

RESEARCH ARTICLE

Open Access



CAVE-based immersive learning in undergraduate courses: examining the effect of group size and time of application

Tycho T. de Back* , Angelica M. Tinga and Max M. Louwerse

*Correspondence:

t.t.deback@uvt.nl

Department of Cognitive

Science & Artificial

Intelligence, Tilburg

University, Dante Building,

Warandelaan 2, 5037

AB Tilburg, The Netherlands

Abstract

Immersive virtual reality is increasingly regarded as a viable means to support learning. Cave Automatic Virtual Environments (CAVEs) support immersive learning in groups of learners, and is of potential interest for educational institutions searching for novel ways to bolster learning in their students. In previous work we have shown that the use of a CAVE-based virtual learning environment yielded higher learning gains compared to conventional textbook study. Yet, few prior studies have explored the circumstances that yield a trade-off between learning gains and the practical feasibility of providing immersive learning to large student numbers. To gain insight into these circumstances the current study examined two factors: (1) group size (small, medium and large), and (2) time of application (pre-, mid- and late-term of a course). Results indicated learning gains were present for all group sizes and application time periods, with highest learning gains in smaller groups. Learning gains were consistent across application time periods. Additionally, structural equation modeling was applied to assess how learning may result from the use of immersive virtual reality. Results indicated technological virtual reality features predicted learning outcomes via self-reported usability but less so via self-reported presence. Based on the findings, recommendations are presented for effective immersive learning for different group size and application time period configurations. Taken together, the current study elucidates factors affecting learning in immersive virtual reality and facilitates its use in educational practice.

Keywords: Virtual learning environments, Improving classroom teaching, Structural equation modeling, Virtual reality, Media in education

Introduction

How can immersive virtual reality (VR) be used in courses as a novel and efficient means of support? Different from conventional non-immersive 2D desktop VR, immersive VR places the viewer inside the virtual content. The immersive effect is further enhanced by reflecting natural body motions into the experience. Immersive VR is most commonly experienced using either of two display types. With VR headsets, the display is contained inside a device worn by the viewer. By contrast, Cave Automatic Virtual Environment (CAVE) mixed reality systems present virtual environments on the walls of a room, which serve as display surfaces (Cruz-Neira et al., 1992). The virtual environments

themselves are then viewed using see-through 3D glasses. Both VR headsets and CAVEs allow for benefits absent from non-immersive 2D desktops, such as stereoscopic 3D, as well as tracking of head and hand movements, allowing free and active interaction with virtual content (Slater & Sanchez-Vives, 2016). The added advantage of CAVE systems over VR headsets is that the latter typically isolate the user from the real surroundings, whereas users in mixed reality CAVEs retain view on the physical locality of the room, themselves, and other users inside it. Consequently, social body language cues are retained, allowing users to learn collaboratively (Birchfield & Megowan-Romanowicz, 2009). CAVEs thus constitute an interesting learning platform by having the potential to yield learning benefits and to do so for multiple users at a time.

When immersive VR is applied for educational purposes, one evident use case is its application to supplement traditional teaching practice. Examples in the literature of using immersive VR in addition to traditional teaching are varied and include topics such as engineering (Buñ et al., 2019; Fogarty et al., 2018; Halabi, 2020; Kamiska et al., 2019), language learning (O'Brien & Levy, 2008; Xie et al., 2019), legal education (McFaul & FitzGerald, 2020) and medical training (Huang et al., 2016; Maresky et al., 2019; Pelargos et al., 2017).

Background

Recent studies into the learning benefits of immersive VR have yielded mixed results (Makransky et al., 2017; Parong & Mayer, 2018) and this raises the question how and why learning may result from VR usage. A theoretical basis for examining the learning process in virtual learning environments (VLEs) may be found in the theoretical model of Lee et al. (2010), foremostly grounded in the framework of Salzman et al., (1999). Consistent with these frameworks, the model of Lee et al. (2010) asserts that technological VR features indirectly affect learning outcomes via a number of psychological factors, which will be detailed later. Using structural equation modeling (SEM), Lee et al. (2010) obtained an acceptable fit for the model in data obtained from students learning frog anatomy using an interactive 2D desktop simulation. The theoretical frameworks of Salzman et al. (1999) and Lee et al. (2010) have since served as starting points for several other SEM studies examining how learning arises from the use of VLE. These studies however predominantly focused on VLE of the non-immersive kind (Fokides, 2017; Fokides & Atsikpasi, 2018; Knutzen, 2019; Makransky & Petersen, 2019; Merchant et al., 2012). A rare exception is Makransky and Lilleholt (2018), who adapted the model of Lee et al. (2010) and used SEM to compare 2D desktop VR and immersive VR headset conditions. Yet, Makransky and Lilleholt (2018) only investigated affective outcomes and did not examine learning aspects. A lack of studies thus remains regarding the learning process when using immersive VR.

In a previous study we explored the potential of using a collaborative immersive VLE in a CAVE for yielding learning gains in students (De Back et al., 2020). The study was conducted in persons recruited from a subject pool, and compared learning gains in two conditions: (1) immersive CAVE learning, (2) conventional textbook learning. Results indicated the immersive CAVE condition induced learning gains, and exceeded those of the textbook condition. However, one can argue that the learning gains obtained can be attributed to the non-ecologically valid settings common in experiments: participants

eager to receive their course credits, carefully tested under specific experimental conditions. Implementing such scenarios in actual course work is far more challenging, given practical issues such as group sizes. For instance, perhaps one would like to implement an immersive VLE at the time of a specific lecture, yet is faced with the practical impossibility of having a large number of students experience the VLE at the same time.

Larger group sizes are associated with reductions in performance (Mullen, 1994; Petty et al., 1977) because of increased difficulty to reach a consensus in larger groups (Strijbos et al., 2004) and social loafing (i.e., free riding) (Suleiman & Watson, 2008). For non-immersive (gamified) settings, several meta-analyses have investigated the connection between group size and learning. Vogel et al. (2006) examined 32 studies spanning 1986–2003 on the effect of games and interactive simulations on learning. No significant learning differences were observed between single- and multi-person groups, albeit the case that the effect size was higher in the single-person groups. Merchant et al. (2014) analyzed 67 studies published up to 2011 using games, simulations and virtual worlds and also found learning was more effective for individual compared to collaborative study for games, and observed no significant difference for simulations and virtual worlds.

For practical implementations in courses, single-person sessions are hardly feasible due to large student numbers and restrictions on available time. Knowing whether learning gains are obtained in small, medium and large groups is therefore desirable. If learning gains are only obtained in small and not in large groups, findings may be promising from a research perspective but might complicate application from an education perspective.

A similar practical issue is when to apply an immersive VLE as part of course work. That is, if an immersive VR lesson is connected to the overarching subject of a course, prior knowledge is likely to vary depending on when the lesson is applied. This raises the question whether the time of application of immersive VLEs modulates learning gains when using these environments. Several non-VR studies have indicated that different levels of prior knowledge may modulate learning when using multimedia, a phenomenon also known as the “expertise reversal effect” (Chen et al., 2017; Kalyuga et al., 2003). In this effect, guided instruction helpful for learners with little prior knowledge becomes progressively redundant and ultimately disadvantageous for learners with high levels of prior knowledge (Kalyuga, 2014). The expertise reversal effect is explained using cognitive load theory, which posits unnecessary taxation of limited cognitive resources may hamper learning (Sweller et al., 2011). To prevent the expertise reversal effect from occurring, instructions could dynamically adapt to the prior knowledge level of the learner (Kalyuga, 2007). The fundamental practical question how the time of application of a VLE as part of course work may affect learning has seldom been investigated.

The current study

The current study aims to determine the circumstances yielding a trade-off between learning gains and practical feasibility for providing immersive VR experiences to large student numbers. To this end, we investigate whether, and if so, how group size and time of application affect learning in immersive VR when used in the ecologically valid setting of an undergraduate course. In addition, we examine the broader picture of how these

and other factors in immersive VR work in tandem to produce learning. The resulting insights are to facilitate educational institutions considering collaborative immersive VLEs as a novel and efficient means to promote learning in their students. An immersive CAVE-based VLE on the topic of 3D human neuroanatomy was employed that leveraged natural collaborative learning. Using this VLE, we examined learning gains while manipulating group size and time of application. Learning gains were expected to be higher in single-person groups compared to multi-person groups, while prior knowledge was expected to change between application time periods. We expected no interaction between group size and time of application.

Method

Study design

We experimentally manipulated group size and time of application to assess a possible effect on learning gains in a balanced, 3 (group size) \times 3 (time of application) between-subjects design. As we did not expect an interaction between the two, group size was allowed to be nested within time of application. Group sizes consisted of single-person, two- to four-person and five- to six-person groups. This allowed a comparison between both the smallest, and approximately the largest number of learners a CAVE system can reasonably hold. The medium-sized group was included to gain a more refined understanding of the effect of group size in computer-mediated learning, a differentiation mostly absent from meta-analyses on this subject. In accordance with the design of the study, each participant took part only once in the study, in one of the three group sizes and in one of the three times of application. The effectiveness of the VLE to incur learning at different levels of prior knowledge was assessed by applying the VLE in three different time periods: the pre-, mid- and late-term of an undergraduate course. The overarching theme of the course was cognitive science with a lecture on neuroscience in the mid-term of the course that was most closely related to the topic of the VLE. In the pre-term participants had not partaken in the course, yet were from the same student population and background and were naïve to the subject of the VLE. The second time period involved students enrolled in the cognitive science course and was conducted in the mid-term of the course, right before the neuroscience lecture. The third time period was conducted in the late-term of the course after the neuroscience lecture. As such, the time periods respectively reflected minimal, medium and highest possible knowledge of the topic of neuroanatomy of the VLE.

Participants

One hundred fifty-eight students took part in the study, either recruited using a subject pool (first time period) or as part of an undergraduate course (second and third time period). Participant candidates younger than 18 or older than 67 years of age, with a past or current condition of migraine or epilepsy, with (expected) pregnancy, without 3D vision and without normal or corrected-to-normal vision were excluded from participation. As the second and third periods of the study were conducted as part of a course, enrolled students not meeting the requirements were offered an alternative learning experience using a non-collaborative 2D desktop version of the VLE. No data was

collected for these students as they were not part of the study. Permission to conduct the study was granted by the Research Ethics Committee at Tilburg University.

Apparatus

A 5.2 m × 5.2 m four-wall WorldViz CAVE, four corner speakers, a position-tracked 3D mouse and active 3D see-through glasses were used to present the immersive virtual environment to the users and to provide interactivity. The VLE was created using Unity 3D version 5.3.4f1. A 3D model of a human brain and its parts was integrated into the environment and was obtained from database BodyParts3D/Anatomography (The Database Center for Life Science, CC Attribution-Share Alike 2.1 Japan). Realistic speech used to provide guidance and feedback to the user was generated using Amazon Polly, a text-to-speech engine.

Materials

A collaborative immersive CAVE-based VLE on the subject of 3D human neuroanatomy was created based on content obtained from a chapter of a conventional textbook (Friedenberg & Silverman, 2006) and concerned the memorization of brain area shape, position, name and function. The VLE was interactive and incorporated a feedback system, enabling immersive learning without the need for teacher supervision. The VLE featured stereoscopic 3D viewed using position tracked see-through glasses, incorporated gamification elements and was designed for both single- and multi-player use. The instructional design of the VLE was structured to foster interdependence and active participation. As an example of this, the environment guided users to take turns in directly interacting with the educational content after a set amount of interactions. Audiovisual stimuli accompanied direct interactions with the environment, thus making clear how the environment was interacted with at any one time and served to foster active discussion. Different stages segmented the educational content. This content could be freely explored and generatively interacted with using embodied actions. For most stages, the structure of the educational content was essentially the same, with different categories of information each shown on a separate wall of the four-wall CAVE. The main wall showed a large size human brain, used to indicate the shape and position of brain areas and their interconnections, while a second wall showed individual brain areas. The third and fourth wall contained large labels, respectively showing the names of the individual brain areas and descriptions of their function.

At the beginning of a stage, large colored lines connected individually presented brain areas to their respective spatial position within the whole brain, as well as indicated their correct name and function. After memorization of the information, the lines were removed. Using a position tracked 3D mouse, the user(s) had to recreate the correct connections between the individual elements by drawing connecting lines, thus proving to have learned the information. A virtual lever contained on the main wall could be pulled when needed to verify if one or multiple answers were correct. For correct answers a green tick mark was shown above the connecting line in question, while errors were shown by coloring the offending connecting line red. At the end of each stage, all educational content was integrated onto one wall, allowing review and consolidation of the educational content.

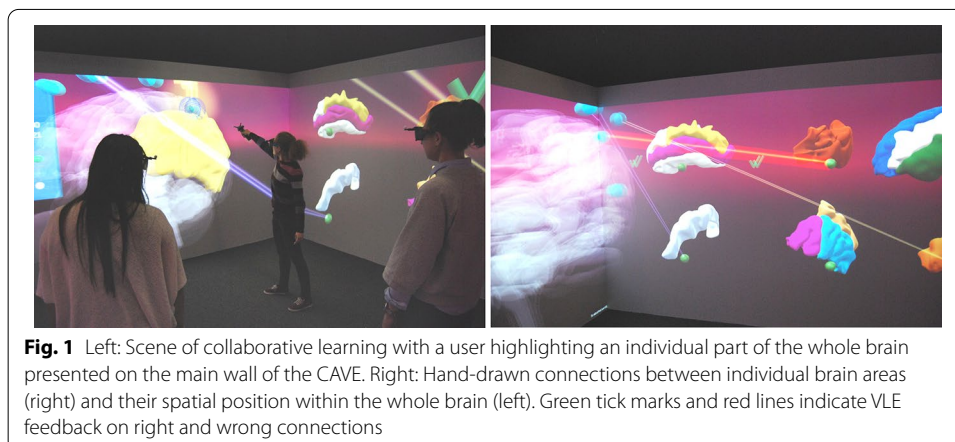
Several design elements of the VLE supported learning at different levels of prior knowledge. Depending on the needs of the user, the large size human brain and its individual components shown on the main wall could be examined from multiple angles to create a better understanding of their structure and spatial relationships. The VLE was flexible in allowing novice users to receive feedback each time an answer was input into the system, while more advanced users could input multiple related answers at a time without receiving redundant intermittent feedback. A scoreboard provided additional feedback on performance to motivate users at different levels of expertise, yet its use was not enforced to prevent hindering those who could perform well without additional support structures. Figure 1 depicts the VLE and its use to enable engaged collaborative learning.

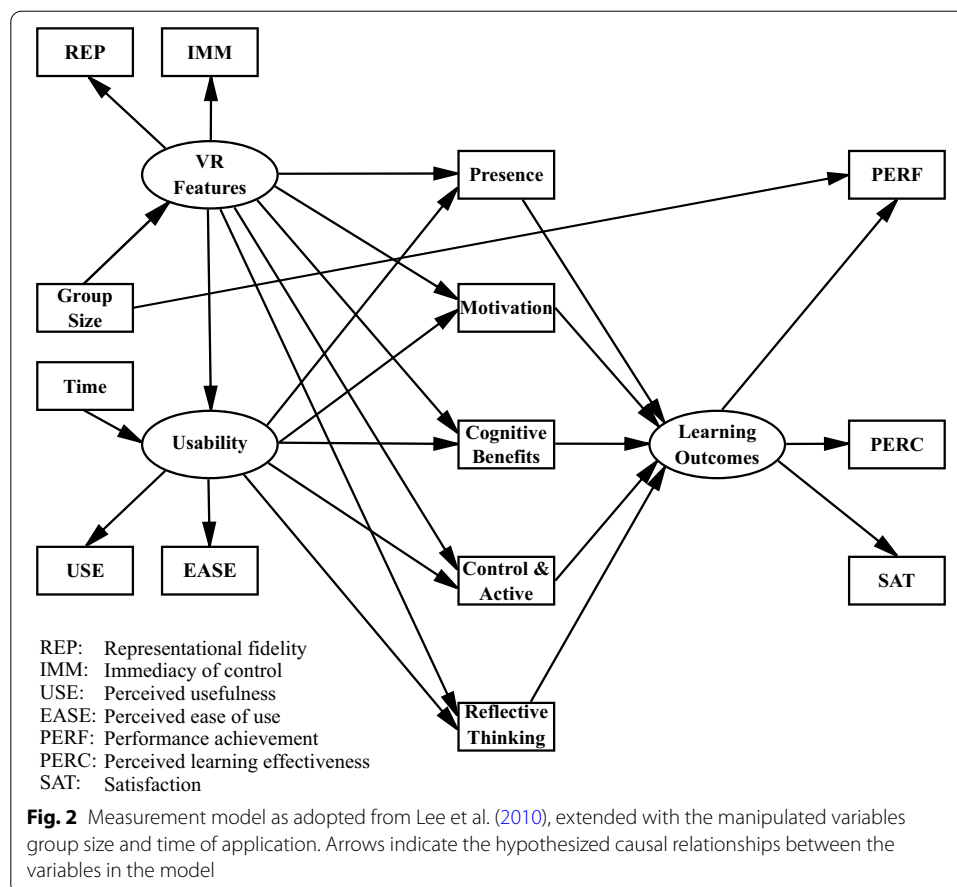
Tests

Two 20-item four-choice multiple choice question tests on the educational content of the VLE were used to assess learning gain performance, identical to the ones used by De Back et al. (2020). Question type (brain area name, function, location) and number were counterbalanced between the two tests. The tests were interchangeably used as pretest and posttest. The order of the tests was counterbalanced such that learning gain performance could not depend on test order.

Measurement model

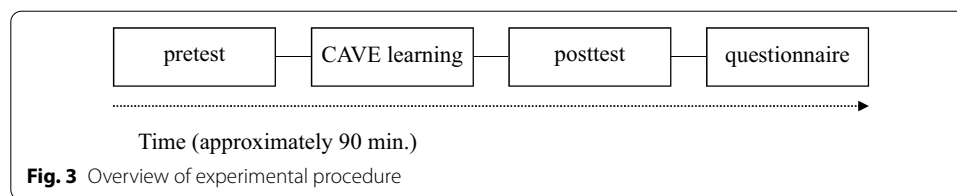
The model of Lee et al. (2010) was adopted for the SEM analysis into how learning results from immersive VR usage. This model is grounded in the model of learning in immersive VR of Salzman et al. (1999) as well as several models of technology-mediated learning, including those of Alavi and Leidner (2001), Sharda et al. (2004) and Wan et al. (2007). Lee et al. (2010) use these studies to support the variables of their model and their predicted relationships. The model assumes technological VR features have an indirect impact on learning outcomes through usability (i.e., the interaction experience), and through several psychological factors (i.e., the learning experience). VR features are measured using representational fidelity (i.e., the realism of both the environment and that of the behavior of the objects within it), as well as using immediacy of control (i.e., the extent of the ability to explore the environment from different perspectives and to





observe and interact with its components). For usability, two aspects are assessed: The quality aspect is measured using perceived usefulness and the accessibility aspect is measured using perceived ease of use. The model also contains five psychological factors together describing the learning experience: presence (i.e., the sense of being part of a computer-generated environment, Heeter, 1992; Steuer, 1992), motivation, cognitive benefits (including perceived benefits to absorb, comprehend and apply the learning material), control and active learning (i.e., perceived control over one's learning, engaged and involved learning) and reflective thinking. The model assumes that VR features affect these five factors both directly as well as through usability. These factors in turn are assumed to be directly predictive of learning outcomes. Learning outcomes are measured using quantitative learning gain performance, as well as using perceived learning effectiveness of and satisfaction with the experience.

For the purpose of the current study, the model of Lee et al. (2010) was extended with the manipulated variables group size and time of application. Group size was assumed to directly predict VR features and usability as well as learning gain performance. Time period was assumed to predict usability only as appraisal of VR features was thought to be time independent. The resulting measurement model is presented in Fig. 2.



Questionnaires

All self-report variables of the SEM model were measured using questionnaires as obtained from Appendix A of Lee et al. (2010)'s paper and were applied in the study of the current paper without alteration. One exception was the presence questionnaire, which was originally measured using a single item, and was replaced using an 18-item spatial presence subscale obtained from the ITC-Sense of Presence Inventory (Lessiter et al., 2001). In addition to the questionnaires of Lee et al. (2010) we used a 4-item questionnaire to assess preference to VR learning over textbook learning. Items of all questionnaires were measured on a 5-point Likert scale.

Procedure

After receiving both oral and written information about the purpose of the study and signing an informed consent form, the participant(s) completed a written pretest. Inside the CAVE, the participant(s) briefly practiced using an introduction stage under the scripted guidance of the experimenter. Next, the experimenter left the CAVE and the participant(s) started with the immersive learning experience. After completing all stages, the participant(s) exited the CAVE and completed a written posttest and questionnaire, concluding the session. Session duration averaged to 90 min. The experimental procedure is depicted in Fig. 3.

Data analysis

Unless specified otherwise, statistical tests were performed using a 3 (group size) \times 3 (time of application) two-way analysis of variance (ANOVA) F -test with SPSS 24 (IBM Corp. in Armonk, NY) and a non-parametric two-way Aligned-Rank ANOVA F -test using the ARTool package (Kay & Wobbrock, 2019) in R (R Core Team, 2019) when parametric assumptions were violated. As prior knowledge was assumed to differ between time periods, learning gain performance was measured by calculating the normalized gain score, which accounts for differences in prior knowledge (Hake, 1998, 2002). Missing answers to the tests were treated as errors. Using the percentage of correct answers in the pretest and posttests, normalized gain was computed using the formula: $(\text{Posttest} - \text{Pretest}) / (100 - \text{Pretest})$. Effect size is indicated using partial eta-squared (η_p^2). Statistical significance is reported two-tailed ($\alpha = 0.05$). All post-hoc pairwise comparisons are corrected for multiple testing using Tukey's HSD.

SEM analyses were conducted in AMOS 24 (IBM SPSS, Chicago) using maximum likelihood estimation. Goodness-of-fit was assessed using normed χ^2 (χ^2/df), the comparative fit index (CFI), Tucker–Lewis Incremental Fit Index (TLI) and root mean square error of approximation (RMSEA), with values ≥ 0.95 for CFI and TLI and ≤ 0.06 for RMSEA taken as indicative of a good fit (Hu & Bentler, 1999).

Table 1 Number of participant groups per group size for the pre-, mid- and late-term

Term	1-person groups	2–4 person groups	5–6 person groups
Pre-term	16	19	18
Mid-term	8	12	11
Late-term	25	23	15

Table 2 Questionnaire scales, with item loading ranges and reliability

Scale	Loadings	Cronbach's α
Representational fidelity	.810–.909	.833
Immediacy of control	.823–.885	.866
Perceived usefulness	.783–.931	.907
Perceived ease of use	.649–.873	.755
Presence	NA	NA
Motivation component 1	.374–.829	.874
Motivation component 2	.341–.836	
Motivation component 3	.610–.793	
Motivation component 4	.434–.641	
Cognitive benefits	.741–.866	.858
Control and active learning	.736–.869	.796
Reflective thinking	.784–.873	.825
Perceived learning effectiveness	.698–.810	.889
Satisfaction	.508–.854	.862

NA = Not applicable, treated as single-item measurement

Technical difficulties occurred in two sessions for four participants, and one participant dropped out due to a headache. In addition, four participants had 20% or more missing data in either the pretest and posttest ($n = 1$) or the questionnaire ($n = 3$). Data of two participants constituted true outliers based on univariate and multivariate tests of non-normality. All further analyses were conducted for the remaining participants ($n = 147$, 76 females, age: $M = 22.122$, $SD = 4.144$). The number of participant groups per group size for the pre-, mid- and late term of the cognitive science course is presented in Table 1.

Measure reliability

A factor analysis was performed on the items of the self-report questionnaires. Presence was treated as a single-item measurement, consistent with Lee et al. (2010). The items of the motivation questionnaire consisted of four categories, and was reflected in the four components found for this questionnaire in the factor analysis. In accordance with Lee et al. (2010), the questionnaire as a whole is used for the SEM analysis. Table 2 presents an overview of the loadings of the items of the questionnaire scales as well as Cronbach's α , which was satisfactory for all questionnaires.

Results

Prior knowledge

As the VLE was applied in three time periods, namely the pre-, mid- and late-term of a course, it was first verified whether prior knowledge differed between these three time periods using the percentage of correct answers on the pretest using a one-way ANOVA F -test. As expected, a significant effect of time of application on prior knowledge was present, $F(2, 144) = 5.30$, $p = 0.006$, $\eta_p^2 = 0.069$. Pairwise comparisons indicated that taking the course led to an increase in prior knowledge, as the percentage of correct answers on the pretest was significantly lower in the first time period ($M = 29.906$, $SD = 10.354$) compared to both the second ($M = 36.290$, $SD = 10.721$), $p = 0.034$, and third time period ($M = 36.111$, $SD = 12.097$), $p = 0.010$, while there was no difference between the latter two time periods, $p = 0.997$.

Learning gains

As significant differences in prior knowledge existed between time periods, these differences were taken into account by using normalized learning gains which adjust for the score on the pretest. No significant interaction between time of application and group size was present for learning gains, $F(4, 138) = 0.88$, $p = 0.480$, $\eta_p^2 = 0.025$. Moreover, no significant main effect of time period on learning gains was observed, $F(2, 138) = 0.21$, $p = 0.807$, $\eta_p^2 = 0.003$, first time period: $M = 0.319$, $SD = 0.244$; second time period: $M = 0.334$, $SD = 0.262$; third time period: $M = 0.353$, $SD = 0.247$, indicating that learning gains due to using immersive VR were comparable at different time periods of the course. By contrast, the main effect of group size on learning gains was significant, $F(2, 138) = 6.41$, $p = 0.002$, $\eta_p^2 = 0.085$, and showed that group size was a relevant factor for learning with immersive VR. Pairwise comparisons revealed that the mean learning gain in the single-person groups ($M = 0.430$, $SD = 0.220$) was significantly higher than that of the two- to four-person groups ($M = 0.307$, $SD = 0.279$), $p = 0.031$, as well as that of the five- to six-person groups ($M = 0.269$, $SD = 0.207$), $p = 0.005$, and that the two–four-person- and five- to six-person groups did not differ significantly, $p = 0.718$. Besides showing that group size significantly modulated learning gains, the findings indicated that learning gains were present for all configurations of group size and time period.

Learning preference

Besides learning gains, student preference for learning with VR over textbook learning was assessed using a 4-item questionnaire to gain insight into student attitude towards the use of these two learning platforms. A non-parametric two-way ANOVA F -test indicated that there was no significant interaction between the effect of time of application and group size on preference for VR learning over textbook learning, $F(4, 138) = 0.87$, $p = 0.486$, $\eta_p^2 = 0.024$. The main effect of time period on learning preference was significant, $F(2, 138) = 3.50$, $p = 0.033$, $\eta_p^2 = 0.048$. Yet, neither the pairwise comparison between the first ($M = 3.731$, $SD = 1.147$) and second time period ($M = 3.694$, $SD = 1.130$), $p = 0.957$, first and third time period ($M = 4.163$, $SD = 0.962$), $p = 0.068$, nor the second and third time period, $p = 0.078$, was significant. Therefore, mean preference for VR learning was highest in the third time period, but this did not differ significantly

Table 3 Internal consistency of the measurement model

Construct	Composite reliability	Average variance extracted
VR features	.89	.90
Usability	.61	.66
Presence	NA	NA
Motivation	NA	NA
Cognitive benefits	NA	NA
Control and active learning	NA	NA
Reflective thinking	NA	NA
Learning outcomes	.79	.59

NA = Not applicable, as these constructs were measured using a single observed variable

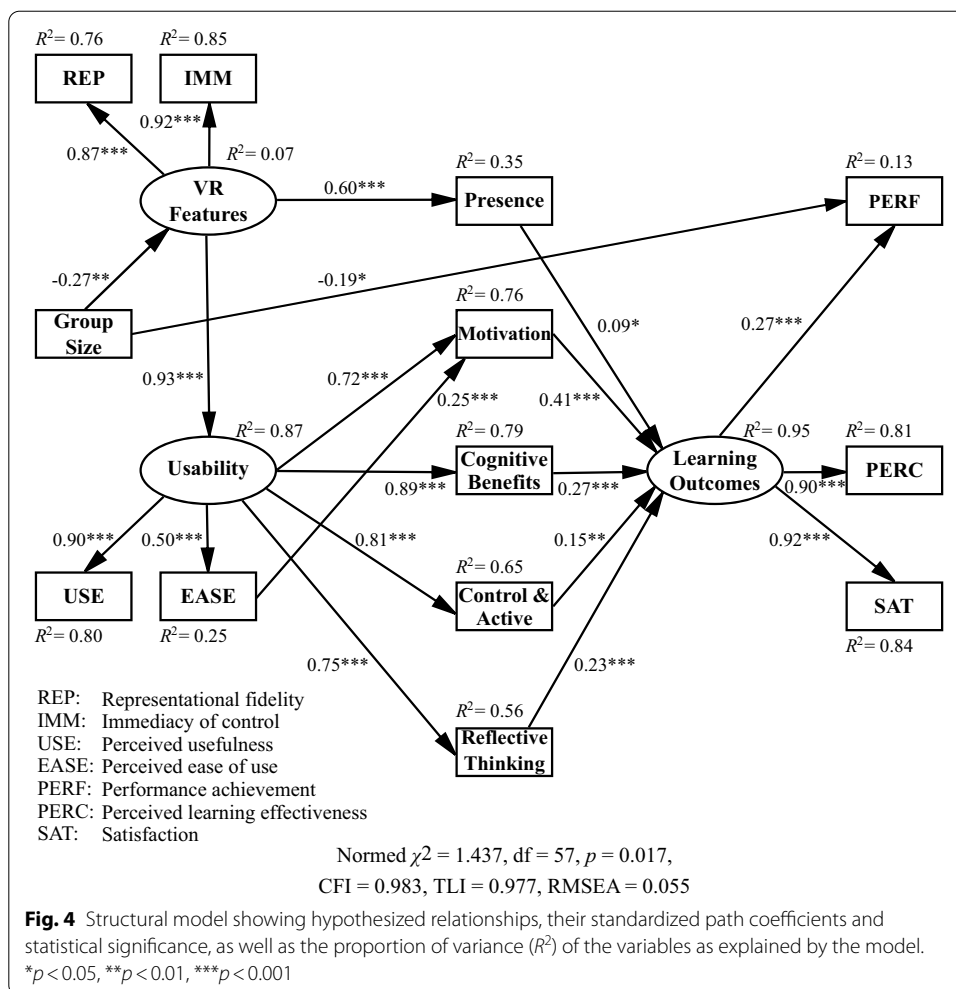
from the preceding time periods. No significant main effect of group size was observed, $F(2, 138) = 2.40$, $p = 0.094$, $\eta_p^2 = 0.034$, indicating student preference for learning with VR was robust for the number of members in the group. The overall mean of preference for VR learning over textbook learning irrespective of time period and group size was 3.908 ($SD = 1.083$).

Measurement model assessment

Internal consistency of the measurement model was verified by assessing composite reliability and average variance extracted of the constructs in the model. Cronbach's α exceeded the common threshold of 0.7 for all constructs and was satisfactory. Consistent with Lee et al. (2010), constructs presence, motivation, cognitive benefits, control and active learning and reflective thinking were measured using a single observed variable, such that computation of composite reliability and average variance extracted was not possible. Composite reliability and average variance extracted of the remaining constructs respectively exceeded the recommended threshold of 0.6 (Bagozzi & Yi, 1988) and 0.5 (Fornell & Larcker, 1981). Table 3 presents an overview of the internal consistency values of the measurement model.

Structural model and analysis

A confirmatory factor analysis of the measurement model yielded a model fit approaching yet exceeding acceptable limits (normed $\chi^2 = 1.956$, CFI = 0.960, TLI = 0.942, RMSEA = 0.081). In a two-step explorative process, (1) one path was added from perceived ease of use to motivation due to a large modification index of 21.093, after which an acceptable model fit was obtained (normed $\chi^2 = 1.625$, CFI = 0.974, TLI = 0.962, RMSEA = 0.065) and (2) non-significant paths were successively removed in order of largest non-significance, consistent with Makransky and Petersen (2019). Most prominently, VR features were restricted to directly predict usability and presence, and time of application did not significantly predict usability, and was therefore removed. This process yielded a more parsimonious model containing significant paths only, with a further improved acceptable fit (normed $\chi^2 = 1.437$, CFI = 0.983, TLI = 0.977, RMSEA = 0.055). The resulting structural model is presented in Fig. 4. The paths in the model are



accompanied by (1) their standardized path coefficients (β), indicating the degree and direction (negative/positive) of the direct relationship between a pair of independent and dependent variables, and (2) the statistical significance of the relationship. Additionally, squared multiple correlations (R^2) are provided, indicating the proportion of the variance of the variable in question explained by the model. Furthermore, standard errors, critical ratios and confidence intervals for the unstandardized path coefficients of the model are presented in [Appendix](#).

The structural model explained 95% of the variance of learning outcomes, 87% of usability, 35% of presence, 76% of motivation, 79% of cognitive benefits, 65% of control and active learning and 56% of reflective thinking.

Group size had a small direct negative effect on VR features, $\beta = -0.27$, $p = 0.003$, and a small direct negative effect on learning gain performance, $\beta = -0.19$, $p = 0.018$. By contrast, VR features strongly predicted usability, $\beta = 0.93$, $p < 0.001$, as well as presence, $\beta = 0.60$, $p < 0.001$, and usability strongly predicted motivation, $\beta = 0.72$, $p < 0.001$, cognitive benefits, $\beta = 0.89$, $p < 0.001$, control and active learning, $\beta = 0.81$, $p < 0.001$ and reflective thinking, $\beta = 0.75$, $p < 0.001$.

Of the five psychological factors directly predicting learning outcomes, the standardized path coefficient of presence was distinctly smaller than the others, presence: $\beta = 0.09$, $p = 0.020$, motivation: $\beta = 0.41$, $p < 0.001$, cognitive benefits: $\beta = 0.27$, $p < 0.001$, control and active learning: $\beta = 0.15$, $p = 0.003$, reflective thinking: $\beta = 0.23$, $p < 0.001$. The model explained a similar 81% and 84% of the respective variance of perceived learning effectiveness and satisfaction, and 13% of learning gain performance.

Discussion and conclusion

The current study investigated factors potentially modulating immersive learning with CAVEs when applied in courses. In earlier work we had established learning gains in a VLE compared to a textbook condition in an experimental lab setting. The current study extended these findings to the more ecologically valid setting of an actual course. Two factors, group size and time of application, were examined to gain insight into the circumstances yielding a trade-off between learning gains and the feasibility for providing immersive VR learning to large student numbers. To this end, immersive VR was applied in small (single-person), medium (two- to four-person) and large size (five- to six-person) groups, both in the pre-, mid- and late-term of an undergraduate course. Additionally, it was examined how learning in immersive VR may arise from technological VR features mediated by learning and interaction experience.

Results replicated those of our previous study in that use of the VLE yielded learning gains. In addition, results confirmed that the course increased prior knowledge as the percentage of correct scores on a pretest directly preceding VR exposure was higher in the mid-term and the late-term of the course compared to the pre-term. Group size had a negative medium size effect on learning gain performance. Consistent with our expectations, learning gains were present for all groups and were highest for single-person groups, and diminished as group size increased. This is in line with the findings of the meta-analyses of Vogel et al. (2006) and Merchant et al. (2014), reporting performance benefits of single-person over multi-person groups for games and interactive simulations, and games yet not simulations and virtual worlds respectively. The current study contributes to the literature in showing that findings for the effect of group size in non-immersive settings hold for immersive settings, and does this for single, medium and large-size groups, a more refined level of differentiation absent from the aforementioned meta-analyses.

Time of application of immersive VR did not significantly affect learning gain performance, implying that students learned just as well with the VLE regardless of their level of prior knowledge. This is indicative of the efficacy of the design of the VLE used, and supports the assertion that an expertise reversal effect in which high prior knowledge learners are disadvantaged by guided instruction benefiting low prior knowledge learners can be countered given appropriate instructional measures (Kalyuga, 2007). The absence of an effect of time of application and thereby of prior knowledge has implications for the ease of use of immersive VR in education and will be examined in detail hereafter.

Previous studies have reported positive student attitudes towards learning with both immersive and non-immersive VR (Jensen & Konradsen, 2018; Mikropoulos & Natsis, 2011). The current study contributes in providing evidence of student preference for learning with immersive VR over textbook learning. Additionally, results revealed that the preference for VR was unaffected by group size and time of application and reflects the robustness of student's positive outlook of VR to support learning.

SEM analyses were conducted using a parsimonious version of the model of Lee et al. (2010) to investigate the broader picture of how immersive VR usage yields learning outcomes, explained by quantitative learning gain performance, perceived learning effectiveness and satisfaction. Results were reflective of the findings of the ANOVA *F*-tests conducted prior, indicating a significant negative effect of group size increase and no significant effect of time of application on learning gains. VR features strongly predicted usability, which in turn strongly affected learning outcomes mediated by psychological factors presence, motivation, cognitive benefits, control and active learning and reflective thinking, and replicates previous findings (Makransky & Lilleholt, 2018; Makransky & Petersen, 2019; Makransky et al., 2017). Presence was directly affected by VR features yet not by usability, and did not strongly predict learning outcomes.

The findings of the current study have implications for the use of immersive VR in education. The findings support the use of VR to complement conventional teaching practice as: (1) the VLE of the current study consistently yielded learning gains across different group sizes and application time periods, (2) these results were obtained in a large student population as part of real-life teaching practice, (3) the application of VR did not require instructional intervention, showing that VR learning benefits can be obtained without increasing the workload of teachers.

Additionally, the findings provide new insight into the circumstances of effective use of immersive VR for learning. Specifically, findings indicated group size to be a factor of relevance for immersive learning, while the time of application of immersive learning as part of a course was not. The implication of this is that for courses with large student numbers, the use of medium size two- to four-person group sessions dispersed across the course is recommended to achieve a trade-off between learning gains and practical feasibility for providing all students with an immersive learning experience. For lower student numbers, the use of smaller size groups is recommended for yielding the highest learning gains.

Finally, the current study has implications for the informed instructional design of immersive VLE. The findings of the SEM analyses indicate the importance of designing for high usability through informed use of technological VR features, shown to be strongly predictive of higher learning outcomes. Designing for high usability in this way may also be beneficial as a countermeasure to a potential negative effect of larger group size on learning. VR features should primarily be incorporated to serve usability with the aim of increasing learning and should not merely be applied for their immersive effect (Dalgarno & Lee, 2010; Fowler, 2015). This is supported by the finding in the SEM

analyses that presence resulting from immersion was not strongly linked to learning outcomes. How to best foster usability through the informed use of VR features will depend on the type and structure of the educational content in question.

In the current study, interaction time with the VLE per user decreased as group size increased. The VLE was designed to keep all users actively engaged in learning regardless, yet the possibility remains that the coupling between interaction time and group size affected learning gains. Future research should aim to disentangle these factors, as well as investigate possibilities allowing multiple users to interact simultaneously with a VLE, thus increasing interaction time of users without increasing session duration. Moreover, collaborative learning in non-immersive computer-mediated environments has been suggested to be modulated by additional factors, including the nature of the learning objective and the type of task used (Strijbos et al., 2004). Future studies are to examine these additional factors for VLE for further understanding of the circumstances for optimal learning in these environments. Additionally, learning preference for VR over textbook learning could have been modulated by additional factors unaccounted for in the current study. An example of such a factor is prior experience with VR. More research is needed to elucidate whether this factor affects learning preference for VR over textbook learning and especially for VR headsets, the use of which is likely to increase given current trends.

For the SEM analyses, an acceptable model fit was obtained after adding one path from perceived ease of use to motivation. Additionally, non-significant paths were removed to achieve a more parsimonious model, consistent with Makransky and Petersen (2019). Due to the modifications made to the model it is necessary for future studies to verify whether the fit of the modified model is observed as well in different populations. The structural model of the current study explained a large proportion of the variance of self-reported learning outcomes, yet a comparatively small proportion of the variance of quantitative learning gain performance. This is consistent with the pattern of results reported by Lee et al. (2010). A worthwhile avenue for further research therefore is to investigate whether additional factors not contained in the current model might explain remaining variance of learning gain performance. A candidate factor is cognitive load, which may negatively affect learning when unnecessarily high (Sweller et al., 2011). An examination of a possible moderating effect of cognitive load seems especially relevant for immersive VR, suggested to have the potential for inducing high load through the inclusion of features irrelevant to learning (Parong & Mayer, 2018).

The current study examined factors affecting learning with immersive VR when used as part of an undergraduate course. The use of the VLE resulted in learning gains. Group size significantly modulated these gains, whereas time of application did not. These findings provide new insights into the use of immersive VR in courses in general and the circumstances for effective learning in VLE specifically.

Appendix

See Table 4.

Table 4 Unstandardized estimates, standard errors, critical ratios and confidence intervals of the paths in the model

Construct	Path	Construct	Estimate	S. E	C. R	95% CI
VR features	←	Group size	− 0.106	0.036	− 2.941	[− 0.177, − 0.035]
Usability	←	VR features	0.939	0.180	5.213	[0.586, 1.292]
Perceived ease of use	←	Usability	0.553	0.092	5.989	[0.373, 0.733]
Motivation	←	Usability	1.240	0.199	6.226	[0.850, 1.630]
Cognitive benefits	←	Usability	2.057	0.319	6.450	[1.432, 2.682]
Control and active learning	←	Usability	1.782	0.287	6.203	[1.219, 2.345]
Motivation	←	Perceived ease of use	0.216	0.043	4.971	[0.132, 0.300]
Reflective thinking	←	Usability	1.810	0.302	5.989	[1.218, 2.402]
Presence	←	VR features	1.065	0.204	5.213	[0.665, 1.465]
Learning outcomes	←	Presence	0.100	0.043	2.320	[0.016, 0.184]
Learning outcomes	←	Motivation	0.449	0.059	7.610	[0.333, 0.565]
Learning outcomes	←	Cognitive benefits	0.217	0.047	4.649	[0.125, 0.309]
Learning outcomes	←	Control and active learning	0.127	0.042	2.986	[0.045, 0.209]
Learning outcomes	←	Reflective thinking	0.179	0.036	5.033	[0.108, 0.250]
Perceived learning effectiveness	←	Learning outcomes	1.003	0.055	18.251	[0.895, 1.111]
Satisfaction	←	Learning outcomes	0.997	0.055	18.251	[0.889, 1.105]
Performance achievement	←	Learning outcomes	0.107	0.032	3.367	[0.044, 0.170]
Perceived usefulness	←	Usability	2.587	0.400	6.467	[1.803, 3.371]
Performance achievement	←	Group size	− 0.058	0.025	− 2.373	[− 0.107, − 0.009]
Representational fidelity	←	VR features	2.612	0.420	6.226	[1.789, 3.435]
Immediacy of control	←	VR features	2.373	0.375	6.322	[1.638, 3.108]

C. R. Critical ratio, CI confidence interval

Acknowledgements

We would like to thank Anton Sluijman, Rob Vossen and Maarten Horden for their help with the development of the immersive VLE, and Daniel Hudson and Andrea Favia for performing the experimental procedure and entering the collected data.

Authors' contributions

TB: Conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, writing—original draft, visualization. AT: Conceptualization, writing—review and editing. ML: Conceptualization, methodology, resources, writing—review and editing, supervision, project administration, funding acquisition. All authors read and approved the final manuscript.

Funding

The current research is funded in part by the PACCAR foundation and DAF Trucks and on a Grant by the European Union, Operational Program Zuid, the Ministry of Economic Affairs, the Province of Noord-Brabant and the municipalities of Tilburg and Gilze Rijen awarded to ML (PROJ-00076). The usual exculpations apply.

Availability of data and materials

The anonymized data used for the analysis of the current study can be obtained from the corresponding author as per the guidelines of the American Psychological Association.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 2 January 2021 Accepted: 19 July 2021

Published online: 07 October 2021

References

- Alavi, M., & Leidner, D. E. (2001). Research commentary: Technology-mediated learning—A call for greater depth and breadth of research. *Information Systems Research*, 12(1), 1–10.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94.
- Birchfield, D., & Megowan-Romanowicz, C. (2009). Earth science learning in SMALLab: A design experiment for mixed reality. *International Journal of Computer-Supported Collaborative Learning*, 4(4), 403–421.
- Buń, P., Trojanowska, J., & Rewers, P. (2019). VR and AR in lean manufacturing classes. In J. Trojanowska, O. Ciszak, J. M. Machado, & I. Pavlenko (Eds.), *Advances in manufacturing II* (pp. 342–351). Springer International Publishing.
- Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, 29(2), 393–405.
- Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6), 65–72.
- Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, 41(1), 10–32.
- De Back, T. T., Tinga, A. M., Nguyen, P., & Louwse, M. M. (2020). Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education*, 17(1), 51.
- Fogarty, J., McCormick, J., & El-Tawil, S. (2018). Improving student understanding of complex spatial arrangements with virtual reality. *Journal of Professional Issues in Engineering Education and Practice*, 144(2), 04017013.
- Fokides, E. (2017). A model for explaining primary school students' learning outcomes when they use multi-user virtual environments. *Journal of Computers in Education*, 4(3), 225–250.
- Fokides, E., & Atsikpasi, P. (2018). Development of a model for explaining the learning outcomes when using 3D virtual environments in informal learning settings. *Education and Information Technologies*, 23(5), 2265–2287.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
- Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology*, 46(2), 412–422.
- Friedenberg, J., & Silverman, G. (2006). *Cognitive science: An introduction to the study of the mind*. Sage Publications.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74.
- Hake, R. R. (2002). Lessons from the physics education reform effort. *Conservation Ecology*, 5(2), 1–33.
- Halabi, O. (2020). Immersive virtual reality to enforce teaching in engineering education. *Multimedia Tools and Applications*, 79(3), 2987–3004.
- Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1(2), 262–271.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Huang, H.-M., Liaw, S.-S., & Lai, C.-M. (2016). Exploring learner acceptance of the use of virtual reality in medical education: A case study of desktop and projection-based display systems. *Interactive Learning Environments*, 24(1), 3–19.
- 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.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539.
- Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 576–597). Cambridge University Press.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.
- Kamiska, D., Fyrch-Nowacka, A., Lefik, M., Zwolinski, G., Wiak, S., Petkovska, L., Cvetkovski, G., Di Barba, P., Mognaschi, M., & Anbarjafari, G. (2019). Learning in virtual reality: A case study in mechatronics. In *Proceedings of the 19th international symposium on electromagnetic fields in mechatronics, electrical and electronic engineering* (pp. 29–31). Nancy, France.
- Kay, M., & Wobbrock, J. (2019). ARTool: Aligned rank transform for nonparametric factorial ANOVAs (R package version 0.10.6). Retrieved from <https://github.com/mjskay/ARTool>.
- Knutzen, K. B. (2019). A hybrid model of experiential learning within the social virtual world of second life. *Journal of Virtual Worlds Research*, 12(2).
- 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*, 55(4), 1424–1442.
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3), 282–297.
- 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.
- Makransky, G., & Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education*, 134, 15–30.
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2017). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236.
- Maresky, H. S., Oikonomou, A., Ali, I., Ditkofsky, N., Pakkal, M., & Ballyk, B. (2019). Virtual reality and cardiac anatomy: Exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education. *Clinical Anatomy*, 32(2), 238–243.
- McFaul, H., & FitzGerald, E. (2020). A realist evaluation of student use of a virtual reality smartphone application in undergraduate legal education. *British Journal of Educational Technology*, 51(2), 572–589.
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29–40.

- Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Kwok, O., Cifuentes, L., & Davis, T. J. (2012). The learner characteristics, features of desktop 3D virtual reality environments, and college chemistry instruction: A structural equation modeling analysis. *Computers & Education*, *59*(2), 551–568.
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, *56*(3), 769–780.
- Mullen, B. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *115*(2), 210–227.
- O'Brien, M. G., & Levy, R. M. (2008). Exploration through virtual reality: Encounters with the target culture. *Canadian Modern Language Review*, *64*(4), 663–691.
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, *110*(6), 785–797.
- Pelargos, P. E., Nagasawa, D. T., Lagman, C., Tenn, S., Demos, J. V., Lee, S. J., Bui, T. T., Barnette, N. E., Bhatt, N. S., Ung, N., Bari, A., Martin, N. A., & Yang, I. (2017). Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery. *Journal of Clinical Neuroscience*, *35*, 1–4.
- Petty, R. E., Harkins, S. G., Williams, K. D., & Latane, B. (1977). The effects of group size on cognitive effort and evaluation. *Personality and Social Psychology Bulletin*, *3*(4), 579–582.
- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from <https://www.Rproject.org/>.
- Salzman, M. C., Dede, C., Loftin, R. B., & Chen, J. (1999). A model for understanding how virtual reality aids complex conceptual learning. *Presence: Teleoperators and Virtual Environments*, *8*(3), 293–316.
- Sharda, R., Romano, N. C., Jr., Lucca, J. A., Weiser, M., Scheets, G., Chung, J.-M., & Sleezer, C. M. (2004). Foundation for the study of computer-supported collaborative learning requiring immersive presence. *Journal of Management Information Systems*, *20*(4), 31–64.
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74.
- Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93.
- Srijbos, J. W., Martens, R. L., & Jochems, W. M. G. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. *Computers & Education*, *42*(4), 403–424.
- Suleiman, J., & Watson, R. T. (2008). Social loafing in technology-supported teams. *Computer Supported Cooperative Work*, *17*(4), 291–309.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). The redundancy effect. *Cognitive load theory* (pp. 141–154). Springer.
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*(3), 229–243.
- Wan, Z., Fang, Y., & Neufeld, D. J. (2007). The role of information technology in technology-mediated learning: A review of the past for the future. *Journal of Information Systems Education*, *18*(2), 183–192.
- Xie, Y., Ryder, L., & Chen, Y. (2019). Using interactive virtual reality tools in an advanced chinese language class: A case study. *TechTrends*, *63*(3), 251–259.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
