Draft version 2023-06-11 submitted for peer review

Virtual reality in primary science education: Improving knowledge of the water cycle

Corinna S. Martarelli[1], Josua Dubach[2], Natalie Schelleis[3], Trix Cacchione[3], & Sebastian Tempelmann[2]

[1] Faculty of Psychology, UniDistance Suisse, Switzerland

[2] Bern University of Teacher Education, Switzerland

[3] Institute of Primary Education, University of Applied Sciences and Arts Northwestern Switzerland, Switzerland

Correspondence concerning this article should be addressed to Corinna Martarelli at corinna.martarelli@fernuni.ch

Acknowledgments

We thank the Ateo company for programming the virtual environment, Katrin Bölsterli and Matthias Probst for their expertise in the development phase of the virtual environment, and all the children and the school of Naters for taking part in the study.

Funding

This study was supported by the Swiss National Science Foundation (grant number 407740_187327), which was awarded to Corinna Martarelli, Trix Cacchione, and Sebastian Tempelmann.

Conflict of interest

The authors declare no competing interests.

Data availability

The data analyzed for the study are available on OSF (https://osf.io/pdwgv/).

Abstract

Virtual reality (VR) can provide access to otherwise inaccessible aspects of the world and thus foster science learning. We developed a VR learning tool about the water cycle, with 11 lessons for classroom teaching at the primary level. We assessed prior knowledge before a four-week intervention and learning outcomes directly after the intervention, as well as eight weeks later. A total of 165 11–12-year-old children participated in the study. We manipulated *immersion* through the use of VR headsets vs. computers and *interaction* through direct engagement with the virtual world vs. observation of an avatar executing the same actions, which allowed us to test the impact of VR on learning about the water cycle. The results showed an effect of immersion but not of interaction. Children who worked with headsets outperformed children who used computers both in the test immediately after the intervention and eight weeks after the intervention. Further, we assessed several affective and cognitive variables during the intervention, including presence, motion sickness, cognitive load, physical load, and children's satisfaction (liking). The findings indicate that immersive VR is a promising tool for teaching science topics about otherwise inaccessible aspects of the world. Moreover, VR can be effectively integrated into a standard primary school classroom setting. Future research is needed to better understand how interactive elements can enhance learning in this context.

Keywords: Virtual reality; water cycle; learning outcome; 11–12-year-olds

Highlights

- 165 11–12-year-olds participated in a virtual reality (VR) study.
- Headsets facilitate learning.
- Headsets showed higher presence, motion sickness, and physical load, but lower cognitive load.
- VR is a promising tool for teaching science.

Virtual reality in primary science education: Improving knowledge of the water cycle

Children learn via experience; *being there* and *handling objects* are important sources of learning from an educational (Collins et al., 1989) and developmental point of view (Carey, 2009). Perceptual input provides children with rich phenomenological information about objects and events in the world. Acting in the environment offers important feedback on how properties of objects and events change across different interventions and is thus considered a particularly important source of causal learning (e.g., Blouw et al., 2016; Gopnik & Wellmann, 2012). Accordingly, knowledge development takes place in a constant interplay with the environment. Children focus their attention on certain objects of interest, manipulate them, and extract statistical regularities from the resulting sensory feedback. In doing so, they spontaneously learn about the characteristics and behavior of living beings and inanimate objects (e.g., Carey, 2009).

However, many science topics are not directly accessible and cannot be interacted with in this way. For example, we cannot fully perceive and experience processes of the microcosm, such as molecules, or those of the macrocosm, such as solar systems. This is also true for processes that unfold over very short (e.g., chemical reactions) or very long periods of time (e.g., evolution). In these situations, we cannot build our knowledge on direct observation and handling, which renders learning about these topics difficult (Johnstone, 1991). As a result of these limitations, empirical evidence from the educational field highlights the difficulty of science learning in school-age children (Reinfried & Tempelmann, 2014; Vaiopoulou & Papageorgiou, 2018) and in adults (Potvin et al., 2020; Shtulman & Valcarcel, 2012). The absence of direct access to aspects of the world makes teaching science in school (and beyond) challenging. Younger children in particular (e.g., primary school age) find it difficult to acquire abstract scientific concepts, as they depend heavily on grounding abstract contents in subjective experiences (Vaiopoulou & Papageorgiou, 2018).

In light of these challenges, there is a need for innovative learning and teaching methods that provide access to these unobservable processes, as well as opportunities to interact with and gain sensory experiences of them. One promising approach is immersive virtual reality (VR), which uses a head-mounted display to completely immerse users in a virtual environment, thus enabling them to engage naturally with the simulated world around them using their own senses and motor skills (Burdea & Coiffet, 2003). Thus, this technology can provide access to processes that are otherwise inaccessible, such as those of the microcosm and macrocosm. Immersive VR allows one to simulate in a dynamic manner not only existing micro processes (e.g., formation of a raindrop) but also difficult-to-access models (e.g., the particle model of water). Furthermore, it is possible to interact with the virtual world via motion controllers or hand-tracking devices. The haptic feedback, which is usually received via vibration to signal contact between the hands and the virtual world, can be realistic, despite being reduced when compared to haptic experiences in the real world. In short, in an immersive VR setting, there is no separation between the individual and the virtual world (high immersion), and it is possible to interact with this new environment (different modes of interaction). Unobservable, inaccessible aspects can be transferred with immersive VR into an accessible learning environment, in which children can enter, act, perceive, and be provided with contingent feedback. Thus, it is conceivable that immersive VR technology could provide ways of learning about science that are better adapted to children's cognitive needs. Our study thus aims to investigate whether VR can be used to teach inaccessible structures and processes of the micro- and macrocosmic world in a way that aligns with the cognitive and behavioral characteristics of school-aged children, thereby enhancing learning. We also aim to assess the practical feasibility of integrating VR into a primary school setting. Specifically, we will examine the respective roles of immersion and interaction in VR-supported learning.

Recent empirical work suggests that both the immersive and interactive aspects of VR support science learning. For example, Liu et al. (2020) developed four immersive VR science units designed to sustain conceptual understanding (e.g., categories of animals and plants) with 11-year-old children. The virtual world was presented via headsets, and the children could interact with it. Around 45 children participated in the experimental group (immersive VR), and their learning outcomes were compared to a control group of about 45 children who were taught with traditional methods. The authors used a pre-test/post-test design, and the results revealed that the children in the experimental group significantly outperformed those in the control group. Chen et al. (2020) found a similar pattern of results when comparing children working with computers (including some immersive equipment, such as passive stereo glasses) and those working with traditional methods. More specifically, the researchers found that the VR group not only outperformed the control group but also showed enhanced hands-on abilities. It seems that both immersive (headsets) and low-immersive (computers) VR help ground abstract scientific knowledge, such as understanding of the water cycle. In a more recent study, Liu et al. (2022) again found a positive impact of immersive VR on science learning when compared to traditional methods. In their study, the children took lessons on weather formation, the human respiratory system, and the human digestive system, all of which included aspects that were not directly accessible. These results support immersive VR as an extremely promising tool for teaching science classes.

Several studies have found that immersive VR has a positive impact on science learning; however, most of these studies were carried out with older children and young adults (e.g., Makransky & Petersen, 2019). With this particular population, it has been shown that immersive VR is not always beneficial due to the difficulty of developing VR environments that do not increase cognitive load. There is indeed an ongoing debate on whether the novelty of the technology, as well as its richness, might actually impair learning (e.g., Meyer et al., 2019) or enhance it by, for example, helping children focus on the VR task (Parong & Mayer, 2021).

Immersive VR also makes learning more motivating and engaging, which is important to consider, given that topics in science suffer from a lack of child interest in primary school (Maltese & Tai, 2010). For example, the study of Liu et al. (2020) mentioned above also found a significant effect on children's engagement in science lessons, as the immersive VR group showed higher engagement when compared to the group working with traditional methods. Overall, it seems that VR enhances learning motivation (Makransky et al., 2019).

Present study

The findings reviewed above suggest that VR might be an extremely promising educational tool; however, empirical evidence is still scarce, especially when it comes to school-age children (Pellas et al., 2021). In the present study, we focused on a specific age group (11–12-year-olds) and a specific content area (water cycle). We selected the topic of the water cycle because it contains several non-directly accessible structures and processes (e.g., formation of raindrops and the particle model of water, including states of aggregation), and it is a foundational topic in primary school that is linked to a wide range of subjects related to environmental problems (e.g., climate change and water pollution) and is thus of high societal relevance. We created a VR tool with 11 lessons about the water cycle for "NMG" (*nature-human-society*) to be taught in Swiss primary schools.[1] Our main goal was to investigate whether primary school children learn more about the water cycle (e.g., in terms of the amount and consolidation of knowledge) with immersive VR (headsets) than with low-immersive VR (computers). We hypothesized that two VR characteristics supporting learning are *immersion* in virtual worlds and *interaction* with them and aimed at disentangling the respective impacts of these two features. We manipulated immersion via the medium (immersive VR headsets vs. low-immersive

In German-speaking parts of Switzerland, the subject NMG includes scientific disciplines and is taught up to the age of 11–12. One of the fundamental themes in NMG is the water cycle, which is addressed in extensive detail. Moreover, important sub-concepts of the water cycle are addressed in several other contexts (e.g., explaining aggregate states with the particle model or understanding weather phenomena).

computers) and interaction via the possibility of interacting with the virtual world (the child directly interacts with the virtual world vs. the child observes an avatar executing the actions). Based on previous theoretical and empirical evidence (e.g., Makransky & Mayer, 2022), we expected the high immersion-high interaction combination (children working with headsets and interacting with the virtual world via controllers) to be most beneficial to the learning process. We anticipated that high immersion and high interaction could foster learning by providing access to otherwise hidden processes and by realizing natural sensory-motor contingencies, thus rendering action in the virtual world similar to action in the real world. Furthermore, we assessed several affective and cognitive variables during the intervention, including presence, motion sickness, cognitive load, physical load, and children's satisfaction (liking). These variables were included for exploratory purposes; we thus refrain here from formulating specific hypotheses.

Methods

Participants

An a priori G*Power analysis indicated a required sample size of 116 participants (at least 29 participants in each of the four conditions) to detect a small-to-medium effect in an analysis of variance with repeated measures (within-between interaction; parameters: *partial* η^2= .03, α = 0.05, $1{-}\beta$ = 0.95, number of groups: 4, i.e., immersion [low vs. high] x interaction [low vs. high], number of measurements: 3, i.e., pre-test, post-test, follow-up test; Faul et al., 2007; see main analysis in the section results under "learning outcome"). To ensure enough valid datasets for each condition, eight classes (two in each condition) took part in the study. Children who missed at least one of the three sessions when learning outcomes were assessed (pre-test, post-test, follow-up test) were excluded (i.e., a total of 13 children, ranging from 3 to 4 in each of the four conditions). In the final sample, 38 children (36.82% female, mean age of 11.42, ranging from 10 to 13 years of age) participated in the high immersion and high interaction condition, 36 children (50.00% female, mean age of 11.42, ranging from 11 to 13 years of age) in the high immersion and low interaction condition, 37 (51.35% female, mean age of 11.30, ranging from 10 to 12 years of age) in the low immersion and high interaction condition, and 36 (52.78% female, mean age of 11.31, ranging from 10 to 12 years of age) in the low immersion and low interaction condition.

During the intervention, we also assessed variables such as presence, motion sickness, cognitive load, physical load, and liking. Four children (one in the high immersion-high interaction condition and three in the low immersion-high interaction condition) had missing values for these variables. Additionally, during the five teaching units where we assessed cognitive and affective variables, there were missing values for 10% to 16% of the total intervention sample (N

= 147). However, we did not exclude children with missing values from our analyses. We also included a control group of 18 children (33.34% female) with a mean age of 11.44 years (range: 10–12) who received no intervention and underwent the three learning outcome measurements.

The study was approved by the local ethics committee. Informed written parental approval was obtained before the study, and the children gave oral consent to participate.

Design and procedure

In the present study, we manipulated the *immersive* and *interactive* aspects of a virtual environment with content about the water cycle. This gave us a 2 × 2 design (immersion and interaction were between-subjects factors) with the following four conditions: high immersion and high interaction (children worked with immersive VR headsets and operated changes with motion controllers), high immersion and low interaction (children worked with immersive VR headsets and observed an avatar operating changes), low immersion and high interaction (children worked with computers and operated changes with a mouse), low immersion and low interaction (children worked with computers and observed an avatar operating changes). We report in Appendix A a brief description of each unit (in terms of objectives and content), as well as the interactive elements used in them. We tested the impact of these manipulations with a pre-test/post-test/follow-up design (measurement point was a within-subjects factor).

Knowledge about the water cycle was assessed three times: before the intervention, immediately after the intervention, and eight weeks after the intervention. For the intervention part, we created full teaching units (11 lessons of 45 minutes each) about the water cycle, including the particle model of water. One part of the teaching unit (about 10 minutes per lesson) was the experimentally critical section using VR, in which we manipulated the variables of immersion and interaction. The second author of this publication, an experienced primary teacher, taught all the units in their entirety in the eight classes, and the third author of this publication assisted with the teaching. This procedure ensured that the teaching was kept as consistent as possible in the eight classes, apart from the experimental manipulation in the virtual environment. During the second, fourth, sixth, eighth, and tenth teaching units (a total of five times), we assessed the cognitive and affective variables (exploratory part of the study). In the following section, we describe the virtual environment, materials, equipment, and procedure of this study.

Virtual environment

The virtual environment was created by the Ateo company (https://www.ateo.ch) using UnityPro software as the main programming tool. The environment contains water in its different forms (e.g., a lake and snow on mountains). We had an avatar named ONOS that gave instructions and accompanied the children through the virtual sessions. Changes in location were possible (i.e., teleporting in the clouds or through the particle model). The virtual sessions involved several activities, such as coloring water, observing the evaporation process, changing temperature, and watching how water particles behave in different states (liquid, ice, and gas). A brief description of the material covered in the units is provided in Appendix A. Images of the environment are shown in Figure 1, and a short video of the environment can be found on OSF (https://osf.io/pdwgv/).

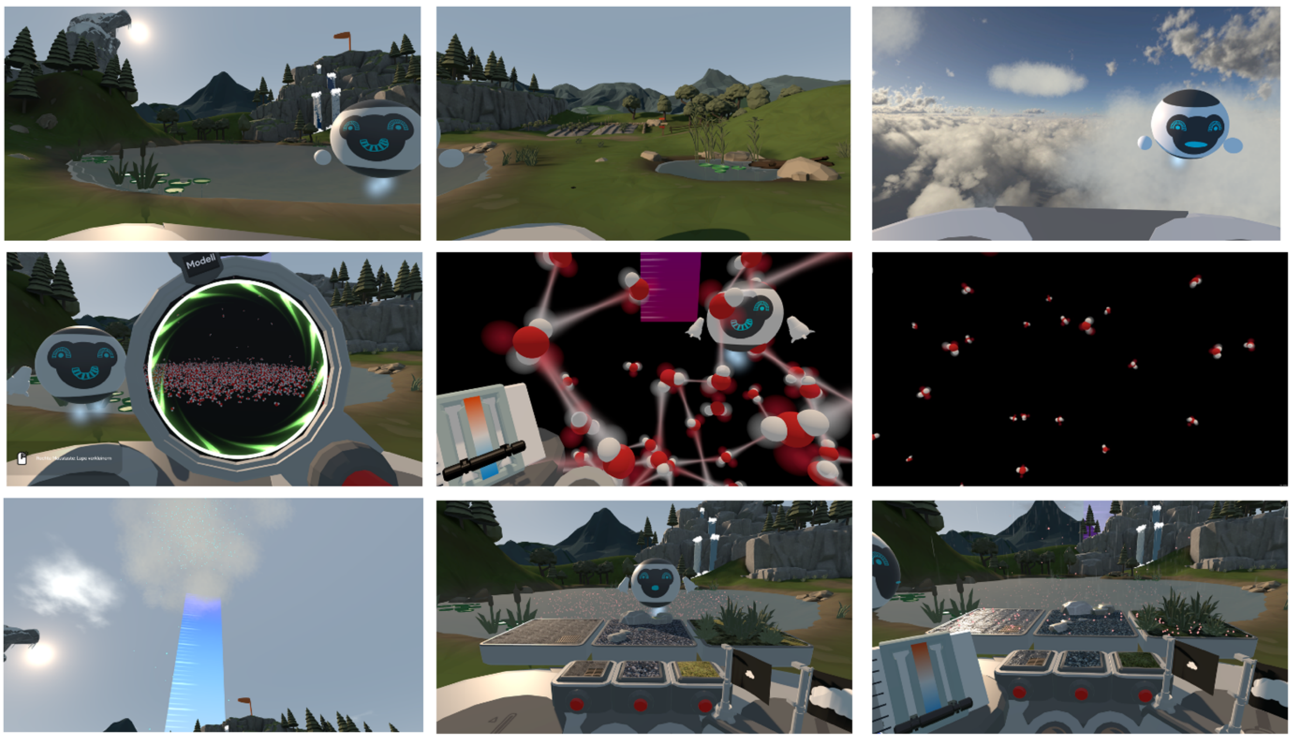


Figure 1. Screenshots of the virtual environment

We manipulated the *immersive* and *interactive* aspects of the virtual environment about the water cycle (2 × 2 design) with four conditions. In the *high immersion-high interaction* condition, the children worked with immersive VR headsets and were able to interact with the virtual world using motion controllers. As examples, children were able to mark water molecules, make wind with their hands to accelerate the formation of water droplets, change the temperature of the environment, and observe the consequences of their actions. In the *high immersion-low interaction* condition, the children also worked with immersive VR headsets, but it was the avatar ONOS who carried out all the actions. In other words, it was the avatar who marked the water molecules, made wind accelerate the formation of water droplets, and changed the temperature of the environment. In this condition, the children were immersed in the virtual world; however, they were spectators of the avatar's actions. In the *low immersion-high interaction* condition, the children worked in with laptops and were able to interact with the virtual world using a mouse. In the *low immersion-low interaction* condition, the children also worked in a computer situation, but they observed the avatar executing the actions. This last condition resembled a video condition.

Materials

Knowledge test

To measure the *learning outcomes*, we developed 38 items based on the addressed learning objectives. Measurement took place before the intervention, immediately after the intervention, and eight weeks after the intervention (same 38 items; item example: "The water droplets formed in the clouds always remain the same size"). The items were to be answered on 3-point Likert scales ("I agree," "I disagree," or "I don't know"). We coded correct answers as 1 and incorrect, as well as "I don't know" answers, as 0. The 38 items are reported in Appendix B. Overall learning outcome scores (proportion of correct answers) for the three measurement points (pre-, post-, and follow-up tests) were then generated.

Cognitive and affective variables

To assess the *cognitive and affective aspects*, during the second, fourth, sixth, eighth, and tenth teaching units, immediately after the 10 minutes spent in the virtual environment, presence, motion sickness, cognitive load, physical load, and liking were assessed. The variables were always assessed in the same order listed here. We also assessed other explorative variables (e.g., epistemic emotions) that were not further considered in this publication.

Presence. We adapted a single pictorial item to assess spatial presence ("How deeply are you immersed in the virtual world right now?") to be answered on a 5-point Likert scale (1 = "not immersed at all" to 5 = "completely immersed"). The item was inspired by the well-validated pictorial presence scale of Weibel et al. (2015). The pictorial item can be found in Appendix C. A mean response score for the five measurements was created, and Cronbach's alpha was determined to be 0.879.

Motion sickness. We assessed motion sickness with a verbal item ("Did you feel nauseous or dizzy") to be answered on a 5-point Likert scale (1 = "not at all" to 5 = "very much"). A mean response score for the five measurements was created, and Cronbach's alpha was 0.857.

Cognitive load. We created a single pictorial item to assess cognitive load ("How much did you need to use your head? E.g., thinking, deciding, searching, remembering, etc.") to be answered on a 5-point Likert scale (1 = "very little” to 5 = “a lot”). The pictorial item is reported in Appendix C. A mean response score for the five measurements was created, and Cronbach’s alpha was 0.835.

Physical load. We further created a single pictorial item to assess physical load (“How much did you need to use your body? E.g., dexterity, strength, etc.”) to be answered on a 5-point Likert scale (1 = “very little” to 5 = “a lot”). The pictorial item is reported in Appendix C. A mean response score for the five measurements was created, and Cronbach’s alpha was 0.904.

Liking. To assess the children’s satisfaction, we used a pictorial scale (“How did you find the environment?”; see Appendix C) to be answered on an 11-point Likert scale (0 = “very bad” to 10 = “very good”). A mean response score for the five measurements was created, and Cronbach’s alpha was 0.885.

Equipment

In the high immersion conditions, we used 12 VR headsets (Oculus Quest 2). In the low immersion conditions, we used 12 laptops (HP ProBook x360 435 G8). The questionnaires were delivered in paper-and-pencil format.

Procedure

Of the nine participating classes, two were attributed to each of the four conditions, and one served as a control class. Except for the type of device and the possibility of manipulation in the environment, the procedure in the eight intervention classes was identical. In the first lesson, the children filled in the questionnaires about the water cycle (i.e., knowledge test). In the second lesson, the children were familiarized with the digital equipment. In the following 11 lessons, the intervention took place, which lasted four weeks. The intervention combined classical teaching (with experiences in the real world) and the virtual environment. Classical teaching was provided to the entire class, whereas the children had to take turns in the virtual environment. While half of the class was using the virtual technology, the other half was completing paper-and-pencil exercises or drawings (and vice versa). In each of the 11 lessons, the children visited the virtual world for approximately 10 minutes. During the intervention, cognitive and affective variables were assessed every second. After the 11 lessons, the post-test took place (i.e., assessment of learning outcomes). To measure the consolidation of knowledge, the same test was carried out eight weeks after the end of the intervention. The procedure used in the control class was the same as the one used in the intervention classes regarding the assessment of learning outcomes; however, the children in the control did not receive any intervention.

Analytical approach

To test our main hypothesis (i.e., the impact of immersion and interaction on learning outcomes), we computed an immersion (low vs. high) x interaction (low vs. high) x measurement point (pre-, post-, follow-up test) analysis of variance, with the learning outcomes as the dependent variable. Given that we included the measurement point (pre-, post-, follow-up test) in the model, the effects of interest were the interaction terms. When Mauchly's test indicated that sphericity was not met ($p < .10$), we used the Huynh–Feldt correction to adjust the degrees of freedom. We reported *partial* η^2 as a measure of effect size. To unpack the significant results, we used *t*-tests with Bonferroni correction and reported Cohen's *d* as a measure of effect size. Furthermore, we analyzed presence, motion sickness, cognitive load, physical load, and children's satisfaction (liking). We computed five separate immersion (low vs. high) x interaction (low vs. high) analyses of variance, with the aforementioned measures as dependent variables. Again, we reported partial η^2 as a measure of effect size. All analyses were computed with jamovi (The jamovi project, 2022) and R (2022), and the dataset is available on OSF (https://osf.io/pdwgv/).

Results

Learning outcomes

Correct answers were coded with 1, whereas wrong answers and "don't know" responses were coded with 0. We averaged the learning outcome (i.e., amount of knowledge about the water cycle) over the 38 items per testing session (pre-, post-, and follow-up tests). To test whether there were condition differences in the learning outcomes, we carried out an analysis of variance with immersion (low vs. high) and interaction (low vs. high) as between-subjects factors, and time point of measurement (pre, post, follow-up) as a within-subjects factor. The main finding of this analysis is the significant interaction between immersion and the time point of measurement ($F(1.581, 221.307) = 6.999$, $p = .003$, partial $\eta^2 = .048$), as depicted in Figure 3. We further explore this significant interaction in the following paragraph. The main effect of time point of measurement also turned out to be significant ($F(1.581, 221.307) = 360.624$, $p < .001$, partial $\eta^2 = .720$), illustrating significant differences (*p-values* < .001) between the pre- ($M = 0.359$; $SD = 0.134$), post- ($M = 0.650$, $SD = 0.151$), and follow-up tests ($M = 0.619$, $SD = 0.154$). All other effects were non-significant, which included the three-way interaction between immersion × interaction × timepoint ($F(1.581, 221.307) = 0.698$, $p = .466$, partial $\eta^2 = .005$), the two-way interaction between interaction and time point ($F(1.581, 221.307) = 0.698$, $p = .466$, partial $\eta^2 = .005$), the two-way interaction between immersion and interaction ($F(1, 140) = 2.843$, $p = .094$, partial $\eta^2 = .020$), and the main effects of immersion ($F(1, 140) = 1.801$, $p = .182$, partial $\eta^2 = .013$) and interaction ($F(1, 140) = 3.163$, $p = .078$, partial $\eta^2 = .022$).

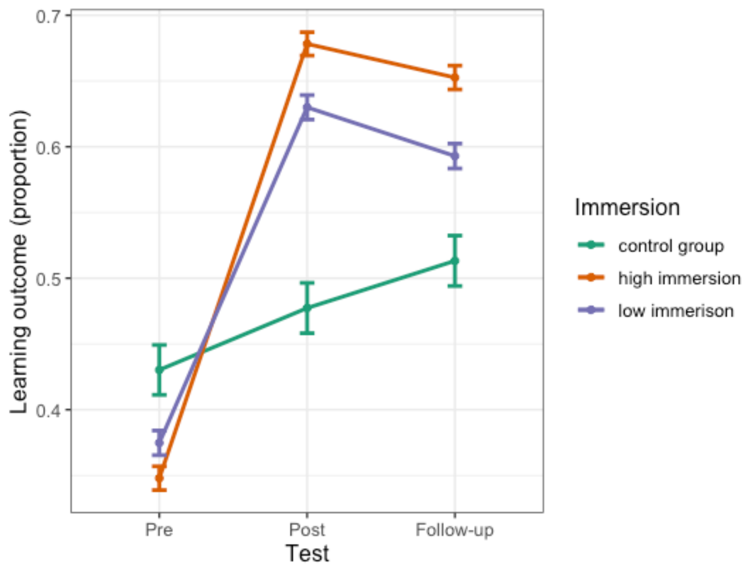


Figure 3. Learning outcome values as a function of immersion (low vs. high) and measurement point (pre-, post-, and follow-up tests). The control group is also plotted. Learning outcome values vary between 0 and 1 (proportions), and higher values represent better performance. Error bars represent one standard error of the mean.

To unpack the significant immersion x timepoint interaction, we computed difference variables between learning outcome in the post-test (immediately after) and learning outcome in the pre-test, and between learning outcome in the follow-up test (eight weeks after) and in the pre-test. This procedure was undertaken to consider baseline knowledge, as assessed in the pre-test. We then computed two independent *t*-tests to compare these different variables between the low immersion and high immersion conditions. The first *t*-test with the difference variable between pre-test and post-test (immediately after the intervention) revealed a significant difference between low immersion and high immersion conditions ($t(142) = 2.724$, $p = .007$, $d = 0.454$), because of the smaller difference between pre-test and post-test in the low immersion condition ($M = 0.257$, $SD = 0.169$) when compared to the high immersion condition ($M = 0.329$, $SD = 0.151$). The second *t*-test with the difference variable between the pre-test and follow-up test (eight weeks after the intervention) revealed a significant difference between the low immersion and high immersion conditions ($t(142) = 3.022$, $p = .003$, $d = 0.504$), which was attributed to the smaller difference between the pre-test and post-test in the low immersion condition ($M = 0.219$, $SD = 0.164$) when compared to the high immersion condition ($M = 0.301$, $SD = 0.160$). These *p*-values were below 0.025 (adjusted Bonferroni alpha level). This result indicates that working with immersive VR (via headsets) enhanced the learning outcomes when compared to non-immersive VR (computer situation) in the post-test directly after the intervention, but also in the follow-up test eight weeks after the intervention, when compared to the pre-test before the intervention. It seems that immersive VR has an advantage over non-immersive VR in teaching the water cycle in terms of learning outcomes and their consolidation.

As for the control group (see Figure 3), we compared the results (knowledge assessment ranging from 0 to 1) in the three sessions with chance level (0.5) by conducting one-sample *t*-tests. The performance was below chance level in the pre-test (M = 0.416, SD = 0.131, $t(20)$ = 2.926, p = .008, d = 0.639), and at chance level in the post-test (M = 0.485, SD = 0.126, $t(20)$ = 0.545, p = .592, d = 0.119) and follow-up test (M = 0.523, SD = 0.128, $t(20)$ = 0.805, p = .430, d = 0.176[2]). A visual comparison between the control group and the intervention groups (see Figure 3) illustrates that all the intervention groups had a learning benefit when compared to the control group.

Cognitive and affective variables

In the following table (see Table 1), we report the results for presence, motion sickness, cognitive load, physical load, and liking divided by condition. Below, we also provide the results of the analyses to test the condition differences.

	Presence (from 1 to 5)	Motion sickness (from 1 to 5)	Cognitive load (from 1 to 5)	Physical load (from 1 to 5)	Liking (from 0 to 10)
High immersion High interaction	4.043 (0.852)	1.399 (0.729)	1.621 (0.539)	2.331 (0.835)	8.422 (1.712)
High immersion Low interaction	4.039 (0.720)	1.485 (0.621)	1.452 (0.538)	1.444 (0.489)	8.400 (1.330)
Low immersion High interaction	3.430 (0.762)	1.127 (0.529)	1.782 (0.889)	1.525 (0.875)	8.234 (1.578)
Low immersion Low interaction	3.485 (0.738)	1.139 (0.405)	1.834 (0.646)	1.249 (0.377)	8.365(1.502)

Table 2. Descriptive statistics (means and standard deviations are reported in brackets) for presence, motion sickness, cognitive load, physical load, and liking. Categories include high immersion and high interaction (n = 37), high immersion and low interaction (n = 36), low immersion and high interaction (n = 37), and low immersion and low interaction (n = 33).

Presence

The children's overall mean rating for presence was 3.755 (SD = 0.817) on a Likert scale ranging from 1 to 5, indicating a generally high level of presence across all conditions. To test whether there were condition differences in presence, we carried out an analysis of variance, with immersion (low vs. high) and interaction (low vs. high) as

[2] It is surprising that children in the control group seemed to learn over time. We thus asked the teacher if she respected the instructions not to teach the water cycle over this period. However, it turned out that she taught the topic of weather and in part the water cycle between the second and third measurement points. What is more, the teacher reported that children asked some questions about the water cycle after the first measurement point. Additionally, the sample size of the control group was relatively small (n = 18), which may have increased the variance of the sample mean. Therefore, caution should be taken when interpreting the results of the control group.

between-subject factors. As expected, we found a significant main effect of immersion (F(1, 139) = 20.436, p < .001, partial η^2= .128), with presence being higher in the high immersion conditions (M = 4.041, SD = 0.784) than in the low immersion conditions (M = 3.450, SD = 0.746). All other effects yielded non-significant results, including the two-way interaction between immersion and interaction (F(1, 139) = 0.053, p = .819, partial η^2< .001), as well as the main effect of interaction (F(1, 139) = 0.037, p = .847, partial η^2 < .001).

Motion sickness

Motion sickness was found to be low across conditions, with an overall mean of 1.290 (SD = 0.602) on a Likert scale ranging from 1 to 5. To test whether there were condition differences in motion sickness, we carried out an analysis of variance, with immersion (low vs. high) and interaction (low vs. high) as between-subject factors. As anticipated, we found a significant main effect of immersion (F(1, 139) = 9.857, p = .002, partial η^2= .066), with motion sickness being higher in the high immersion conditions (M = 1.442, SD = 0.675) than in the low immersion conditions (M = 1.133, SD = 0.472). All other effects yielded non-significant results, which included the two-way interaction between immersion and interaction (F(1, 139) = 0.137, p = .712, partial η^2= .001), as well as the main effect of interaction (F(1, 139) = 0.246, p = .621, partial η^2 = .002).

Cognitive load

With an overall mean across conditions of 1.669 (SD = 0.680) on a Likert scale ranging from 1 to 5, cognitive load was low. To test whether there were condition differences in cognitive load, we carried out the same analysis of variance, with immersion (low vs. high) and interaction (low vs. high) as between-subject factors, but with cognitive load as the dependent variable. We found a significant main effect of immersion (F(1, 139) = 5.870, p = .017, partial η^2= .041), with cognitive load being lower in the high immersion conditions (M = 1.538, SD = 0.541) than in the low immersion conditions (M = 1.807, SD = 0.779). The other effects turned out to be non-significant, including the two-way interaction between immersion and interaction (F(1, 139) = 0.970, p = .326, partial η^2= .007), as well as the main effect of interaction (F(1, 139) = 0.271, p = .604, partial η^2 = .002).

Physical load

On a Likert scale ranging from 1 to 5, the children gave a generally low rating for physical load across all conditions, with an overall mean of 1.649 (SD = 0.797). To analyze the impact of immersion (low vs. high) and interaction (low vs. high) as between-subject factors on physical load, we conducted a similar analysis of variance, with physical load as the dependent variable. This analysis revealed a significant interaction between immersion and interaction ($F(1, 139) = 7.053$, $p = .009$, partial $\eta^2 = .048$), as well as a main effect of immersion ($F(1, 139) = 18.895$, $p <$ .001, partial $\eta^2 = .120$) and a main effect of interaction ($F(1, 139) = 25.535$, $p < .001$, partial $\eta^2 = .155$), as depicted in Figure 4. Physical load was greater in the high immersion-high interaction condition when compared to the other conditions.

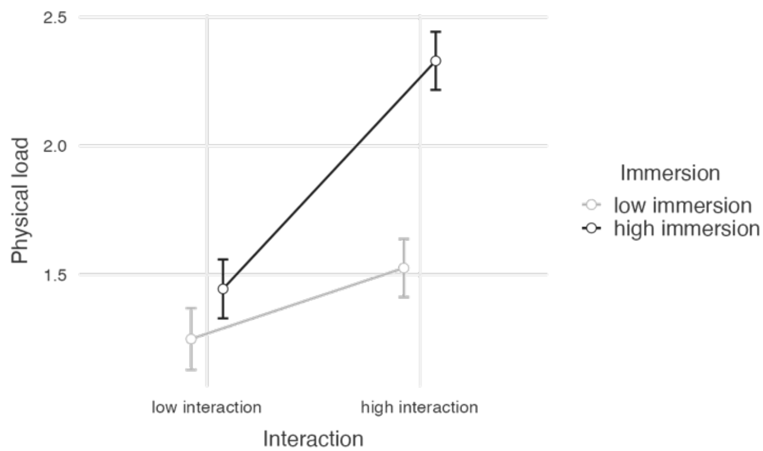


Figure 4. Physical load (y-axis) as a function of immersion (low vs. high) and interaction (low vs. high). Physical load varies between 1 and 5 (Likert scale). Error bars represent one SEM.

Liking

With an overall mean across conditions of 8.354 ($SD = 1.525$) on a Likert scale ranging from 0 to 10, liking was generally high. In this step, we carried out the same analysis of variance, with immersion (low vs. high) and interaction (low vs. high) as between-subject factors and liking as the dependent variable. This analysis revealed no significant effects. Specifically, the two-way interaction between immersion and interaction ($F(1, 139) = 0.088$, $p = .767$, partial $\eta^2 =$ .001), the main effect of immersion ($F(1, 139) = 0.186$, $p = .667$, partial $\eta^2 = .001$), and the main effect of interaction, $F(1, 139) = 0.045$, $p = .832$, partial $\eta^2 < .001$) were not significant.

Discussion

The main aim of this study was to investigate whether the use of VR devices can enhance the learning of 11–12-year-olds in a primary school classroom setting on the topic of the water cycle. Specifically, we aimed to assess the relative contributions of *immersive* and *interactive* aspects of VR to the learning process. We found that high immersion VR compared to low immersion computer conditions enhanced understanding of the water cycle in primary classes. Children who worked with headsets showed a better learning outcome immediately and eight weeks after the intervention when compared to children who worked with computers. This effect was observed even though most of the teaching was carried out using conventional methods. This demonstrates that the use of VR technology in standard school settings is effective and feasible. Moreover, this result fits well with previous research investigating the use of VR in science teaching (e.g., Chen et al., 2020; Liu et al., 2020, 2022) . *Being in the virtual world* seems to foster learning more than observing the virtual world on a computer. Thus, engaging in a naturalistic manner with the virtual environment may facilitate learning, as suggested by previous research (Makransky & Mayer, 2022). Motivational components of immersive VR, such as an increased sense of presence, could also contribute to enhanced learning (see, e.g., Makransky et al., 2019).

Contrary to our predictions, the interactive elements did not seem to play a role in our study. This is surprising, because handling is an important source of learning (Carey, 2009; Collins et al., 1989). Actions promote learning for two reasons. First, they provide intervention knowledge (i.e., they inform about environmental changes associated with the action), and second, they allow for intervention knowledge to be grounded in multimodal bodily experiences, thus promoting strong representations. A possible explanation for the results of our study can be found in the way we manipulated the interactive aspects of VR. In the high interaction conditions, the children manipulated the virtual world by themselves (e.g., changing the temperature), whereas in the low interaction conditions, the children observed the avatar carrying out the actions. Although children in the low interaction conditions did not receive any proprioceptive action information, they had access to intervention knowledge via observation of the avatar. Learning by enactment has not always been shown to outperform observational learning, and it is indeed uncertain whether observing is linked to enactment, and whether children in low interactive conditions might actually be simulating actions (e.g., Koßmann et al., 2023). Moreover, research on VR has shown that direct interaction with virtual environments offers the benefit of providing first-person experiences; however, it can also impose cognitive and physical demands, especially since children may not be used to interacting with virtual environments (Parong & Mayer, 2018). Indeed, some previous research has found contradictory results regarding the impact of interactive elements in VR on learning. For example, Makransky et al. (2019) found that immersive VR under specific conditions (e.g., practicing laboratory activities) elicited a stronger feeling of presence and more liking but less learning. In this sense, it is possible that we did not find a difference between our high and low interactive conditions because the benefit of handling was canceled out by the constraints of the technology. Indeed, although *cognitive load* was significantly lower in the immersive conditions compared to the computer-based situation, our study revealed that *physical load* was higher in the immersive conditions, particularly in the high immersion-high interaction condition. In other words, observing an avatar carrying out the actions (low interactive conditions) compared to carrying out the actions oneself (high interactive conditions) might have reduced the demands of the task and thus freed attentional resources that were then allocated to the material to be learned.

Finally, it is important to note that the intervention durations for immersion and interaction differed in our study. Immersion was consistently provided during the children's stay in the virtual world, whereas interaction occurred only during certain phases of the intervention. This difference in the timing of the immersion and interaction parts may have contributed to the more robust impact of immersion on learning outcomes. It is worth pointing out that our study found *no significant difference* between the two interactive conditions (low vs. high interaction), and what is more, it was the children who expressed a desire for more interaction in the non-interactive conditions. Indeed, a qualitative analysis of the children's responses to two open-ended questions about criticism of the virtual environment and ideas for improvement showed that a recurring theme was a desire for more activity in the different conditions. In the high immersion-high interaction condition, four out of 38 children (10.5%) mentioned a desire for more interactive activities. In the high immersion-low interaction condition, 25 children out of 36 (69.4%) reported a desire for more interactive activities. In the non-immersive environments, the pattern was similar: in the low immersion-high interaction condition, only one child out of 37 (2.7%) reported a desire for more activity, whereas in the low immersion-low interaction condition, 18 children out of 36 (50.5%) reported such a desire. The desire for more activity was higher in the low interaction conditions when compared to the high interaction conditions, thus suggesting that interactive elements have an impact, at least on the motivational components of the learning processes. Nevertheless, this analysis and interpretation are explorative and warrant further investigation.

Our study has several limitations that should be considered when interpreting the results. While integrating VR into actual teaching has applied relevance, our study lacks experimental control (e.g., random attribution of children to the conditions or individual testing in quiet rooms) due to the longitudinal design (11 teaching units) and classroom setting. Moreover, the applied setting made it difficult to investigate the underlying mechanisms that could clarify the reasons for the effects we found. Due to our rather small sample size, we were unable to compute more complex statistical analyses, such as mediation analyses. A further limitation is the straightforward manipulation of immersion (headsets vs. computers) versus the more complex manipulation of interaction, which required numerous discussions in our team about which interactive elements to implement in the application. Lastly, our results are specific to 11–12-

year-olds learning about the water cycle. Therefore, future studies should investigate whether our findings can be generalized to other age groups and materials.

Conclusion. VR is rapidly gaining interest in educational settings; however, there are important gaps in the literature concerning the educational value of VR and its potential to stimulate learning in specific age groups. What is more, few studies have investigated the specific impact of VR-related features on learning outcomes, as well as the long-term retention of knowledge learned from immersive VR compared to other digital media. Our study found that a learning environment with both headsets and computer-based VR had a positive impact on learning outcomes. We focused on a comparison of immersive and interactive VR elements and found a medium effect of immersion on learning outcomes (Cohen's *d* of 0.454 and 0.504 for immediate learning outcomes and delayed learning outcomes, respectively). In the specific context of our study (11–12-year-olds learning about the water cycle), it seemed that the experience of *being there* in the virtual world was an important source of learning. Given that we kept everything constant, except for the digital intervention, we believe that the effect indicates practical relevance for the field.

In summary, VR provides access to unobservable processes and structures; it has the power to connect us to inaccessible aspects that would otherwise be beyond our direct experience. Our results suggest that immersive VR is a promising tool for enhancing science learning in primary schools. This finding is reinforced by the fact that VR also appeared to be easily applicable in a primary school setting, that children generally liked the intervention, and that motion sickness was low. As technology continues to advance, we can expect VR to play an increasingly important role in science education, thus enabling children to explore and learn about the world in ways that were previously unimaginable.

References

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology*. Wiley.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Chen, J.-C., Huang, Y., Lin, K.-Y., Chang, Y.-S., Lin, H.-C., Lin, C.-Y., & Hsiao, H.-S. (2020). Developing a hands-on activity using virtual reality to help students learn by doing. *Journal of Computer Assisted Learning*, *36*(1), 46–60. https://doi.org/10.1111/jcal.12389

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship teaching and the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453-494). Lawrence Erlbaum Associates.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Gopnik, A. & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*, 1085–1108. https://doi.org/10.1037/a0028044

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, *7*, 75–83. https://doi.org/10.1111/j.1365-2729.1991.tb00230.x

Koßmann, C., Straatmann, T., Mueller, K., & Hamborg, K.-C. (2023). Effects of enactment in virtual reality: A comparative experiment on memory for action. *Virtual Reality, 27,* 1025-1038. https://doi.org/10.1007/s10055-022-00701-y

Liu, R., Wang, L., Koszalka, T. A., & Wan, K. (2022). Effects of immersive virtual reality classrooms on students' academic achievement, motivation and cognitive load in science lessons. *Journal of Computer Assisted Learning*, *38*(5), 1422–1433. https://doi.org/10.1111/jcal.12688

Liu, R., Wang, L., Lei, J., Wang, Q., & Ren, Y. (2020). Effects of an immersive virtual reality-based classroom on students' learning performance in science lessons. *British Journal of Educational Technology*, *51*(6), 2034–2049. https://doi.org/10.1111/bjet.13028

Makransky, G., & Mayer, R. E. (2022). Benefits of taking a virtual field trip in immersive virtual reality: Evidence for the immersion principle in multimedia learning. *Educational Psychology Review, 34*(3), 1771–1798. https://doi.org/10.1007/s10648-022-09675-4

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. https://doi.org/10.1016/j.compedu.2019.02.002

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). 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

Maltese, A. V., & Tai, R. H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education*, *32*(5), 669-685,. https://doi.org/10.1080/09500690902792385

Medin D., L. & Rips L. J. (2005). Concepts and categories: Memory, meaning, and metaphysics. In K. J. Holyoak and R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 37–73). Cambridge University Press.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, *140*, 103603. https://doi.org/10.1016/j.compedu.2019.103603

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

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, *37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Pellas, N., Mystakidis, S., & Kazanidis, I. (2021). Immersive virtual reality in K-12 and higher education: A systematic review of the last decade scientific literature. *Virtual Reality*, *25*(3), 835–861. https://doi.org/10.1007/s10055-020-00489-9

Potvin, P., Malenfant-Robichaud, G., Cormier, C., & Masson, S. (2020). Coexistence of misconceptions and scientific conceptions in chemistry professors: A mental chronometry and fMRI study. *Frontiers in Education*, *5*, 542458. https://doi.org/10.3389/feduc.2020.542458

Reinfried, S., & Tempelmann, S. (2014). Wie Vorwissen das Lernen beeinflusst – Eine Lernprozessstudie zur Wissenskonstruktion des Treibhauseffekt-Konzepts. *Zeitschrift für Geographiedidaktik*, *42*(1), 31–56.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209–215. https://doi.org/10.1016/j.cognition.2012.04.005

Vaiopoulou, J., & Papageorgiou, G. (2018). Primary students' conceptions of the Earth: Re-examining a fundamental research hypothesis on mental models. *Preschool and Primary Education*, *6*(1), 23–34. https://doi.org/10.12681/ppej.14210

Weibel, D., Schmutz, J., Pahud, O., & Wissmath, B. (2015). Measuring spatial presence: Introducing and validating the pictorial presence SAM. *Presence: Teleoperators and Virtual Environments*, *24*(1), 44–61. https://doi.org/10.1162/PRES_a_00214

Appendix A

Table describing the 11 units (including their content, objectives, and interactive elements) of the virtual environment.

We had six general modules, each of which was divided in two units, apart from the last module (with only one unit). In low interactive conditions, it is the avatar ONOS that is executing the interactive elements instead of the child.

Unit	Levels	Content	Objectives	Interactive elements
1. On the trace of water	1.1 SOS from outer space	The avatar ONOS and its history, as well as water in different states of matter. The question of the origin of clouds when rain begins is raised.	Children familiarize themselves with the VR environment and practice handling the controllers (headset conditions) or the mouse (computer conditions).	Virtual hands can be used to indicate and highlight the presence of detected water within the environment (e.g., water of the lake or snow on the mountains).
	1.2 A cycle?	Temperature as a driver of evaporation.	Children can explain the effect of temperature on evaporation.	Children can test their hypotheses about evaporation by *marking* individual water particles in the lake with luminous paint, then *manipulating the temperature with the temperature controller,* and observing the marked particles.
2 The world of water particles	2.1 The world of water particles I	Particle motions and bonds in the three states of matter.	Children can explain how particles behave depending on the temperature, and how the temperature can cause changes in states.	Children can use the temperature controller in the model level to observe the impact of temperature on the particles while being surrounded by the particle model.
	2.2 The world of water particles II	Relationship between bonds and states of matter.	Children can explain the process of water particles being released from bonds.	A particle can be released from a bond via the children's hands. The energy of the particle can be felt as a vibration in immersive VR.
3 Clouds and how they are formed	3.1 Clouds and how they are formed I	Clouds consist of droplets. Condensation depending on temperature and condensation nuclei.	Children can explain how clouds consist of (liquid) droplets, and that condensation depends on the temperature and causes condensation nuclei.	As in Unit 1, water particles can be marked with luminous paint. At the research station in the clouds, children can use a magnifying glass to witness how the illuminated spots gather around a condensation nucleus.
	3.2 Clouds and how they are formed II	Condensation requires condensation nuclei.	Children can explain how condensation occurs on a solid substance (condensation nucleus), and that condensation is more likely to happen in cooler temperatures.	A condensation nucleus can be captured, with which water particles can be "collected" to create a new droplet.
4 It is raining	4.1 Raindrops arise	Droplets continue to aggregate and become larger until they reach a critical weight and fall.	Children can explain how small water droplets become large water drops that then fall.	Through the magnifying glass, children can observe how small water droplets combine to form larger drops. Children can simulate wind *by waving their*

				hands and thus accelerate the process.
	4.2 Self-made rain	Relationship between temperature and evaporation/condensation and the link to the temperature gradient.	Children can explain why clouds form faster at higher temperatures, although low temperatures are required for cloud formation.	The temperature controller can be used to test at which temperatures clouds form most quickly.
5 Wind, clouds, and soil	5.1 Wind and clouds: A great couple	Displacement of gaseous and condensed water in the atmosphere.	Children can explain how air masses are displaced and what this means for the water cycle.	The temperature controller is used to create a cloud as quickly as possible. By generating wind through hand waving, the cloud can be directed toward fields, resulting in the watering of the crops.
	5.2 The soil is crucial: Evaporation	Water percolates differently on different surfaces.	Children can explain what happens to water when it falls to the ground.	Material samples from different surfaces (e.g., asphalt, gravel, or grassland) should be sorted according to their infiltration capabilities.
6 Masters of the water cycle	Masters of the water cycle	The water cycle and its processes.	Children can explain the different processes in the water cycle using examples.	Free testing and practice are available in all locations (meso-level, model level, and in the clouds) within the VR world.

Appendix B

List of 38 items used during the pre-test, post-test, and follow-up test to assess learning outcomes. We report in parentheses whether the item is correct (C) or incorrect (I).

1) The particle model is an idea proposed by researchers, and it is not known whether all its considerations are correct. (C)
2) Researchers know exactly what water particles look like. (I)
3) There are objects that are not made up of particles. (I)
4) Water is made up of particles that are smaller than water droplets. (C)
5) The warmer it gets, the more the water particles move. (C)
6) When it gets colder, the water particles make bigger movements. (I)
7) When water is frozen into ice, the water particles are so close together that they touch each other. (I)
8) When water is frozen, the water particles have a fixed place. (C)
9) When water is liquid, the water particles are closer together than when water is frozen. (C)
10) When water is liquid, each water particle has a fixed place. (C)
11) When water evaporates, the water particles are much farther apart than when it is liquid or frozen. (C)
12) When water has evaporated, the water particles stay in one place. (I)
13) Water evaporates because the water particles move so little that no new bonds can form between them. (I)
14) The stronger the movement of the water particles, the more liquid water evaporates. (C)
15) Only when the water particles of evaporated water move less strongly can bonds form between them. (C)
16) Droplets form from evaporated water when the movement of the water particles increases. (I)
17) Cooling water has no effect on how much it evaporates. (I)
18) Increasing the temperature of water makes it even more liquid. (C)
19) Liquid water does not need to boil in order to evaporate. (C)
20) Only the water at the surface of a lake evaporates. (C)
21) When the water of a lake evaporates, tiny water droplets rise into the sky. (I)
22) Less water evaporates when the sun shines than when it does not. (I)
23) Clouds are mainly composed of water that is not in liquid form. (I)

24) Clouds form when evaporated water condenses into small droplets. (C)

25) Evaporated water condenses into droplets more easily near the earth's surface than higher up in the sky. (I)

26) For evaporated water to condense into liquid in the air, it needs other particles, such as pollen. (C)

27) The water droplets formed in clouds remain the same size. (I)

28) In clouds, water droplets collide and merge into larger drops. (C)

29) Once a water droplet has formed in a cloud, it falls to the ground as a raindrop. (I)

30) The water droplets formed in clouds must reach a certain size before they fall down. (C)

31) If the sun shines less often, there is usually more rain. (I)

32) Sunlight has a great influence on whether it rains or not. (C)

33) Water from a lake in Spain cannot fall as rain in Valais. (I)

34) If you spill water outside, it is possible that someone will eventually get rained on by it. (C)

35) Clouds are more likely to form during prolonged cold weather, as evaporated water can then condense more easily. (I)

36) Clouds are more likely to form when the temperature on the earth's surface is high. (C)

37) If the earth were to get warmer forever, there would eventually be less rain. (I)

38) If the earth were to get colder forever, there would eventually be fewer floods. (C)

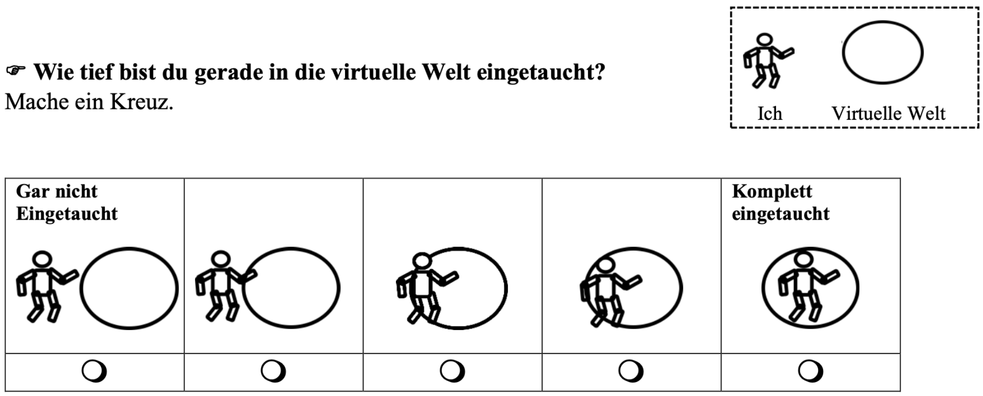
Appendix C

Examples of pictorial items used to assess presence, cognitive load, physical load, and liking during the intervention.

We report the original German versions and provide a translation.

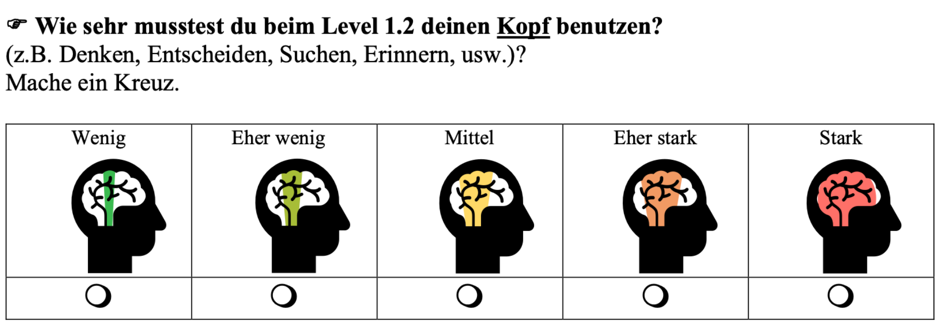
Presence

The single item "How deeply are you immersed in the virtual world right now?" is to be answered on a 5-point Likert scale (1 = "not immersed at all" to 5 = "completely immersed").



Cognitive load

The single item "How much did you need to use your head? E.g., thinking, deciding, searching, remembering, etc." is to be answered on a 5-point Likert scale (1 = "very little" to 5 = "a lot").



Physical load

The single item "How much did you need to use your body? E.g., dexterity, strength, etc." is to be answered on a 5-point Likert scale (1 = "very little" to 5 = "a lot").

☞ **Wie sehr musstest du beim Level 1.2 deinen <u>Körper</u> benutzen?**
(z.B. Geschicklichkeit, Kraft, usw.)?
Mache ein Kreuz.

Wenig	Eher wenig	Mittel	Eher stark	Stark
				
❍	❍	❍	❍	❍

Liking

The single item “How did you find the environment?” was to be answered on a 11-point Likert scale (0 = “very bad” to 10 = “very good”).

☞ **Wie fandest du das Level 1.2?**
Mache ein Kreuz auf dem Zahlenstrahl.