019) 'Realizing virtual reality learning environment for industry 4.0', *Procedia CIRP*, vol. 79, pp. 712–717. doi: 10.1016/j.procir.2019.02.025

Madani, A., *et al*., (2014) 'Impact of a hands-on component on learning in the Fundamental Use of Surgical EnergyTM (FUSE) curriculum: a randomized-controlled trial in surgical trainees', *Surgical Endoscopy*, vol. 28, no. 10, pp. 2772–2782. doi: 10.1007/s00464-014-3544-4

Madani, A., *et al*., (2016) 'Long-term knowledge retention following simulation-based training for electrosurgical safety: 1-year follow-up of a randomized controlled trial', *Surgical Endoscopy*, vol. 30, no. 3, pp. 1156–1163. doi: 10.1007/s00464-015-4320-9

Madden, J., *et al*., (2020) 'Ready student one: exploring the predictors of student learning in virtual reality', *PLoS One*, vol. 15, no. 3, p. e0229788.

Makransky, G., *et al*., (2020) 'Immersive Virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive Virtual reality', *Journal of Educational Psychology*. doi: 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*, vol. 35, no. 6, pp. 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*, vol. 66, no. 5, pp. 1141–1164. doi: 10.1007/s11423-018-9581-2

Makransky, G., Terkildsen, T. & Mayer, R. (2019) 'Adding immersive virtual reality to a science lab simulation causes more presence but less learning', *Learning and Instruction*, vol. 60, pp. 225–236. doi: 10.1016/j.learninstruc.2017.12.007

Makransky, G., Thisgaard, M. & Gadegaard, H. (2016) 'Virtual simulations as preparation for lab exercises: assessing learning of key laboratory skills in microbiology and improvement of essential non-cognitive skills', *PLoS One*, vol. 11, no. 6, p. e0155895. doi: 10.1371/journal.pone.0155895

Merchant, Z., *et al*., (2014) 'Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: a meta-analysis', *Computers & Education*, vol. 70, no. C, pp. 29–40. doi: 10.1016/j.compedu.2013.07.033

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*, vol. 140, p. 103603. doi: 10.1016/j.compedu.2019.103603

Mikropoulos, T. & Natsis, A. (2011) 'Educational virtual environments: a ten-year review of empirical research (1999–2009)', *Computers & Education*, vol. 56, no. 3, pp. 769–780. doi: 10.1016/j.compedu.2010.10.020

Monteiro, V., Mata, L. & Peixoto, F. (2015) 'Intrinsic motivation inventory: psychometric properties in the context of first language and mathematics learning', *Psicologia: Reflexão e Crítica*, vol. 28, no. 3, pp. 434–443. doi: 10.1590/1678-7153.201528302

Moos, D. & Marroquin, E. (2010) 'Multimedia, hypermedia, and hypertext: motivation considered and reconsidered', *Computers in Human Behavior*, vol. 26, no. 3, pp. 265–276. doi: 10.1016/j.chb.2009.11.004

Moro, C., Štromberga, Z. & Stirling, A. (2017) 'Virtualisation devices for student learning: comparison between desktop-based (Oculus Rift) and mobile-based (Gear VR) virtual reality in medical and health science education', *Australasian Journal of Educational Technology*, vol. 33, no. 6. doi: 10.14742/ajet.3840

OECD. (2018) Ruth Benander, EDU/EDPC(2018)45/ANN3, 'Future of education and skills 2030: conceptual learning framework', in *A Literature Summary for Research on the Transfer of Learning 8th Informal Working Group (IWG) Meeting,* [online] Available at: https://www.oecd.org/education/2030/A-Literature-Summary-for-Research-on-the-Transfer-of-Learning.pdf

Pande, P. & Chandrasekharan, S. (2017) 'Representational competence: towards a distributed and embodied cognition account', *Studies in Science Education*, vol. 53, no. 1, pp. 1–43. doi: 10.1080/03057267.2017.1248627

Pande, P. (2020) 'Learning and expertise with scientific external representations: an embodied and extended cognition model', *Phenomenology and the Cognitive Sciences,* pp. 1–20. doi: 10.1007/s11097-020-09686-y

Parong, J. & Mayer, R. E. (2018) 'Learning science in immersive virtual reality', *Journal of Educational Psychology*, vol. 110, no. 6, p. 785. doi: 10.1037/edu0000241

Pekrun, R. (2006) 'The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice', *Educational Psychology Review,* vol. 18, no. 4, pp. 315–341. doi: 10.1007/s10648-006-9029-9

Pekrun, R. & Stephens, E. J. (2010) 'Achievement emotions: a control-value approach', *Social and Personality Psychology Compass,* vol. 4, no. 4, pp. 238–255. doi: 10.1111/j.1751-9004.2010.00259.x

Plass, J. L. & Kaplan, U. (2016) 'Emotional design in digital media for learning', in *Emotions, Technology, Design, and Learning*, eds S. Y. Tettegah & M. Gartmeier, Elsevier, Amsterdam, pp. 131–161. doi: 10.1016/B978-0-12-801856-9.00007-4

Radianti, J., *et al*., (2020) 'A systematic review of immersive virtual reality applications for higher education: design elements, lessons learned, and research agenda', *Computers & Education*, vol. 147, Article 103778. doi: 10.1016/j.compedu.2019.103778

Rambøll Management Consulting. (2014). *SurveyXact*. https://www.surveyxact.com/

Renninger, K. A. & Hidi, S. (2015) *The Power of Interest for Motivation and Engagement*, Routledge, New York, USA. doi: 10.4324/9781315771045

Ryan, R. M. & Deci, E. L. (2000) 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being', *American Psychologist*, vol. 55, no. 1, p. 68. doi: 10.1037/0003-066X.55.1.68

Sánchez, J., Lumbreras, M. & Silva, J. (1997) 'Virtual reality and learning: trends and issues', *Proceedings of the 14th International Conference on Technology and Education*, Oslo, Norway, pp. 10–13.

Schiefele, U. (1991) 'Interest, learning, and motivation', *Educational Psychologist*, vol. 26, no. 3–4, pp. 299–323. doi: 10.1080/00461520.1991.9653136

Schott, C. & Marshall, S. (2018) 'Virtual reality and situated experiential education: a conceptualization and exploratory trial', *Journal of Computer Assisted Learning*, vol. 34, no. 6, pp. 843–852. doi: 10.1111/jcal.12293

Southgate, E. (2020) *Virtual Reality in Curriculum and Pedagogy: Evidence from Secondary Classrooms,* Routledge, New York, USA. doi: 10.4324/9780429291982

Tan, S. & Waugh, R. (2013) 'Use of virtual-reality in teaching and learning molecular biology', in *3D Immersive and Interactive Learning*, ed Y. Cai, Springer, Singapore, pp. 17–43. doi: 10.1007/978-981-4021-90-6_2

Tokunaga, R. S. (2013) 'Engagement with novel virtual environments: the role of perceived novelty and flow in the development of the deficient self-regulation of Internet use and media habits', *Human Communication Research*, vol. 39, pp. 365–393. doi: 10.1111/hcre.12008

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

Torff, B. & Tirotta, R. (2010) 'Interactive whiteboards produce small gains in elementary students' self-reported motivation in mathematics', *Computers & Education*, vol. 54, no. 2, pp. 379–383. doi: 10.1016/j.compedu.2009.08.019

Vergara, D., *et al*., (2019) 'On the importance of the design of virtual reality learning environments', *International Conference in Methodologies and Intelligent Systems for Technology Enhanced Learning,* Springer, Cham, pp. 146–152. doi: 10.1007/978-3-030-23990-9_18

Wigfield, A. & Eccles, J. S. (2002) 'The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence', in *Development of Achievement Motivation*, eds A. Wigfield & J. S. Eccles, Academic Press, New York, USA, pp. 91–120. doi: 10.1016/B978-012750053-9/50006-1

Willson, V. L. & Putnam, R. R. (1982) 'A meta-analysis of pretest sensitization effects in experimental design', *American Educational Research Journal*, vol. 19, no. 2, pp. 249–258.

Wu, B., Yu, X. & Gu, X. (2020) 'Effectiveness of immersive virtual reality using head-mounted displays on learning performance: a meta-analysis', *British Journal of Educational Technology*, vol. 21, no. 6, pp. 1991–2005. doi: 10.1111/bjet.13023

Appendix

Appendix 1 (Description of simulations)

Pigment extraction simulation

This simulation focuses primarily on the following learning objectives: understanding the importance of photosynthesis, properties of light and pigments as colors, data analysis of the pigment spectra and chemical properties of pigments. The main techniques demonstrated in this simulation are centrifugation, pigment extraction, and spectrophotometry. In this simulation, students can experience some aspects of these techniques while analysing a selected pigment (sample). During the simulation, the students are required to answer several questions in order to proceed further in the simulation.

Food webs simulation

This simulation focuses on helping students understand the concepts of food webs, trophic cascades and energy relations between the two (e.g. calculating the energy required to maintain cascades). The student engages in several interactive quizzes where they can change different parameters/effectors in a food web/chain, and gradually observe the impact of these changes, thereby gaining knowledge of the importance of various levels of a trophic cascade.

Biodiversity simulation

In this simulation, students are exposed to the following lab techniques: how to use quadrats, camera traps, pitfall traps, biodiversity index and elevational biodiversity gradients. Students are placed on an exoplanet where they can collect a sample for biodiversity identification. While doing so, they experience different sampling and analysis techniques, and answer various quizzes on species identification.

Appendix 2 (Affective questionnaire)

(Knowledge questionnaires – students were provided with four choices/options for each question. They had to choose the right answer from those options.)

Topic 1: Stream Ecology

1. Which of the following is not a pigment?
2. What gives the microalgae different colors?
3. Are there any advantages (for organisms) in having more than one type of pigment?
4. You want to extract chlorophyll from a water sample you have collected, how would you proceed (among the following four options)?
5. Which solvent, among the following, would you use for pigment extraction?
6. Which color light, among the following, do the chlorophyll pigments not absorb?
7. Why is it relevant to know chlorophyll concentration in natural waters?
8. You return from the field with water samples containing a very high concentration of chlorophylls. What does it tell you about the health of the ecosystem?
9. Which apparatus would you use to measure chlorophyll concentration?

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482

Topic 2: Lake ecology and food webs

1. What sources does an autotroph primary producer use to create biomass?
2. What, among the following, is the main difference between primary and secondary consumers?
3. Which of the following statements is true for FCE (food conversion efficiency) values?
4. Which term describes the change in trophic levels resulting from disturbances in other trophic levels?
5. Which term describes disturbance in the food web resulting from the introduction or removal of primary-producers in an ecosystem?
6. What is the purpose of using a respirometer in water research?
7. An ecosystem (e.g. Gulf of Mexico) has been disturbed (e.g. large oil spill). How long does it take this ecosystem to get back to the original state?
8. Which strategy will you apply to restore a eutrophic lake?
9. Which of the following organisms can give us a hint about the trophic state of a lake?

Topic 3: Fjords and biodiversity

1. Which term describes the occurrence of many species?
2. If there are many organisms biodiversity is:
3. Why would you use quadrats for sampling instead of sampling all specimens in the investigated sampling site?
4. How should quadrats be used in order to test biodiversity (to examine sessile organisms)
5. How does the density of organisms influence the optimal number of quadrats?
6. How are biodiversity data collected with different methods compared?
7. How are mobile terrestrial organisms best sampled quantitatively?
8. What is the purpose of using a dichotomous key?
9. What is important to consider when sampling biodiversity?

Appendix 3

For each of the following statements, please indicate on the scale of 1–7 how true it is for you; where 1 = not true at all, while 7 = Very true.

Intrinsic motivation (adapted from intrinsic motivation inventory – Monteiro, Mata, and Peixoto 2015)

1. I enjoy working with (name of the topic).
2. (Name of the topic) activities are fun to do.
3. (Name of the topic) is boring.
4. (Name of the topic) does not hold my attention at all.
5. I would describe (name of the topic) as very interesting.
6. I believe (name of the topic) could be of some value to me.

7. I think that working on (name of the topic) is useful for (name a skill critical to the topic).
8. I would be willing to work on (name of the topic) again because it has some value to me.
9. I think working on (name of the topic) is an important activity.
10. I believe this activity could be beneficial to me.

Self-efficacy (adapted from Makransky, Thisgaard, and Gadegaard 2016)

1. I am confident and can understand the basic concepts of (name of the topic).
2. I am confident that I understand the most complex concepts related to (name of the topic).
3. I am confident that I can do an excellent job on the assignments and tests in the (name of the topic) exercises.
4. I expect to do well in (name of the topic).
5. I am certain that I can master the skills being taught in (name of the topic).
6. I believe I will receive an excellent grade in (name of the topic).
7. I am certain I can understand the most difficult material presented in this course.
8. Considering the difficulty of the course, the teacher and my skills, I think I will do well in the class.

Perceived learning (adapted from Lee, Wong, and Fung 2010)

1. I was more interested to learn the different topics in (name of the topic).
2. I learned a lot of factual information in (name of the topic).
3. I gained a good understanding of the basic concepts in (name of the topic)
4. I learned a lot of practical skills in the topic of (name of the topic)
5. I was interested and stimulated to learn more.
6. I was able to summarise and conclude what I learned.
7. The activities were useful.

Intervention-specific enjoyment questions Intrinsic Motivation Inventory (IMI)

1. I thought the virtual laboratory case was quite enjoyable.
2. Virtual laboratory was fun to do.
3. The virtual laboratory case was boring.
4. The virtual laboratory case did not hold my attention at all.
5. I would describe the virtual laboratory case as very interesting.

Citation: Research in Learning Technology 2021, **29**: 2482 - http://dx.doi.org/10.25304/rlt.v29.2482