Immersive virtual reality in STEM: is IVR an effective learning medium and does adding self-explanation after a lesson improve learning outcomes?

Liisalotte Elme[1] · Maria L. M. Jørgensen[2] · Gert Dandanell[2] · Aske Mottelson[1] · Guido Makransky[1]

Abstract

The goal of the current study was to investigate the effects of an immersive virtual reality (IVR) science simulation on learning in a higher educational setting, and to assess whether using self-explanation has benefits for knowledge gain. A sample of 79 undergraduate biology students (40 females, 37 males, 2 non-binary) learned about next-generation sequencing using an IVR simulation that lasted approximately 45 min. Students were randomly assigned to one of two instructional conditions: self-explanation (n=41) or control (n=38). The self-explanation group engaged in a 10 min written self-explanation task after the IVR biology lesson, while the control group rested. The results revealed that the IVR simulation led to a significant increase in knowledge from the pre- to post-test ($\beta_{\text{Posterior}}=3.29$). There were no differences between the self-explanation and control groups on knowledge gain, procedural, or conceptual transfer. Finally, the results indicate that the self-explanation group reported significantly higher intrinsic cognitive load ($\beta_{\text{Posterior}}=.35$), and extraneous cognitive load ($\beta_{\text{Posterior}}=.37$), and significantly lower germane load ($\beta_{\text{Posterior}}=-.38$) than the control group. The results suggest that the IVR lesson was effective for learning, but adding a written self-explanation task did not increase learning after a long IVR lesson.

Keywords Virtual reality · Self-explanation · STEM education · Generative learning strategies

Introduction

The teaching methods of Science, Technology, Engineering and Mathematics (STEM) are required to change rapidly due to the speedy development of technology and accumulation of knowledge within the applied STEM fields (Kelley & Knowles, 2016). STEM areas

✉ Guido Makransky
   gm@psy.ku.dk

[1] Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353, Copenhagen, Denmark

[2] Department of Biology, University of Copenhagen, Copenhagen, Denmark

are therefore challenged with the task of ensuring that graduating students have the corresponding knowledge, skills, and techniques to match the demands of the modern job market (Jelks & Crain, 2020). However, providing specialized labs, equipment, and enough experts to guarantee a quality knowledge transfer are resource-demanding and not always possible to obtain (Bonde et al., 2014). Educational facilities may struggle finding necessary staff or funding and students may experience a temporary inability to access a professional learning environment, such as has been the case internationally with the current COVID-19 pandemic.

There has also been a decades-long educational challenge with students' low motivation to engage in learning STEM subjects, and governments have made effortful attempts to change the way STEM subjects are being taught (National Research Council, 2012). Pedagogical sciences have therefore developed specific STEM practices that can help with raising students' awareness about the goals, applicability, and relevance of the STEM subjects through combining different STEM subjects, providing situational learning, applying technology, using manipulatives and hands-on learning, applying a problem solving approach, and implementing an inquiry approach to instruction (e.g., Kelley & Knowles, 2016; Zemelman, et al., 2005). In order to boost students' personal motivation, STEM education should also provide socially mediated learning contexts, where the students' sense of self-efficacy is supported (e.g., Clark & Dwyer, 1998; Foote, 1999; Pintrich & De Groot, 1990; Pintrich & Schunk, 1996, 2002; Schunk & DiBenedetto, 2016). To achieve this, Zemelman et al. (2005) advocate having teachers be the learning facilitators and integrating formative assessment into instruction.

STEM education with its current challenges is speculated to possibly benefit from the rapid advancement of Immersive Virtual Reality (IVR) technology that has already seen the development of various educational applications (Radianti et al., 2020; Suh & Prophet, 2018; Wu et al., 2020). It has especially been seen as a promising technology for adoption in K-12 and higher education settings (Makransky & Mayer, 2022; Patterson & Han, 2019; Suh & Prophet, 2018; Tilhou et al., 2020). IVR is predicted to be used broadly in classroom teaching within a few years (Freeman et al., 2017) as a cost-friendly aid for teaching STEM subjects in an effective manner (Dalgarno & Lee, 2010; Hew & Cheung, 2010; Johnson-Glenberg et al., 2011; Kaufmann et al., 2000; Nadan et al., 2011). Compared to other technological knowledge transmission mediums, such as video, slides, or online virtual worlds, IVR has several unique affordances including: the enhancement of spatial knowledge and sense of presence, the ability to provide experiential learning and embodiment, the affordance of creating situated learning, and the capacity to increase intrinsic motivation (i.e., DiNatale et al., 2020). Previous studies have already coupled IVR technology with various learning theories, such as multimedia learning, situated learning and embodied learning, and the cognitive affective model of immersive learning (Dede, 2009; Johnson-Glenberg, 2019; Makransky & Parong & Mayer, 2018; Petersen, 2021; Petersen et al., 2022a). In the next section, we present the specific IVR affordances that support the effective STEM practices, exemplified with the IVR biology simulation on next-generation sequencing (NGS) that was used in the current study.

IVR affordances in secondary and higher educational settings for STEM subjects

In order to effectively show learners how a certain STEM knowledge or skill can be put in practice, a learning context must be created that corresponds to the conditions in which real-life STEM knowledge can be applied (Putnam & Borko, 2000). IVR allows for the

AECT

creation of realistic learning environments that engage the user and create a sense of presence and authenticity (DiNatale et al., 2020). In the case of IVR, a high sense of presence is supported through a head mounted display (HMD) with its head position tracking, its sufficient depth perception, and its ability to isolate learners from viewing the outside world (Baceviciute et al., 2021b; Cummings & Bailenson, 2016). Recent research has shown, for example, that a sense of presence can be especially beneficial in the exploration phases of inquiry-based science learning interventions in STEM (Makransky & Mayer, 2022; Petersen et al., 2020).

Applying a problem solving approach to STEM lessons may be beneficial in order to promote deeper learning, meanwhile promoting the students' awareness about the relevance and practicality of a STEM subject (e.g., Bonde et al., 2014; Fowler, 2015; Zemelman et al., 2005). Dede (2009) warns against separating tasks of learning and doing, which could lead to the fragmentation of the learning process and hinder the transfer of knowledge and skills. Unlike many other media, Immersive Virtual Reality Learning Environments (IVRLEs) allow for the seamless combination of practical goal-directed learning assignments and additional theoretical knowledge. The simulated biology lesson used in the current study has applied this IVRLE affordance and created a real-life practical problem of having to locate the gene that is linked to obesity for the learners to solve. The students' progress in solving the assignment is aided by additional theoretical information presented in a lab pad, thus tapping into many effective STEM practices.

Various STEM fields also place a high value on practical assignments, because hands-on activities help to achieve better knowledge transfer than purely theoretical learning (e.g., Zemelman et al., 2005). The benefits of practical assignments stem from the phenomenon of embodied learning. Embodiment is one of the key features within any learning process (e.g., Barsalou, 2008; Gallagher, 2005; Varela et al., 1992), because the brain's sensory systems, bodily states, and actions in connection to the situation form the basis of our cognition (Barsalou, 2008). Positive findings on learning outcomes have been reported when using embodied approaches in teaching STEM subjects, such as physics (Kontra et al., 2012), chemistry (Ping et al., 2011), mathematics (Howison et al, 2011; Nemirovsky & Ferrara, 2009; Nuñez et al., 1999)**,** geology (Bircheld & Johnson-Glenberg, 2012), and neuroscience (Decety & Grezes, 2006). The virtual manifestations of the user's own body within the IVR—self-avatars—are thought to help create a sense of embodiment (Taylor, 2002). Thanks to this IVR-specific affordance, the user can achieve a sense of self-location, global motor control, and body ownership (Kilteni et al., 2012). Moreover, a high level of interactivity with the virtual environment is often mediated by physical controllers placed in the users' hands, which can support a sense of active engagement and allow one to feel in control of one's own actions, thus supporting the sense of agency within the virtual environment (Makransky & Petersen, 2021; Petersen et al., 2022b; Piccione et al., 2019). The STEM practice of hands-on learning can only effectively be replicated in IVRLE when utilizing the affordances that unite the physical and embodied learning activities with the virtual sensory experience. The biology simulation used in the current study depicts a realistic navigable 3D virtual biology laboratory that allows the students to experience the activities, manipulate the tools and follow the safety measures that are a crucial part of a real NGS process in a biology laboratory. The user is set in a first-perspective view and is coupled with virtual hands wearing the equipment necessary for the lab procedures. The learning simulation is accessible through an HMD, and the interaction is facilitated by two physical controllers.

It is not always possible to provide students with necessary meaningful and motivating practical STEM learning activities that encourage students to be curious and inquisitive

towards the learning content approaches (e.g., National Research Council, 2012; Zemelman, et al., 2005). IVRLEs, however, allow students to perform tasks that would be impossible or impractical to conduct in the real world, leading to a feeling of safety within the learning experience (Dalgarno & Lee, 2010). The biology simulation used in the current study provides a safe environment, where making mistakes does not impose a high cost, allowing students to freely explore the learning contents and manipulate with the tools within the learning environment. This IVR affordance can, in turn, encourage the users to acquire skills of scientific inquiry (e.g., National Research Council, 2012), as well as provide cognitive benefits of hands-on learning (e.g., Zemelman et al., 2005), and improve technological literacy (e.g., Kelley & Knowles, 2016).

Next to the environmental design, learning content naturally also plays an important role for learning effectiveness. Integrating STEM subjects is thought to provide students with the opportunity to learn from relevant and stimulating experiences, improve critical thinking and problem-solving skills, as well as boost knowledge retention (Stohlmann et al., 2012). The IVR simulation used in our learning experiment incorporates knowledge and practices from the fields of genetics, biomedicine, and molecular biology. Thus, this specific IVRLE dismantles the traditional educational boundaries between different STEM subjects in order to provide an experience of the realistic application of the integrated STEM knowledge (e.g., Kelley & Knowles, 2016; Klingenberg et al., 2020; Zemelman et al., 2005). This, in turn, can lead to a higher motivation towards STEM subjects.

Supporting students' sense of self-efficacy is seen as another motivational factor for effective learning (e.g., Clark & Dwyer, 1998; Foote, 1999; Pintrich & De Groot, 1990; Pintrich & Schunk, 1996, 2002; Schunk & DiBenedetto, 2016). Academic accomplishments may depend on the students' self-efficacy even more than on their study motivation (Pintrich & De Groot, 1990; Pintrich & Schunk, 1996, 2002). Research points to feedback as being one of the most significant sources of information for helping students to achieve an optimal sense of self-efficacy (Clark & Dwyer, 1998; Foote, 1999; Warden, 2000; Zimmerman & Martinez-Pons, 1992). Feedback that emphasizes mastery, improvement, and achievement is suggested to enhance the self-efficacy of the learner (Pintrich & Schunk, 2002). Students work harder when they see themselves as competent for the task (Schunk & DiBenedetto, 2016). The STEM practices that can support students' self-efficacy are, among others, having a teacher as a facilitator and integrating assessments into instruction (Makransky et al., 2020b; Zemelman et al., 2005). IVRLEs allow for the utilisation of virtual agents—with the affordance of adapting their form to fit the learning demand—whose presence can mimic the educational dynamics associated with the more traditional pedagogical settings (). Moreover, Parong and Mayer (2018) propose that IVR may even exceed traditional pedagogical teaching practices thanks to the possibility for personalized immediate and adaptive feedback to the students within the virtual learning environments. The biology IVR simulation used in the current study offers the user real-time interaction with a mostly auditory pedagogical agent who provides not only factual and conceptual information and guidance in the practical assignments, but also knowledge checks and timely feedback for the student's actions. Many IVR learning experiments have resulted in higher self-efficacy when learning in IVR as opposed to less immersive media (e.g., Plechata et al., 2022; Pulijala et al., 2018).

To summarize, IVRLEs have the potential to provide embodied, engaging, immersive, and empirical learning experiences, which can lead to a better knowledge transfer (e.g., Makransky et al., 2019a; Meyer et al., 2019) and a higher sense of motivation for STEM learning (e.g., Makransky et al., 2020c; Parong & Mayer, 2018; Petersen et al., 2020; Stepan, et al., 2017; Zhao et al., 2020).

Research on IVR in education

Luo et al. (2021) reported in their recent meta-analysis that: "With the launch of commercial products such as the Oculus Rift and HTC Vive, HMD-based VR interventions have witnessed the largest increase in publication in the last decade (from 14 to 24%)." (p. 7). The existing research has revealed that IVRLEs can have many affordances in educational settings, including several motivational and affective benefits (e.g., Lee & Wong, 2014; Makransky et al., 2020a; Merchant et al., 2014; Parong & Mayer, 2018; Zhao et al., 2020). A recent meta-analysis found a small effect size advantage ($ES = .24$) of using VRLEs compared to non-immersive learning approaches when summarizing results from studies conducted between 2013 and 2019. Results of newer studies comparing learning outcomes for IVRLEs compared to less immersive media in science education have been mixed (e.g., Baceviciute et al., 2021a; Fealy et al, 2019; Luo et al., 2021; Parong & Mayer, 2018; Wu et al., 2020). One challenge which has been repeatedly highlighted when applying IVR in everyday teaching practices is that lessons presented in IVR can lead to higher levels of cognitive load compared to non-IVR based learning mediums (Makransky et al., 2019b; Parong & Mayer, 2020). Cognitive load can occur due to an increased level of visual stimulation that is not always relevant for learning, or due to the unfamiliarity of the control devices (Makransky et al, 2019b). In some cases, the increased cognitive load has even been associated with poorer learning outcomes (e.g., Makransky et al., 2019b; Parong & Mayer, 2018, 2020).

To overcome the increased cognitive strain, several studies have investigated the role of scaffolding techniques around IVR lessons (e.g., Klingenberg et al., 2020; Makransky et al., 2020a; Meyer et al., 2019, 2020; Parong & Mayer, 2018, 2020). It has been suggested that scaffolding techniques can help facilitate learners in IVR to recollect the given material more efficiently, structure it better, and help students associate the information with pre-existing knowledge and apply it in novel situations (Parong & Mayer, 2018).

Zhao et al. (2020) investigated the effects of using the Generative Learning Strategy (GLS) of summarization after a biology IVR and video lesson, and found that the GLS strategy significantly reduced learners' perceived effort both in the IVR and video conditions but did not lead to better content comprehension. The researchers reason that "*writing summaries facilitated the cognitive process of selecting relevant information in the working memory, thus resulting in the reduced effort*." (p. 480). Meyer et al. (2019) investigated the effect of pre-training on learning biology in IVR and through video, and found that knowing the names and characteristics of the main concepts before the session helped both knowledge acquisition and transfer of the learned material in IVR but not the video condition. The authors suggest that without pre-training, the IVR learning experience could have been more cognitively demanding than with video, leaving fewer resources available for information processing and encoding it into the long-term memory. Parong and Mayer (2018) investigated the instructional effectiveness of adding a prompt to write summaries at different points in an IVR-based biology lesson. They found that the students who summarized after each IVR learning segment, performed significantly better on the post test, compared to the participants who did not summarize and who learned from the IVR lesson without segmentation. Makransky et al. (2020a), examined the efficacy of enactment—a GLS that involves engaging in task-relevant movements during learning—and in combination with an IVR science simulation compared the same lesson observed as a video. They found that the IVR and enactment

group had the highest performance in the procedural knowledge and transfer tests and concluded that enactment is specifically relevant for learning in IVR. Finally, Klingenberg et al. (2020) investigated the efficiency of using the GLS of teaching following a higher education biology simulation presented through IVR or on a desktop computer. The authors found an interaction effect between media (IVR/desktop) and method (teaching/no GLS) for the outcomes of retention, transfer, and self-efficacy. The authors concluded that the GLS of teaching is specifically relevant for improving self-efficacy, retention, and transfer in IVR.

Objectives of this study

The different positive findings of scaffolding techniques applied to the IVR science lesson on learning, motivational, and affective measures show that scaffolding is important when applying IVR based lessons in STEM. However, the specific affordances of different scaffolding techniques have not been fully analyzed. To our knowledge, the GLS of self-explanation has not been investigated in the context of IVR science learning.

Moreover, even though various studies investigate the value of IVR in science education (e.g., Chittaro & Buttussi, 2015; Chittaro & Ranon, 2007; Gunn et al, 2017; John et al, 2018; Kesim & Ozarslan, 2012; Monahan et al., 2008; Pan et al, 2006; Rauch, 2007; Yang et al., 2018), many fail to disentangle the instructional design, or pedagogical principles that are essential for understanding the value of this medium from an instructional science perspective. Furthermore, many of these studies are conducted within laboratory settings, which may limit their generalizability to ecological science learning settings.

The current study is a value-added study that investigates the effect of adding a supplementary feature—a self-explanation task—to an IVR based science simulation on perceived cognitive load and learning outcomes. Testing this in an ecological valid setting (a higher education biology course in the beginning of the education) increases the practical value of the results for science education. The self-explanation principle currently under scrutiny was selected based on literature that suggests that it is important to include scaffolding strategies when implementing IVR based science lessons (e.g., Klingenberg et al., 2020; Makransky & Petersen, 2021; Parong & Mayer, 2018, 2020; Zhao et al, 2020). Self-explanation is relevant in applied contexts because it can be easily implemented in a real classroom environment in addition to an IVR learning intervention. We will hereby proceed with introducing the theoretical background of the study and reveal the research questions in the next section, then explain the study methodology in detail, report the experimental results and finally discuss the outcomes and provide our future directions.

Theoretical background

This theory section concentrates on the two main theories supporting the current study. It starts with the self-explanation from generative learning theory, which constitutes the background of our experimental intervention, and binds it with the research goals. The section proceeds with the cognitive load theory in association with the cognitively demanding IVR learning environment, and ends with the study's research questions.

AECT

Self-explanation

Self-explanation is one of eight Generative Learning Strategies (GLS) proposed by Fiorella and Mayer (2015, 2016). Self-explanation is theorized to be a process by which learners form inferences about causal connections or conceptual relationships within the learned material (Bisra et al., 2018). It is a constructive cognitive activity that can be applied at will or in response to a prompt, for understanding novel information or learning new skills (Fonseca & Chi, 2011). Self-explanation is seen as separate from other scaffolding techniques, such as summarizing, explaining to others, or talking aloud, as it is self-focused and has the purpose of making new information personally meaningful. Its process may be completely covert or, if expressed overtly, it may not be intelligible for anyone else but the learner (Chi, 2000).

Positive relationships between self-explanation and learning have been found in early self-explanation research (Chi et al., 1989; Renkl, 1997). Bisra et al. (2018) found a statistically detectable advantage (g = .35) of self-explanation over instructional explanations in their meta-analysis. They hypothesize that by retrieving relevant previously acquired knowledge from memory and elaborating on it using the relevant features in the new information, meaningful associations emerge. Constructing the explanation is thought to engage the fundamental cognitive processes involved in understanding the explanation, recalling it later, and using it to form further inferences.

Previous research has also investigated the learning effectiveness of self-explanation while using prompts either before (e.g., Ainsworth and Burcham 2007), during (Hausmann & VanLehn, 2010) or at the end of the learning activity (Tenenbaum et al., 2008). Evidence points towards better learning achievement when self-explanation has been taught or prompted (e.g., Lin & Atkinson, 2013; Rittle-Johnson et al., 2017). A meta-analysis by Rittle-Johnson et al. (2017) revealed that prompting students to self-explain while learning mathematics had a small to moderate effect on learning outcomes. Moreover, it is theorized that if learners are allowed to decide what to self-explain, constructing explanations may be better adapted to idiosyncratic prior knowledge (Bisra et al., 2018). The researchers also claim that "the beneficial effects of self-explanation may be evident more on transfer tests than recall tests, and more on long-form test items such as essays and problems than multiple-choice questions" (p. 707).

A study by Schworm and Renkl (2006) showed that the participants who received self-explanation prompts only outperformed the participants who were provided with self-explanation prompts together with supplementary instructional explanations on demand. This finding may refer to a boundary condition of the self-explanation prompts, where allowing for relevant instructional explanation may undermine the learners' incentive to exert effort in self-explaining.

The prompt format may affect the elaboration quality generated in the self-explanation process (Bisra et al., 2018). For example, an open-ended prompt format that contains fewer cues than multiple-choice and fill-in-the-blank formats, can allow for more elaborative processing that is better adapted to the learners' unique deficiencies in knowledge. Alternatively, stronger cues in a multiple-choice format may more effectively point out the specific misconceptions the learner should address. Bisra et al. (2018) found that using multiple-choice prompts did not have a statistically significant effect on learning, which may mean that the greater cue strength of multiple-choice prompts can undermine self-explanation effects. Bisra et al. (2018) argue that optimal cue strength may depend on the students' own ability to self-explain the concepts or

procedures at hand. Novices may benefit more from strongly cued self-explanation prompts, but as the level of expertise increases, they may exceedingly benefit from less strongly cued prompts. In conclusion, research on self-explaining suggests that a successful self-explaining task should be prompted by open-ended cues, loosely structured and devoid of supplementary instructional explanations. Since the students in our experiment had some previous knowledge of the subject at hand, they were given an open-ended prompt to self-explain the content of the IVR lesson in a loose format, with a time limit of 10 min. No extra material was provided to the self-explanation group. Thus, all of the recommendations from the self-explanation literature were followed.

Cognitive load theory

Cognitive load theory posits that a high mental load in information processing demands the allocation of mental resources (Kirschner et al., 2011a, 2011b; Sweller, 1988, 2010). The theory proposes that the working memory load (cognitive load) is determined by the number of information elements (Cowan, 2001; Miller, 1956) that need to be processed simultaneously within a certain amount of time (Barrouillet et al., 2007). Cognitive load theory is in its essence a theory of instructional design that proposes that the working memory limitations should be taken into account when developing learning instructions (Sweller, 2010; Sweller et al., 1998, 2011; Van Merriënboer & Sweller, 2005).

Several studies have highlighted the relevance of cognitive load in IVR based STEM lessons. Parong and Mayer (2018) investigated the instructional effectiveness of IVR versus a desktop slideshow for teaching scientific knowledge. They found that students who viewed the slideshow performed significantly better on the posttest than the IVR group. The researchers argued that this effect may have been elicited by the more cognitively demanding IVRLE. Similar results were found in their later study (Parong & Mayer, 2020), where the students who viewed the IVR biology lesson as an interactive animated journey performed significantly worse on transfer tests, reported more extraneous cognitive load and showed less engagement based on EEG measures than those who viewed the slideshow lesson. The authors argue, based on mediational analysis, that IVRLEs may create an affective and cognitive distraction, which may lead to poorer learning outcomes than desktop environments. Makransky et al. (2019b) found that a science simulation presented in IVR led to more presence but less learning compared to a desktop version of the simulation. They measured cognitive load using EEG and found that students in the IVR condition had higher levels of cognitive activity compared to the desktop condition late in the session. The authors highlight the importance of being able to differentiate between different forms of cognitive load in future IVR based research. Makransky et al. (2020a) propose that due to the increased size of the visual field of view of the IVR systems compared to a monitor, the IVR users' sense of presence is also enhanced, but that this feature also increases extraneous cognitive load because learners need to locate relevant content, and often the environment includes seductive details that are not necessary for learning.

Many authors have attempted to divide the working memory load into categories depending on its function (see also: Kalyuga, 2011; Paas et al., 2003, 2004; Sweller, 2010; Sweller et al., 1998; van Merriënboer & Sweller, 2005). Three types of cognitive load have been identified by Sweller et al. (1998) including intrinsic, germane, and extraneous cognitive load. Intrinsic cognitive load (IL) is enforced by the basic structure (the intrinsic nature) of the information that learners are required to obtain. This type of cognitive load

AECT

is based on the interaction between the nature of the material at hand and the learner's level of relevant expertise (Sweller, 2010; Sweller et al., 1998, 2011).

Similar to IL, extraneous cognitive load (EL) requires working memory resources. The manner in which the to-be-learned information is presented (the nature of the instructional design or procedures) or the actions which learners have to perform, impose EL. If the instructional methods are suboptimal, the learners need to apply cognitive processes that do not directly contribute to the construction of their cognitive schemata and do not serve the learning goals (Brünken et al., 2003; Sweller & Chandler, 1994; Sweller et al., 1990). Therefore, EL is related to the manner in which the material is presented, and IL is related to the innate complexity of the information (Sweller et al., 2011). It is assumed that IL and EL are additive (Sweller, 1994; Um et al., 2012). The more working memory resources have to be devoted to EL, the fewer will be available to dealing with IL, thus reducing learning (Sweller, 2010). Research implies that learning in IVRLE may lead to an essential overload more quickly than learning in a less immersive format because the extraneous load is thought to be higher due to the increased amount of sensory information in this medium (Meyer et al., 2019; Richards & Taylor, 2015).

Finally, germane cognitive load (GL) refers to the working memory resources that the person devotes to deal with a lesson. It is the result of helpful cognitive processes, such as abstractions and elaborations that are promoted by the instructional presentation in the learning environment. An efficient learning environment should allow learners to apply available resources to advanced cognitive processes that are associated with germane cognitive load while reducing extraneous cognitive load. (Gerjets et al., 2004).

Research questions

The current study aims to answer two general research questions:

- Research Question 1a: Does an IVR based science lesson significantly increase knowledge about Gene Expression when measured using a pre- to post-test design?
- Research Question 1b: Does the addition of a self-explanation task after an IVR lesson lead to better learning outcomes compared to a control group?
- Research Question 2: Does the addition of a self-explanation task increase students' perceived intrinsic (Research question 2a), extraneous (2b), and germane (2c) cognitive load?

Method

This section provides a detailed description of the methodology of the experiment, including the characteristics of the sample, the full experimental procedure, as well as the account of the study materials, measures and apparatus.

Participants and procedure

The participants were 79 undergraduate students from the biology department at a large European university. The sample consisted of 37 males, 40 females, and two non-binary. The mean age of the students was 23.8 (SD=5.5). The experiment took place in a higher education biology classroom where the students were a part of four lab groups with

approximately 20 participants per group. The experimental session took place during a planned biology lesson where students had to learn about gene transcription using IVR. The procedure in each of the four sessions was identical. The students were first introduced to the content of the lesson and randomly assigned a number which determined their placement in the self-explanation or control group. Students were seated next to each other at long study desks, leaving at least 50 cm between each other in order to provide them with enough space to physically move their upper body and hands when engaging with the learning content in the IVRLE. The first task was to respond to a web-based pre-test questionnaire that took between 5 and 10 min to complete. Then students were given the IVR headsets and controllers, along with verbal instructions on how to use the equipment. The students were encouraged to start the IVR simulation when they felt ready. Three lab assistants/researchers were available to assist the students in case of technical questions or problems. While being immersed in the IVR learning simulation, the students could either sit in their seats or stand up to allow for better movement. The mean time of the IVR lesson completion was 47 min (SD = 14).

Students were individually approached by an experimenter immediately after finishing the simulation. The students were either left seated in their original seats (the control group) or quietly guided to a remote table (the self-explanation group). They were then given a pen and a sheet with appropriate instructions (see Appendix 1). All students received brief oral instructions from the experimenter to read the given instructions carefully, follow the guidelines, without having any contact with other students during the task. The participants were not aware of the goals of the experimental manipulation and did not discover the differences between the two conditions during the experiment. The experimenters ensured that the people engaged in the required activities by making monitoring rounds. The research assistants instructed the students to respond to a post-test questionnaire as soon as the simulation was finished.

Materials

The immersive VR simulation

The learning intervention consisted of an IVR simulation "Gene Expression Unit: Use sequencing to unveil a gene linked to obesity" developed by the educational technology company Labster (for a preview of the simulation see Labster, 2019). The main objective of the simulation was for students to learn how to sequence DNA using pig samples as a model, and to find the genes in the DNA that are linked to obesity. The simulation depicted a biology laboratory with the necessary equipment that learners could manipulate in order to complete the assignments. The main tasks in the IVR simulation were: preparing samples for next-generation sequencing, understanding the principles behind the next-generation sequencing (NGS) technique and performing a quantitative polymerase chain reaction (qPCR) experiment with proper controls (see Fig. 1 for screenshots of the simulation). First, the students were required to perform a real-time quantitative polymerase chain reaction (RT-PCR) to generate complementary DNA (cDNA) from messenger RNA (mRNA) and prepare the samples for NGS by adding the proper genetic labels. After sequencing, they were guided to analyze the data to single out a gene that could be linked to obesity. To confirm the findings, they needed to design a quantitative polymerase chain reaction (qPCR) with proper controls and analyze the resulting curves. The simulation allowed students to see the process inside the NGS machine in the form of 3D close-up animations

Fig 1. Screenshots of the IVR simulation " The Gene Expression Unit:Use sequencing to unveil a gene linked to obesity"

of the process and hear the auditory explanations of the ongoing processes. During the simulation, students interacted with the auditory virtual agent and a lab pad, where they received both oral and written instructions, as well as multiple-choice questions that served as a retrieval practice activity and provided a running score that students could use as a form of formative assessment. The students received 2 points for every correctly answered multiple-choice question at the end of each phase of the learning simulation. The students were allowed to read complementary material within the lab pad to answer the questions, thus the score mostly reflects the attention paid by the student to the study material at hand. The maximum score a student could obtain in the IVR simulation was 110 points and the minimum score was 0 points.

Post-IVR simulation self-explanation and control instructions

The paper-based materials consisted of two instructive texts: one for the self-explanation condition and the other for the control condition. The instructions were handed out to the students after the IVR learning session. The participants in the self-explanation group were required to write the self-explanation task of the learned material on the same sheet of paper that provided them with the instructions. The people in the control group only needed to read the instructions on the paper and continue with other activities. The instructions are presented in Appendix 1.

The purpose of adding the 10-min long self-explanation task to the IVR learning experience for one group of students in this study was to help them reflect over the most critical points of the IVR lesson that they had encountered. They were asked to select the most relevant information for them from the next-generation sequencing lesson and write it down in a self-organized structure. Three features of the task were highly subjective: the precise IVRLE content that was selected for the reflection, the amount of information the students needed to report, and the quality of the written assignments. As all the students were presented with the same IVR lesson that explicitly highlighted the most important aspects

of next-generation sequencing by providing cued prompts during the session, participants were expected to have a fairly similar understanding of what constituted crucial points of the lesson.

The self-explanation task was meant to be self-directed (i.e., Bisra et al., 2018) and its written format was meant to help structure the information students were processing. Even though most self-explanation processes are expected to happen internally, which means that there is no objective way to evaluate the true quality of the self-explanation process of an individual, the written reports provided the raters with necessary cues to evaluate the nature of the reflected contents.

An open-ended prompt format was chosen for the post-learning self-explanation task, since the students were already somewhat familiar with the learning content, which would allow for a more idiosyncratic reflection (e.g., Bisra et al., 2018). The self-explanation process was also deliberately not taught to the students, but merely prompted by a written cue (see Appendix 1), in order not to make them follow a cumbersome format of the reflection assignment and not hinder the formation of their own associations. The task was purposefully set at the end of the learning activity so as to separate it temporally from the multiple-choice questions provided during the IVR learning session. Lastly, no learning materials were present at the self-explanation task, since providing one group with 10 more minutes of the learning content would have rendered the conditions incompatible (Bisra et al., 2018), and/or it could have led to the boundary condition explained by Schworm and Renkl (2006), and undermined the students' incentive to exert effort in self-explaining.

To score the self-explanation assignment, clear subject-centered 10-point coding criteria were developed by a biology lecturer who was responsible for the course. The ratings were performed by two independent raters who had knowledge of the subject matter. A complete coding scheme is presented in Appendix 2. According to the coding schemes, describing each necessary concept in the simulation (not particularly in the correct order), amounted to one point. One self-explanation could maximally receive 10 points. The interrater reliability of the scoring was $\alpha = .96$, which shows that the criteria were objective enough to consider the coding form successful. The mean score of the self-explanation task was 1.6 (SD = 1.4; maximum score = 6). Based on a content matter expert judgement, 2 was used as a threshold to indicate that students had engaged meaningfully enough with the self-explaining task for it to have a difference from the resting group. This meant that both independent raters had to find at least two key concepts stated correctly in the student's written self-explanation task. Using this criterion meant that 22 out of the 41 students in the self-explanation group were classified as engaging meaningfully with the self-explaining task. This threshold helped us distinguish the students who accurately and with enough thoroughness reflected upon the critical points of the simulation in the self-explanation activity, and those who misunderstood important aspects of the learned concepts, were too general or may have been too unmotivated to fulfill the task demands.

Pre- and post-test measures

Students administered a pre- and post-test in an online survey platform. The pre-test consisted of demographics and a 10-item multiple-choice knowledge test about gene transcription, which was the subject of the IVR learning session. The post-test included the same 10-item multiple-choice knowledge test that was used in the pretest, with two additional transfer test items, and a cognitive load measure (Andersen et al., 2020).

The pre- and post-knowledge test was designed to assess the students' conceptual and procedural knowledge retention (e.g., "Why do we need to reverse transcribe the mRNA to cDNA before PCR and Sequencing? (A) Because the RNA cannot be used by the Taq polymerase, (B) Because cDNA is more abundant than mRNA, (C) Because the mRNA cannot be heated, (D) Because the mRNA cannot be sequenced"). Content validity was prioritized by ensuring that the items in the test measured the broad content of the lesson. The test was scored by giving students one point for each correct answer. A significant positive correlation between the scores obtained in the IVR learning simulation and the retention post-test [$r = .59,** n = 69$, $p < .001$] indicate that the knowledge test is measuring the effects of the IVR learning environment. That is, there was a positive correlation between how well students performed in the simulation and the post-test.

In addition to the retention test, we employed two transfer tests which consisted of open-ended questions. The aim of the transfer tests was to assess the participants' ability to apply what they had learned during the lesson to new situations. The first transfer test item assessed conceptual knowledge transfer (*"Describe the flow cell/chip surface and explain how the DNA molecules are bound to it"*). The second transfer test assessed procedural knowledge transfer (*"Describe the three main steps of the bridge PCR step in NGS"*). The transfer tests were graded by two biology lecturers. Each of the questions were divided into three sub-questions which were rated with 0 or 1 based on a scoring rubric agreed upon before scoring. Therefore, one student could obtain a maximum of 3 points in each test. The correlation between the raters' scores was .87, so the mean score across raters was used as the final measure of transfer. There was a significant positive correlation between the conceptual and procedural transfer tests [$r = .51**, n = 69, p < .001$] (see Table 3). Also, the correlation between the score in the simulation and the procedural transfer test was positive and significant [$r = .27*, n = 69$, $p = .03$]. However, the correlation between the score in the simulation and the conceptual transfer test was non-significant [$r = .17, n = 68, p = .17$].

The post-questionnaire also included a cognitive load measure by Andersen (2020). Students responded on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) to three intrinsic cognitive load items (e.g., "The topic covered in the simulation was very complex"), three extraneous cognitive load items (e.g., "The instructions and/or explanations used in the simulation were very unclear"), and four germane cognitive load items (e.g., "The simulation really enhanced my understanding of the topics covered"). The scales had Chronbach's alpha reliability coefficients of .85, .75, and .82 respectively.

Apparatus

The VR simulation was stereoscopically displayed through a Lenovo Mirage Solo (2018) standalone headset with headphones attached. The HMD has two separate lenses for each eye that give access to a high graphical fidelity screen, and attached headphones that carry the sound of the visual display. A head-motion tracking system controls the interaction in the VR set-up, allowing users to move their field of view inside of the virtual 360-degree environment as they turn their heads.

Results

Initially, an analysis was conducted to investigate if the students in the two conditions differed on basic characteristics of age and gender. An independent samples t-test indicated that the groups did not differ significantly in age, $t_{(77)}=.25$, $p=.62$. Furthermore, a Chi-square test indicated that the groups did not differ significantly in the proportion of different genders, χ^2 (2, N=79)=4,76, p=.09. In conclusion, we were unable to find an effect of the basic characteristics on the conditions.

RQ 1: Did the IVR simulation increase students' knowledge, and were there differences between the conditions on knowledge acquisition and transfer?

We used a Robust Bayesian estimation model, with a prior t-distribution instead of a Gaussian function, such that the model is less sensitive to outliers. In Tables 1 and 2, we report the mean predictions and effects of the model of the form $y_{ik} \sim T\left(v, \mu_{ik}, \sigma_{ik}\right)$.

The results indicate that there is a large increase in retention test scores over time, with an effect size of $\beta_{Posterior}=3.29$ (95% CI ranging from 1.78 to 4.80, see Table 1). Therefore, we can conclude that the IVR science simulation was effective in facilitating learning about next-generation sequencing.

The results reveal that the self-explanation group had a higher retention post-test score compared to the control group $\beta_{Posterior}=.37$ however the 95% confidence interval for the effect size ranged from $-.34$ to 1.06 indicating that there is not enough support to conclude that the groups differed significantly on this outcome. The Bayesian estimation model did not converge for the conceptual transfer and procedural transfer outcomes, so we report the results for independent samples t-tests instead. The results indicate that the difference between the self-explanation group (M=.46, SD=.58) and the control group (M=.50, SD=.77) were not significantly different for conceptual transfer ($t_{(76)}=.243$, $p=.808$, d=.054). Furthermore, the results indicate that the difference between the self-explanation group (M=.67, SD=.91) and the control group (M=.54, SD=.76) were not significantly different for procedural transfer ($t_{(77)}=-.696$, $p=.488$, d=.16). Therefore, no significant differences were found between the self-explanation and control groups on the learning outcomes measured in this study.

A post-hoc analysis was performed to investigate if there were differences in the outcomes of knowledge acquisition, conceptual transfer, and procedural transfer when only including the students who engage meaningfully in the self-explanation task (students who scored 2 or more points on the written self-explanation task). The first column in Table 2 illustrates that the students in the self-explanation group who meaningfully engaged with the task had a significantly higher knowledge retention post-test score than the control group (Engaging self-explanation: $\beta_{Posterior}=.64$, 95% CI [.15; 1.45], control group: $\beta_{Posterior}=.44$, 95% CI [$-.09-1.01$]), however the difference was due to a higher pre-test score. The students who engage meaningfully with the self-explanation task, had a significantly larger effect on procedural transfer than the control group, albeit there is a chance that the effect is non-significant. Finally, the results did not converge for the conceptual transfer test, but an independent samples t-test indicated that there was not a significant difference between the self-explanation group of students who engaged meaningfully with the task (M=.69, SD=.61) and the control group (M=.50, SD=.77) ($t_{(60)}=-1.01$, p=.24, d=.27). We can therefore conclude that the students who engaged meaningfully with the self-explanation task did not perform better than the control

Table 1 Posterior predictions and 95% confidence intervals for the dependent variables measured in the pre- and post-test

Outcome	Pre-test M (SD)		Post-test M (SD)		Time [95% CI]	Effect [95% CI]
	Self-explanation group (2)	Control group (1)	Self-Explanation group (2)	Control group (1)		
Knowledge retention	5.49 (.30)	4.78 (.26)	8.09 (.24)	7.72 (.22)	3.29 [1.78; 4.80]	.37 [− .34; 1.06]
Conceptual transfer*			N/A	N/A		N/A
Procedural transfer*			N/A	N/A		N/A
IL			4.01 (.08)	3.66 (.14)		**.35 [.05; .68]**
EL			2.55 (.13)	2.18 (.11)		**.37 [.03; .73]**
GL			2.91 (.15)	3.29 (.11)		**− .38 [− .75; − .01]**

Results are obtained using a Robust Bayesian Estimation using an unequal variances t-distribution model

Bold values respresent significant effects

Table 2 Posterior predictions and 95% confidence intervals for the dependent variables measured in the pre- and post-test

Outcome	Pre-test M (SD)		Post-test M (SD)		Time [95% CI]	Effect [95% CI]
	Self-Explanation group who exerted effort (2)	Control group (1)	Self-Explanation group who exerted effort (2)	Control group (1)		
Knowledge retention	5.99 (.38)	4.78 (.26)	8.38 (.30)	7.72 (.22)	3.53 [1.96; 5.12]	**.64 [.15; 1.45]**
Conceptual transfer*			N/A	N/A		N/A
Procedural transfer*			.91 (.25)	.47 (.13)		.44 [−.09; 1.01]
IL			4.03 (.09)	3.66 (.14)		**.37 [.03; .73]**
EL			2.41 (.14)	2.18 (.11)		.23 [−.12; .58]
GL			3.06 (.18)	3.29 (.11)		−.22 [−.65; .21]

Results are obtained using a Robust Bayesian Estimation using an unequal variances t-distribution model

The analysis is conducted only including the participants in the self-explanation group whose self-explanation task scores were over 2

Bold values respresent significant effects

Table 3 Summary of Intercorrelations, Means and Standard Deviations of Scores on the Conceptual Transfer, Procedural Transfer, Knowledge Retention, Intrinsic Cognitive Load, Extraneous Cognitive Load, Germane Cognitive Load and IVR simulation scores

Measure	1	2	3	4	5	6	7	8	M	SD
1. Knowledge retention	–	**.33***	**.41****	.28	–.08	–.17	**.71****	.29	8.05	1.49
2. Conceptual transfer	**.24***	–	**.54****	.18	.02	–.19	.21	.33	.49	.59
3. Procedural transfer	**.40****	**.51****	–	.16	.00	–.15	**.39***	**.47****	.71	.92
4. Intrinsic cognitiveload	.17	.14	.10	–	–.28	**–.39***	.16	–.04	4.00	.58
5. Extraneous cognitive load	–.15	–.07	–.05	.14	–	**–.42****	.06	–.24	2.56	.80
6. Germane cognitive load	–.10	–.03	–.10	–.21	**–.50****	–	–.29	.17	2.97	.85
7. IVR simulation score	**.59****	.17	**.27****	.06	–.04	–.13	–	.21	152.44	17.21
8. Self-explanation score	–	–	–	–	–	–	–	–	1.62	1.43
M	7.86	.48	.61	3.76	2.39	3.04	1.63	–	–	–
SD	1.42	.68	.84	.70	.75	.81	1.41	–	–	–

The results to the lower diagonal include the correlations for the self-explanation group and control group together. The results to the upper diagonal include results for only the self-explanation group

Bold values respresent significant effects

group on knowledge retention, there was however a positive effect for procedural transfer, albeit it is possible that the effect is not significant, and no significant difference between the groups on conceptual transfer.

In further analyses we investigated the correlation between students' performance in the simulation and other outcomes in the study (see Table 3). There were significant positive correlations between the score obtained in the IVR simulation and the retention post-test score [$r=.60$, ** $n=69$, $p<.001$], as well as the procedural transfer test [$r=.27$,* $n=69$, $p=.03$]. However, the simulation score is not significantly correlated with the amount of pre-to-post-test gain in the retention test [$r=.22$, $n=58$, $p=.11$]. Thus, the engagement in the IVR lesson significantly predicts knowledge retention and procedural knowledge transfer test outcomes, but not the scope of pre-to-post-test change in the retention test scores.

Table 3 also illustrates how students' self-explanation scores were correlated with their learning results. The correlation between self-explanation assignment scores and knowledge retention test scores was positive but not significant [$r=.29$, $n=34$, $p=.09$]. However, the correlation between self-explanation assignment scores and the procedural transfer test was positive and significant [$r=.47$**, $n=34$, $p=.01$]. Finally, the correlation between self-explanation assignment scores and the conceptual knowledge test was positive but did not reach significance [$r=.33$, $n=33$, $p=.07$]. Thus, the quality of the self-reflection task significantly correlates with the procedural knowledge transfer test, but not the other outcomes.

RQ 2: Did the addition of the self-explanation task increase students' intrinsic, extraneous, and germane cognitive load?

Table 1 also shows that the self-explanation group reported significantly higher intrinsic cognitive load than the control group $\beta_{Posterior}=.35$ (95% CI [.05; .68]). Similarly, the self-explanation group reported significantly higher extraneous cognitive load compared to the control group $\beta_{Posterior}=.37$ (95% CI [.03; .73]). Finally, the self-explanation group reported significantly lower germane load $\beta_{Posterior}=-.38$ (95% CI [$-.75;-.01$]) than the control group. In conclusion, students in the self-explanation group reported having higher intrinsic and extraneous cognitive load, but reported lower germane load compared to the control group. This is a third major empirical finding in this study.

Table 2 presents these results for the group of students who engaged meaningfully in the self-explanation task (scored above 2). These results indicate that the students in the self-explanation group who engaged meaningfully with the task report higher intrinsic load than the control group $\beta_{Posterior}=.37$ (95% CI [.03; .73]), but the results for extraneous load $\beta_{Posterior}=.23$ (95% CI [$-.12$; .58]), and germane load $\beta_{Posterior}=-.22$ (95% CI [$-.65$; .21]), were not significant.

Finally, Table 3 illustrates how the quality of students' self-explanation tasks were correlated to perceived cognitive load. The correlation between the students' self-explanation score was negatively correlated with IL ($r=-.42$), and EL ($r=-.24$), but positively correlated with GL ($r=.17$), but none of the correlations reached statistical significance.

Discussion

Firstly, the current paper aimed to contribute to further evidence related to the effectiveness of IVRLEs in science education (e.g., Klingenberg et al., 2020; Meyer et al., 2019; Petersen et al., 2020). The results of the current study support the notion that IVR can be

an efficient learning medium in science education as reflected by the significant and large effect size improvement of knowledge retention from before to after the IVR lesson. This finding may encourage secondary and higher educational establishments to adapt IVR as a science learning tool. Moreover, the finding that the amount of active engagement with the IVRLE predicts the later learning results may motivate pedagogical instructors to encourage the students to interact actively with the IVRLEs during the sessions.

Secondly, the study set out to test whether a supplemental self-explanation assignment could enhance learning. Contrary to expectations, giving students the task of self-explanation did not increase knowledge retention, conceptual transfer, or procedural transfer. However, students who had engaged meaningfully in the self-explanation task (i.e., achieving a score of 2 or more points) had a medium yet non-significant higher procedural transfer score than the control group but the increase was not significant for knowledge retention. Contrary to these results, most studies have found positive learning effects of using scaffolding strategies. For instance, Parong and Mayer (2018) found that asking students to perform the GLS of summarization led to better conceptual knowledge retention. Meyer et al. (2019), showed that adding a pre-training task achieved better knowledge acquisition and transfer in the IVR than a video. Makransky et al. (2020a) found that when students engaged in the GLS of enactment, it led to higher scores in procedural knowledge and transfer tests in IVR. Finally, Makransky et al. (2020a), found that assigning students the GLS of teaching increased knowledge retention and transfer in IVR.

One possible explanation for the general non-significant self-explanation results can be that the instructions for the self-explanation task may have not provided some students with prompts that were structured enough for a thorough engagement in the self-explanation activity. Some students may not have been motivated enough to spontaneously work on the topic with full dedication if they had to come up with their own structure. However, results show that some of the students visibly worked on the task more than others, and also gained from it. These inconsistencies may be explained by the reasoning by Bisra et al. (2018), who propose that optimal prompt strength may depend on the students' own internal abilities to self-explain.

What differentiates this study from the previous studies is that the IVR lesson in our study lasted over 45 min while IVR lessons in the previous studies were much shorter. The IVR simulation used in the Makransky et al. (2020a), Meyer et al. (2019) and Parong and Mayer (2018)experiments lasted approximately 10 min, and the IVR simulation in the Klingenberg et al. (2020) experiment lasted 36 min. Indeed, the students in the self-explanation condition in our study reported significantly higher intrinsic and extraneous cognitive load and lower germane cognitive load compared to the control condition. Moreover, the self-explanation scores significantly predicted the procedural transfer scores, and were positively, albeit not significantly correlated with both knowledge retention and declarative transfer scores, which indicates that the amount of engagement with a GLS seems to affect learning outcomes.

Our findings of a higher level of cognitive load within the self-explanation group contradict the results of Zhao et al. (2020), who reported that a conceptually similar GLS—summarization—at the end of an IVR biology lesson significantly reduced learners' perceived effort, while not leading to better comprehension. One difference between the two studies is that our experiment used a biology IVRLE where students could actively explore and manipulate the virtual objects while solving a scientific problem, thus tapping into some of the most recognized effective practices of STEM education (e.g., Kelley & Knowles, 2016; Zemelman et al., 2005). In contrast, the biology simulation used by Zhao et al. (2020) was a virtual animation where students were passive viewers. Therefore, the characteristics of

the IVRLE may play a role in whether the cognitive load would be reduced or enhanced by a scaffolding GLS technique.

Another difference between our study and most previous studies (with the exception of Klingenberg et al., 2020) is that this study was conducted as part of a mandatory course rather than a low-stakes lab experiment. In general, it has been argued that conducting more research in natural learning settings will increase the validity of the experimental findings on the topic of IVR learning (e.g., Gerjets & Kirschner, 2009). Also, our self-explanation task was a post-intervention task that was applied after a long and cognitively demanding IVR learning simulation, while Parong and Mayer (2018) used the summarization task after individual segments of the simulation. This design was not possible in our setting because it was conducted in an applied educational context using an existing commercially available IVRLE with approximately 20 students simultaneously, which would have made the organization of such a procedure problematic.

Future research

Our study experimented with only one type of learning strategy applied after an IVR lesson and is thus inconclusive about the effectiveness of similar, but slightly varied strategies, in supporting learning in a virtual medium. Since the self-explanation task can be conducted in various forms, which can lead to different outcomes, future research should investigate different prompt formats (e.g., adding multiple-choice prompts, fill-in-the-blank prompts or diagrams), as well as look into the benefits of having students conduct self-explanation at different times in relation to the IVR learning simulation (Bisra et al., 2018).

Since one major difference between the current study and previous studies is the length of the IVR learning experience, it would be beneficial to investigate the value of using GLS strategies for lessons that differ in terms of length and cognitive demand. Our study also purposefully tested the students' learning in terms of both retention and transfer, because using the GLS strategy of self-explanation may benefit one type of information acquisition more than the other. Future studies should construct tests that extensively and separately measure conceptual and procedural learning. This would allow for further scrutiny of the affordances of different types of self-explanation processes.

As our study only used the GLS of self-explanation and did not compare the effectiveness of other GLS strategies in combination with a cognitively demanding IVRLE, there is yet no understanding as to which GLS strategy would lead to better STEM learning results. Thus, in the future studies, self-explanation should be compared to different GLS strategies such as drawing, explaining, or enacting, which have been shown to increase procedural knowledge and transfer in previous IVR research (e.g., Makransky et al., 2020a).

Conclusion

This study contributes to the evidence that supports the use of IVR as an effective learning medium when it comes to acquiring complex scientific conceptual and procedural knowledge. This information could encourage educational institutions to offer IVR based interventions as part of secondary and higher education STEM lessons. If educators are interested in the information structuring techniques that could make learning in IVRLE more effective, this study proposes that a self-explaining task is only profitable for those who engage meaningfully with the activity. However, adding the scaffolding technique leads to

additional perceived intrinsic and extraneous cognitive load and lower germane load when it is applied after a long and cognitively demanding IVRLE session. Finally, comparing different prompt formats in applying self-explanation after an IVRLE, as well as comparing it to other forms of scaffolding techniques within an authentic STEM higher education setting, is highly recommended.

Appendix 1

Instructional texts for the self-explanation and control conditions.
Self-explanation

"In order to reinforce your knowledge, please use 10 min to reflect over and summarize the main principles of next-generation sequencing (NGS). Do this by selecting the most relevant information from the lesson, organizing it into a coherent structure and writing it down on this sheet of paper. Do this quietly without talking to or disturbing other students. Please monitor the time (10 min) yourself and let the experimenters know when you are finished with the assignment. They will give you the last task."

Control

"You now have a 10 min break to engage in free time activities quietly, without talking to or disturbing other students. You are allowed to use the VR headset, phone, computer, etc.for entertainment. Please monitor the time (10 min) yourself and let the experimenters know when you are finished with the assignment. They will give you the last task."

Appendix 2

Self-explanation coding criteria:
Reverse transcription mRNA to cDNA: 1 point
Adenylate DNA:1 point
Generate DNA with adaptor: 1 point
Run PCR: 1 point
Generate cluster: 1 point
Mention flow cell: 1 point
Mention bridge DNA: 1 point
Trim data for adaptors: 1 point
Remove data with low score: 1 point
Map to genome: 1 point

Declarations

References

Andersen, M. S., & Makransky, G. (2020). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning, 37*(1), 183–196. https://doi.org/10.1111/jcal.12478

Baceviciute, S., Lopez-Cordoba, A., Wismer, P., Jensen, T. V., Klausen, M., & Makransky, G. (2021a). Investigating the value of immersive VR tools for organizational training: An applied international study in the biotech industry. *Journal of Computer Assisted Learning, 38*(2), 470–487. https://doi.org/10.1111/jcal.12630

Baceviciute, S., Terkildsen, T., & Makransky, G. (2021b). Remediating learning from non-immersive to immersive media: Using EEG to investigate the effects of environmental embeddedness on reading in virtual reality. *Computers & Education*. https://doi.org/10.1016/j.compedu.2020.104122

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 570–585. https://doi.org/10.1037/0278-7393.33.3.570

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Bircheld, D., & Johnson-Glenberg, M. (2012). *A next gen interface for embodied learning: Smallab and the geological layer cake. Interdisciplinary advancements in gaming, simulations and virtual environments: Emerging trends: Emerging trends* (p. 51). IGI Global.

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*, 703–725.

Bonde, M. T., Makransky, G., Wandall, J., Larsen, M. V., Morsing, M., Jarmer, H., & Sommer, M. O. (2014). Improving biotechnology education through simulations and games. *Nature Biotechnology, 32*(7), 694–697. https://doi.org/10.1038/nbt.2955

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychology, 38*(1), 53–61.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Lawrence Erlbaum Associates.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.

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, 21*(4), 529–538. https://doi.org/10.1109/tvcg.2015.2391853

Chittaro, L., & Ranon, R. (2007). Web 3D technologies in learning, education and training: Motivations, issues, opportunities. *Computers & Education, 49*, 3–18. https://doi.org/10.1016/j.compedu.2005.06.002

Clark, K., & Dwyer, F. M. (1998). Effect of different types of computer-assisted feedback strategies on achievement and response confidence. *International Journal of Instructional Media, 25*(1), 55–63.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 152–153. https://doi.org/10.1017/s0140525x01003922

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

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.

Decety, J., & Grezes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research, 1079*(1), 4–14.

Dede, C. (2009). Immersive interfaces for engagement and learning [Online Magazine]. *Science, 323*, 66–69.

Varela, F. J., Thompson, E., & Rosch, E. (1992). The embodied mind. CogNet.

Zhao, J., Lin, L., Sun, J., & Liao, Y. (2020). Using the summarizing strategy to engage learners: Empirical evidence in an immersive virtual reality environment. *The Asia-Pacific Education Researcher, 29*, 473–482. https://doi.org/10.1007/s40299-020-00499-w

Fealy, S., Jones, D., Hutton, A., Graham, K., McNeill, L., Sweet, L., & Hazelton, M. (2019). The integration of immersive virtual reality in tertiary nursing and midwifery education: A scoping review. *Nurse Education Today, 79*, 14–19. https://doi.org/10.1016/j.nedt.2019.05.002

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*, 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fonseca, B. A., & Chi, M. T. (2011). The self-explanation effect: A constructive learning activity. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 270–321). Routledge Press.

Foote, C. J. (1999). Attribution feedback in the elementary classroom. *Journal of Research in Childhood Education, 13*(2), 155–166.

Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology, 46*(2), 412–422.

Freeman, A., Adams Becker, S., Cummins, M., Davis, A., & Hall Giesinger, C. (2017). NMC/CoSN horizon report: 2017 k–12 edition. Austin, Texas. Retrieved from https://www.nmc.org/publication/nmccosn-horizon-report-2017-k-12-edition/

Gallagher, S. (2005). *How the body shapes the mind*. Cambridge Univ Press.

Gerjets, P., & Kirschner, P., et al. (2009). Learning from multimedia and hypermedia. In N. Balacheff (Ed.), *Technology- enhanced learning* (pp. 251–272). Springer.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*(1/2), 33–58. https://doi.org/10.1023/B:TRUC.0000021809.10236.71

Gunn, T., Jones, L., Bridge, P., Rowntree, P., & Nissen, L. (2017). The use of virtual reality simulation to improve technical skill in the undergraduate medical imaging student. *Interactive Learning Environments, 26*(5), 613–620. https://doi.org/10.1080/10494820.2017.1374981

Hausmann, R. G., & VanLehn, K. (2010). The effect of self-explaining on robust learning. *International Journal of Artificial Intelligence in Education, 20*(4), 303–332.

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, 41*(1), 33–55.

Howison, M., Trninic, D., Reinholz, D., & Abrahamson, D. (2011). The mathematical imagery trainer: From embodied interaction to conceptual learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1989–1998).

Jelks, S. M. R., & Crain, A. M. (2020). Sticking with STEM: Understanding STEM career persistence among STEM bachelor's degree holders. *Journal of Higher Education, 91*(5), 805–831. https://doi.org/10.1080/00221546.2019.1700477

John, N. W., Pop, S. R., Day, T. W., Ritsos, P. D., & Headleand, C. J. (2018). The implementation and validation of a virtual environment for training powered wheelchair manoeuvres. *IEEE Transactions on Visualization and Computer Graphics, 24*(5), 1867–1878. https://doi.org/10.1109/tvcg.2017.2700273

Johnson-Glenberg, M. C. (2019). The necessary nine: Design principles for embodied VR and active stem education. In P. Díaz, A. Ioannou, K. Bhagat, & J. Spector (Eds.), *Learning in a digital world. Smart computing and intelligence.* Springer.

Johnson-Glenberg, M. C., Bircheld, D., Savvides, P., & Megowan-Romanowicz, C. (2011). *Semi-virtual embodied learning-real world stem assessment* (pp. 241–257). Springer.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*(1), 1–19.

Kaufmann, H., Schmalstieg, D., & Wagner, M. (2000). Construct3D: A virtual reality application for mathematics and geometry education. *Education and Information Technologies, 5*(4), 263–276.

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*, 11. https://doi.org/10.1186/s40594-016-0046-z

Kesim, M., & Ozarslan, Y. (2012). VLE in education: Current technologies and the potential for education. *Procedia - Social and Behavioral Sciences, 47*(222), 297–302. https://doi.org/10.1016/j.sbspro.2012.06.654

Kilteni, K., Groten, K. R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments, 21*(4), 373–387.

Kirschner, F., Kester, L., & Corbalan, G. (2011b). Cognitive load theory and multimedia learning, task characteristics, and learner engagement: The current state of the art. *Computers in Human Behavior, 27*, 1–4. https://doi.org/10.1016/j.chb.2010.05.003

Kirschner, P., Ayres, P., & Chandler, P. (2011a). Contemporary cognitive load theory research: The lab-stgood, the bad and the ugly. *Computers in Human Behavior., 27*(1), 99–105. https://doi.org/10.1016/j.chb.2010.06.025

Klingenberg, S., Jørgensen, M. L. M., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a general learning strategy when learning through

desktop and immersive VR: A media and methods experiment. *British Journal of Educational Technology, 51*(6), 2115–2138. https://doi.org/10.1111/bjet.13029

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). Embodied learning across the life span. *Topics in Cognitive Science, 4*(4), 731–739.

Labster. (2019). Labster - Gene Expression Unit: Use sequencing to unveil a gene linked to obesity. Retrieved June 3, 2022 from https://www.labster.com/simulations/gene-expression-unit-2/

Lee, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education, 79*, 49–58. https://doi.org/10.1016/j.compedu.2014.07.010

Lenovo Mirage Solo. (2018). Standalone VR headset, Retrieved September 9, 2020 from https://www.lenovo.com/us/en/virtual-reality-and-smart-devices/virtual-and-augmented-reality/lenovo-mirage-solo/Mirage-Solo/p/ZZIRZRHVR01

Lin, L., & Atkinson, R. K. (2013). Enhancing learning from different visualizations by self-explanation prompts. *Journal of Educational Computing Research, 49*(1), 83–110.

Luo, H., Li, G., Feng, Q., Yang, Y., & Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12538

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. M. (2020a). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology, 113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019a). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*(6), 691–707. https://doi.org/10.1111/jcal.12375

Makransky, G., & Mayer, E. R. (2022). Benefits of taking a virtual field trip in immersive virtual reality: Evidence for the immersion principle in multimedia learning. *Educational Psychology Review*. https://doi.org/10.1007/s10648-022-09675-4

Makransky, G., Mayer, R. E., Nørremølle, A., Lopez Cordoba, A., Wandall, J., & Bonde, M. (2020b). Investigating the feasibility of using assessment and explanatory feedback in desktop virtual reality simulations. *Educational Technology Research and Development, 68*, 293–317. https://doi.org/10.1007/s11423-019-09690-3

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937–958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Petersen, G. B., & Klingenberg, S. (2020c). Can an immersive virtual reality simulation increase students' interest and career aspirations in science? *British Journal of Educational Technology, 51*(6), 2079–2097. https://doi.org/10.1111/bjet.12954

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019b). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

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. https://doi.org/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*. https://doi.org/10.1016/j.compedu.2019.103603

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63*, 81–97. https://doi.org/10.1037/0033-295X.101.2.343

Monahan, T., McArdle, G., & Bertolotto, M. (2008). Virtual reality for collaborative e-learning. *Computers & Education, 50*, 1339–1353. https://doi.org/10.1016/j.compedu.2006.12.008

Nadan, T., Alexandrov, V., Jamieson, R., & Watson, K. (2011). Is virtual reality a memorable experience in an educational context? *International Journal of Emerging Technologies in Learning (iJET), 6*(1), 53–57.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.

Nemirovsky, R., & Ferrara, F. (2009). Mathematical imagination and embodied cognition. *Educational Studies in Mathematics, 70*(2), 159–174.

Yang, C., Kalinitschenko, U., Helmert, J. R., Weitz, J., Reissfelder, C., & Mees, S. T. (2018). Transferability of laparoscopic skills using the virtual reality simulator. *Surgical Endoscopy, 32*(10), 4132–4137. https://doi.org/10.1007/s00464-018-6156-6

Zemelman, S., Daniels, H., & Hyde, A. (2005). *Best practice: New standards for teaching and learning in America's school* (3rd ed.). Heinemann.

Nuñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics, 39*(1–3), 45–65.

Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, *51*(6), 2006–2033.

Piccione, J., Collett, J., & De Foe, A. (2019). Virtual skills training: The role of presence and agency. *Heliyon*, *5*(11), e02583.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*(1/2), 1–8.

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., & Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & Graphics, 30*, 20–28. https://doi.org/10.1016/j.cag.2005.10.004

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785–797. https://doi.org/10.1037/edu0000241

Warden, C. A. (2000). EFL business writing behaviors in differing feedback environments. *Language Learning, 50*(4), 573–616.

Parong, J., & Mayer, R. E. (2020). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12482

Patterson, T., & Han, I. (2019). Learning to teach with virtual reality: Lessons from one elementary teacher. *TechTrends, 63*(4), 463–469.

Petersen, G. B., Klingenberg, S., & Makransky, G. (2022b). Pipetting in virtual reality can predict real-life pipetting performance. *Technology, Mind, and Behavior*.

Zimmerman, B. J., & Martinez-Pons, M. (1992). Perceptions of efficacy and strategy use in the self-regulation of learning. In D. H. Schuck & J. L. Meece (Eds.), *Student perceptions in the classroom.* Lawrence Erlbaum.

Petersen, G. B., Klingenberg, S., Mayer, R. M., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology, 51*(6), 2098–2114. https://doi.org/10.1111/bjet.12991

Petersen, G. B., Petkakis, G., & Makransky, G. (2022b). A study of how immersion and interactivity drive VR learning. *Computers & Education*. https://doi.org/10.1016/j.compedu.2021.104429

Ping, R., Decatur, M., Larson, S., Zinchenko, E., & Goldin-Meadow, S. (2011). Gesture-speech mismatch predicts who will learn to solve an organic chemistry problem. In *Annual Meeting of the American Educational Research Association*.

Pintrich, P. R., & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40.

Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theory, research, and applications*. Merrill/Prentice Hall.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Prentice Hall.

Plechatá, A., Morton, T., Perez Cueto A. F. J., & Makransky, G. (2022). Why just experience the future when you can change it: Virtual reality can increase pro-environmental food choices through self-efficacy. *Technology, Mind and Behavior*.

Pulijala, Y., Ma, M., Pears, M., Peebles, D., & Ayoub, A. (2018). Effectiveness of immersive virtual reality in surgical training: A randomized control trial. *Journal of Oral and Maxillofacial Surgery, 76*(5), 1065–1072. https://doi.org/10.1016/j.joms.2017.10.002

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning:Recent developments and future directions. *Educational Psychology Review, 17*, 147–177. https://doi.org/10.1007/s10648-005-3951-0

Putnam, R., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4–15.

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, 51*, 1991–2005. https://doi.org/10.1111/bjet.13023

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers in Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rauch, U. (2007). Who owns this space anyway? The Arts 3D VL Metaverse as a network of imagination, In. *Proceedings of ED-MEDIA* (pp. 4249–4253).

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*(1), 1–29.

Richards, D., & Taylor, M. (2015). A Comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: An experiment to find the right representation involving the Marginal Value Theorem. *Computers & Education, 86*, 157–171. https://doi.org/10.1016/j.compedu.2015.03.009

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In *Handbook of motivation at school* (pp. 34–54). Routledge.

Rittle-Johnson, B., Loehr, A., & Durkin, M. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM Mathematics Education, 49*(4), 599–611.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education, 46*(4), 426–445.

Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., & Iloreta, A. (2017). Immersive virtual reality as a teaching tool for neuroanatomy. *International Forum of Allergy & Rhinology*, *7*(10), 1006–1013.

Stohlmann, M., Moore, T., & Roehrig, G. (2012). Considerations for teaching integrated STEM education. *Journal of Pre-College Engineering Education Research, 2*(1), 28–34. https://doi.org/10.5703/12882 84314653

Suh, A., & Prophet, J. (2018). The state of immersive technology research: A literature analysis. *Computers in Human Behavior, 86*, 77–90.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–223. https://doi.org/10.1207/s1532690xci1203_1

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology, 119*, 176–192. https://doi.org/10.1037/0096-3445.119.2.176

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Taylor, T. L. (2002). Living digitally: Embodiment in virtual worlds. *The social life of avatars* (pp. 40–62). Springer.

Tenenbaum, H. R., Alfieri, L., Brooks, P. J., & Dunne, G. (2008). The effects of explanatory conversations on children's emotion understanding. *British Journal of Developmental Psychology, 26*(2), 249–263.

Tilhou, R., Taylor, V., & Crompton, H. (2020). 3D virtual reality in K-12 education: A thematic systematic review. In S. Yu, M. Ally, & A. Tsinakos (Eds.), *Emerging technologies and pedagogies in the curriculum bridging human and machine: Future education with intelligence.* Springer.

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485–498. https://doi.org/10.1037/a0026609

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