Betty Tärning*
Division of Cognitive Science
Lund University, Sweden

Agneta Gulz
Division of Cognitive Science
Lund University, Sweden

and

Department of Computer Science
Linköping University, Sweden

Jens Nirme
Division of Cognitive Science
Lund University, Sweden

Department of Design Sciences
Lund University, Sweden

and

Lund University Humanities Lab,
Lund University, Sweden

Presence, *Vol. 33, Winter 2024*, 367–388
https://doi.org/10.1162/PRES_a_00420

Bridging the Headset: Engagement, Collaboration, and Learning in and around Virtual Reality

Abstract

To study engagement and learning, we developed an asymmetric collaborative version of an existing VR-experience about environmental sustainability, enabling student teams of three, with only one wearing a headset at any time, to collaborate in three different virtual reality subgames. We compared this collaborative version to a non-collaborative version in a between-subjects experimental study ($n = 20$), finding that student teams using the collaborative version interacted to a larger extent while performing the subgames. The students in these teams self-reported a higher level of engagement, and observations suggested that interactions extended beyond necessary instructions. We did not find any significant differences when it comes to learning measured by scores on pre- and post-knowledge tests, but quantitative analysis of responses indicates that the VR-experience affected students overall and a qualitative analysis of an open question suggested a particular effect on the collaborative teams. Logged performance data indicated that collaborative teams managed to solve the tasks involved in the subgames at least as efficiently as the teams using the non-collaborative version.

1 Introduction

Virtual reality (VR) and other related technologies, such as augmented reality (AR), have a potential to bring a dimension to learning that conventional methods cannot. With VR, educational content can be projected all around the learner in a unique way, which can be very useful from an educational point of view. For example, grasping the relations between planets in the universe can be hard. A picture or a film of how the planets relate to one another can of course provide relevant information, but actually being in space (if only in VR) adds another dimension and a sense of being there. It provides an opportunity to interact with objects and triggers emotions (Markowitz & Bailenson, 2023).

Rapid and continuous development of immersive VR technology have increased the accessibility of VR. However, how and when to use it is still debated and it is rare that individual school classes have a set of VR-headsets for all students. A traditional approach has been to have one student wear the head-mounted display (HMD) while the others act as passive bystanders. Recently, it has become more common to work together in VR, either as a group where all have a VR-headset or with some students working together, for example via

*Correspondence to betty.tarning@lucs.lu.se.

a tablet, with the person wearing the HMD. Collaboration in VR might be realized in different ways and whether it leads to increased engagement and/or learning is not yet mapped out.

In previous work, collaborative experiences are most often defined by instructions to collaborate rather than by interface differences or with parallel development of a non-VR application, for example, running on an external tablet.

The work presented here represents a different approach and describes how an existing educational VR application was adapted to encourage collaboration by relatively simple measures, with the basic idea of redistributing the display of some information that facilitates task performance. Members of small teams not currently wearing HMDs (*out-VR*) were engaged in virtual tasks by showing relevant information on a separate screen. Teams using this collaborative version are compared to teams that use a version where team members wearing the headset (*in-VR*) in principle can perform the tasks alone and the other team members (*out-VR*) are shown the same view as is shown in the HMD on the separate screen. We compared students' engagement, learning, and task performance between the two different conditions in a between-subject designed study, finding evidence of higher engagement in collaborative teams, and a general learning effect (irrespective of group). Thus, with simple measures we could make more students engaged in a VR experience originally designed to be performed individually.

2 Background

It has been established that VR can be used to create engagement and allow for experiences outside the ordinary. However, when leaving the amusement arena and entering classrooms and education the questions of when and how to use VR is still discussed.

2.1 Virtual Reality in Education

VR as an educational tool has yet to reach the same maturity and usefulness as conventional teaching methods, and research comparing VR to real classrooms experiences has produced mixed results. There are results showing beneficial effects for learning in VR (Alhalabi, 2016; Webster, 2016; Maresky et al., 2019; Yoganathan et al., 2018; de Boer et al., 2016) while others do not show such results (Makransky et al., 2019; Parong & Mayer, 2018; Mayer et al., 2022). Most studies are, however, unanimous in showing that VR can be a good way to capture students' attention and gain their engagement (Bailenson, 2018; Mikropoulos et al., 1998; Bennett & Saunders, 2019; Allcoat & von Mühlenen, 2018). One classroom study by Liu and colleagues (2020) showed how both academic achievement and engagement increased when a specifically developed immersive VR classroom was used.

The areas in which one uses VR for education include biology (Lalley et al., 2010), design (Ştefan, 2012; Sugiura et al., 2018), and engineering (Abulrub et al., 2011; Dinis et al., 2017). Here, we focus on environmental issues and sustainable behavior, a topic that always is current and that calls for massive transformation of habits and behavior. The importance of direct experience—and not just, for example, being presented with statistics—has been identified as a means of raising concern about climate change (Weber, 2006) and according to Scurati and colleagues (2021), VR is a promising tool in this respect with potential to transform accustomed patterns. Previous research has, for example, shown that screen-based VR with haptic feedback on actions promoted participants' consideration of the environment (deforestation) on both behavioral and self-reported measures (Ahn et al., 2015).

A review by Markowitz and Bailenson (2021) highlighted the capacity of VR to directly visualize things that otherwise would have to be imagined and how this can inform and affect attitudes towards environmental issues (specifically related to climate change), for example, the ability to deal with ocean acidification (Markowitz et al., 2018). Further on, Bailey et al. (2015) found that the degree of vividness and personalization of feedback messages on the amount of energy used to heat and transport virtual water for taking virtual showers affected the amount of hot water participants used when washing their hands in the real world.

Participants exposed to the more vivid messages used cooler water than the group that did not receive these messages. A similar study examined the consequences of exposing individuals to virtual experiences with vivid and personal relevance aiming to teach them about water conservation. The VR experience here had a positive influence on self-reported attitudes and behavioral intentions towards water conservation (Hsu et al., 2018).

In summary, immersive VR experiences have been demonstrated to increase engagement on a number of topics, and have beneficial effects on factual learning, as well as influencing attitudes and behavior. Several authors raise the immersive nature of VR as an explanation for these effects (e.g., Alhalabi, 2016; Maresky, 2019; Markowitz et al., 2018). Immersion, however, has a potential drawback with relevance to education, in that it separates the user from their social environment where learning and teaching normally takes place.

Introducing VR into school contexts is a complicated task (Tärning et al., 2022), by requiring available hardware setups, teacher knowledge, as well as being able to maintain a class of students attentive and on task. In combination with the fact that not many schools have class set-ups, this makes it even more complex and solutions that could engage more than one student at a time are warranted.

2.2 Collaborative Learning in VR

Traditionally, VR is used by individuals alone, but in recent years an increasing number of papers focus on collaborative use. That the person wearing the HMD is isolated from their physical surroundings and peers is a general issue with VR (Kaimara et al., 2022). The person in VR can see and interact with the virtual environment but is isolated from what is going on in their surroundings (Deakin, 2021; Jindal et al., 2024), while the surrounding peers cannot directly experience what is happening in the virtual environment. This implies that the users in the two roles (in-VR and out-VR) are not typically aware of what their peers in the other role are perceiving or attending to. This impedes their mutual communication, and in turn, collaboration (Monahan et al., 2008). Also, the person in-VR might experience a feeling of being isolated or alone in another reality (Hayes, 2022; Merkx & Nawijn, 2021) without common frames of reference or recourse to request help from peers. It is thus difficult to create a sense of being socially included. However, if we are to use VR in schools or at other learning venues social inclusion of more than one person at a time is needed. If this can be solved, more people are likely to be engaged by a VR-experience, hopefully leading to increased learning.

Previous research has shown that students are more motivated during collaborative learning than individual learning (Slavin, 1983; Johnson & Johnson, 1999). Collaborative learning is, according to Dillenbourg (1999), a learning situation in which more than two people learn or attempt to learn together. Regarding possibilities for collaborative learning in virtual reality some studies compare symmetric vs. asymmetric use of VR. In symmetric collaborations all collaborators view one and the same VR view by using individual head displays. In asymmetric collaborations one collaborator is viewing an HMD and the other some type of table-top interface, typically a screen. Drey and colleagues (2022) found that symmetric learning results in higher presence, immersion, player experience, motivation, cognitive load, and learning outcome. However, according to Zheng and colleagues (2018), it is not clear whether collaborative learning in VR can increase learning or not.

Many studies on collaboration in VR involve pair-wise learning. Research on larger groups is scarcer. Some notable exceptions include de Back and colleagues (2020) who showed how student groups of around four people had significantly higher learning gains after collaborative learning in a virtual environment compared to conventional learning. Further, Wen (2021) studied possibilities for augmented reality to enhance active vocabulary learning in classrooms. When comparing students learning Chinese characters through the use of an augmented reality game to students who learned via PowerPoint, results were that students in the experimental group (AR group) showed an increased engagement compared to students in the PowerPoint group. More recently, Queiroz and colleagues (2023) studied the effects on cognitive and affective aspects of learning when comparing different levels of interaction in VR. The results showed that higher levels of interaction in VR corresponded to higher scores on the learning tests.

Despite some evidence that symmetric collaboration (Drey et al., 2022) may have beneficial effects, we opted to focus on asymmetric collaboration for the current study. The main motivation for this was to adhere to the constraints we identified with regards to introducing VR into school contexts above. Instead of designing a collaborative VR experience from the ground up (as done in previous work, e.g., de Back et al., 2020; Queiroz et al., 2023) we opted to explore if we could adapt an existing educational VR experience, where students perform tasks individually, to promote collaboration.

2.3 Research Questions

For the study presented in this paper, we focus on user experience and to what extent a VR environment is engaging (both from an outside and inside perspective). In addition, we investigate in what ways and how different teams of students collaborate. Our reason for wanting students to collaborate is foremost a wish to increase overall engagement, in particular for out-VR users. We therefore developed the existing VR experience (*EnVRment*) by creating a collaborative and a non-collaborative version. The aim was to discover if peers surrounding the student immersed in VR could become more engaged in the experience. We asked ourselves: Can an existing VR experience be made more social and engaging without adding more headsets to the learning environment?

Our more specific research questions were the following:

RQ1: Do students in the collaborative version seem more engaged during the VR experience than students in the non-collaborative version?

RQ2: Do students in the collaborative version learn more compared to students in the non-collaborative version?

RQ3: Will the different conditions affect how students perform the tasks involved in the VR experience?

3 Method

We investigated the research questions by a between-subject designed study.

3.1 Participants

Students from one 6th grade class (10 girls, 9 boys, and one who did not choose to state gender) from a city school in the south of Sweden participated in the experiment. The experience took place in the *House of Technology and Shipping* (*Teknikens och sjöfartens hus* in Swedish), a science center open to the public. Most students were familiar with VR, as indicated by raising their hand when asked after being welcomed to the science center. None, however, indicated having used VR in a learning or collaborative situation. We did not collect further data on individual students' experience with VR.

3.2 Setup

Three different VR stations were set up at the museum. Two stations were located at opposite ends in the spacious hallway at the museum. The third station was located in a separate room connected to the same hallway. All stations were sufficiently spaced out to avoid distractions from other groups. On each station the group members stood next to each other while playing the games. Usually the in-VR-member was located in the center and the out-VR members a few steps to the side, this in order to avoid possible collisions.

The VR equipment used at each station consisted of an HTC Vive Pro headset and controllers (one controller was needed for the experience, and the other was a backup) tracked by two HTC SteamVR Base Stations 2.0 and connected to PCs running the SteamVR client, as well as the VR experience executable. The PCs were also connected to displays (one station to a wall-mounted 55-inch and two 28-inch screens standing on 1.25-m-tall tables) and pairs of external speakers mirroring the output of the headsets' integrated speakers.

At each station, an experimenter (one of the authors or a museum pedagogue) was responsible for running the VR experiment. Moreover, one observer was present at each station and took notes of what went on during the experiment and more importantly what the students discussed.

3.3 Material

For the VR-experience, we used an existing educational VR application about environmental issues (EnVRment), implemented in Unity3D (version 2018.4). EnVRment was already designed to be performed in teams of three, each taking turns to perform different subgames in VR. For the purpose of the current study, we developed a collaborative version where team members were urged to collaborate to complete the subgames, by making some crucial information visible only to the out-VR-members. In the non-collaborate version the in-VR team member could figure out what to do on their own (but was not instructed not to collaborate). Note that all three subgames were technically possible to solve by extensive trial-and-error, also in the collaborative version. The teams received instructions from their assigned experimenter. The instructions included how to use the controls, what to look for at the start of each subgame, and (broadly) what was required to successfully complete it. In both versions the surrounding team members, that is, out-VR-members, could hear the dialogue from the VR-experience, including real-time instructions and factual information given verbally via external speakers.

3.3.1 VR-Experience. The full EnVRment experience consisted of three different subgames; “waste recycling,” “energy conservation,” and “renewable energy sources.” The first in-VR member was first presented a big room with a door. There was a polar bear present who acted as instructor throughout the experience, telling the participants what to do in each subgame. The participant was instructed to open a door after hearing a constant knocking. After passing through the door, the participant arrived at a kitchen where they got acquainted with their virtual hand and practiced how to interact with objects in the environment. After opening a door under the sink, the first subgame (waste recycling) started. The first two subgames (waste recycling and energy conservation) were scripted to be very difficult at the first attempt, making it virtually impossible to succeed. This was done purposefully since we wanted everyone to see at least one negative “consequence scene,” that is, a short scene illustrating negative consequences of, for example, not recycling your waste. These scenes contained some factual information that was later covered in the test (see 3.3.4 Pre-Knowledge Test).

Waste recycling: In the first subgame, participants were faced with the task of recycling wastes in three different bins: plastic, paper, and metal. In a large box in front of them, either plastic bottles, paper boxes, or metal tins appeared at random intervals. To succeed, participants had to sort four plastic bottles, four paper boxes, and four metal tins before time runs out (there is a digital clock on the wall counting down).

In the collaborative version, the in-VR team member could not see the labels indicating what material belong to what bin (Figure 1a). They thus required help from their out-VR team members who could see this information. The out-VR members could also see how many items were left to recycle to be able to succeed with the task. In the non-collaborative version, the in-VR member could themselves see what material and how many items should go in each bin (Figure 1b).

If not succeeding with the task at the first attempt (by design), one of three consequence scenes was shown. Depending on what material the in-VR member failed to sort, a scene showing the negative consequences of not sorting enough plastic, paper, or metal was shown (see Figures 2a, b, and c).

When succeeding with the task, a positive consequence scene followed, showing a blossoming and green nature with a happy polar bear explaining what is good about recycling waste. The polar bear then proclaimed that it was time to take off the headset and for another person to get ready for the next subgame.

Energy conservation: For the second subgame, the in-VR member had to save energy by turning off a heater, an oven, a dripping water tap, a lamp, and closing a refrigerator door. On the kitchen counter, a battery was shown, depicting that it is running out of energy. If the student were not able to save enough energy and the battery runs out, the student was asked to try again after seeing a consequence scene of what could happen

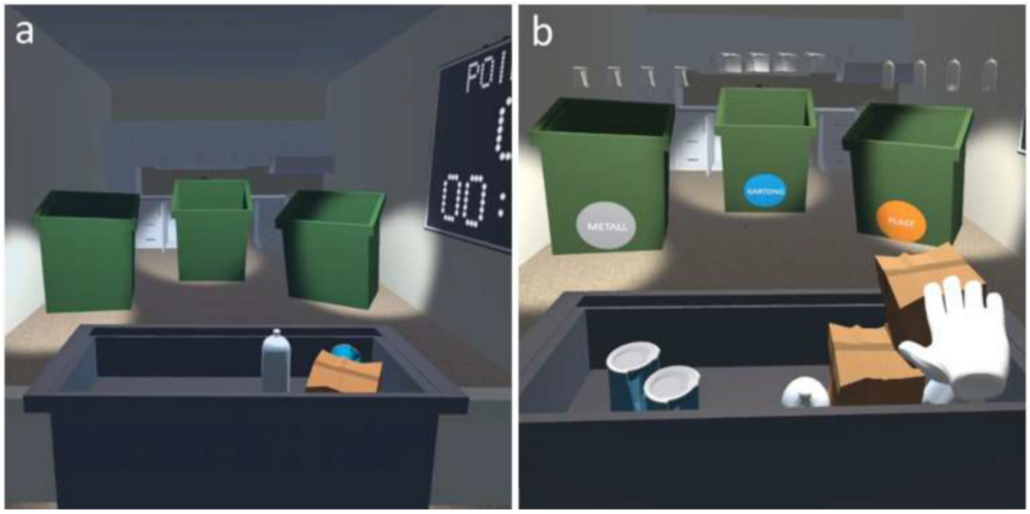

Figure 1. *Screenshots from the waste recycling game where students throw plastic, paper, and metal into one of the three bins. a: What the in-VR member sees in the collaborative version. b: What the in-VR member sees in the non-collaborative version, and what the out-VR members see in both versions.*

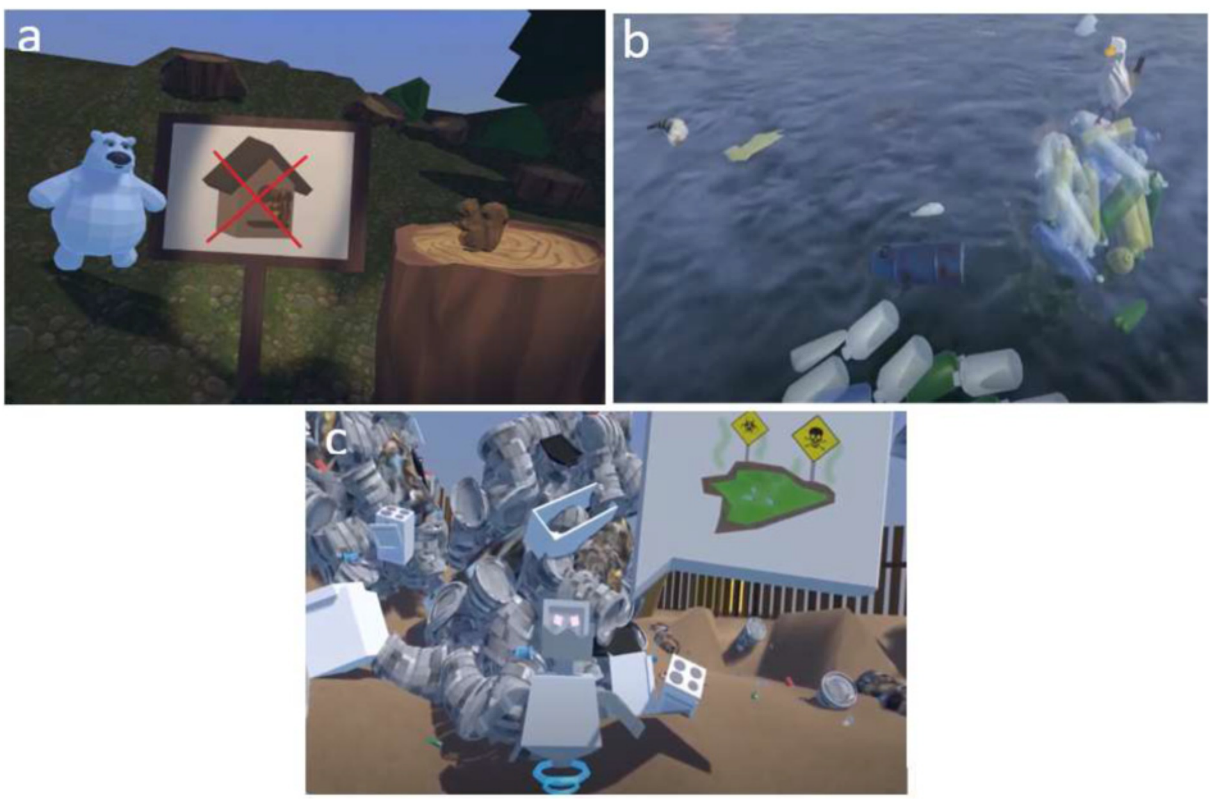

Figure 2. *Screenshots from the different sequences when failing to recycle enough a: paper, b: plastic, and c: metal.*

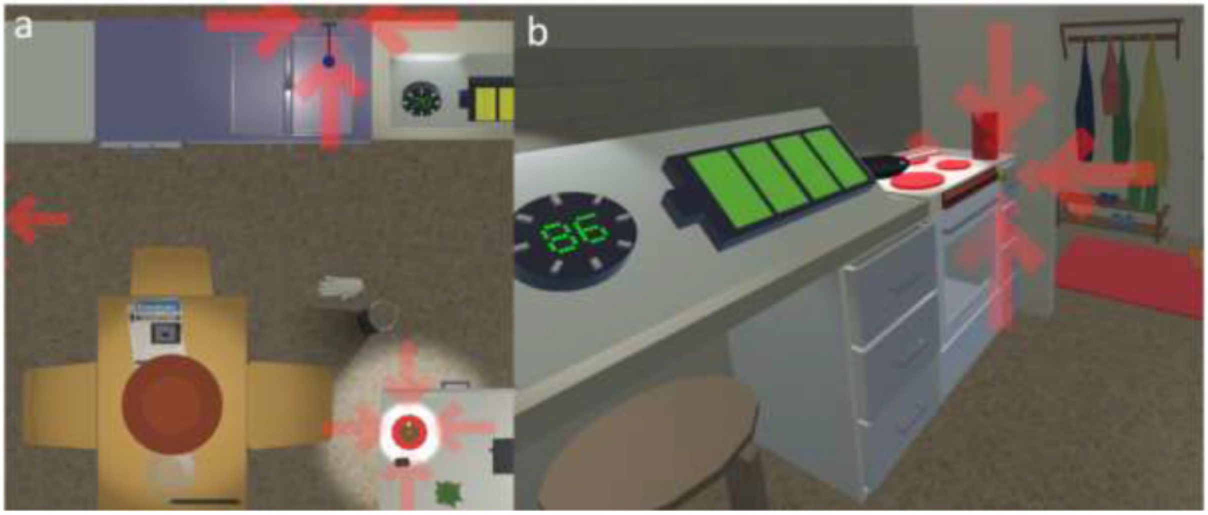

Figure 3. *Screenshots from the energy conservation subgame. a: indications shown to the out-VR team members in the collaborative version. b: indications shown to the in-VR team member in the non-collaborative version.*

when not conserving enough energy; in this scenario, the scene related to the coal industry. When succeeding with the task, the in-VR member was asked to push a button that lead to closing down a coal power plant and illustrating the positive effects of using less coal.

In the collaborative version, the in-VR member did not have any visual indications of which things in the kitchen needed to be turned off. Instead, the other out-VR team members could see animated red arrows indicating what needed to be turned off (see Figure 3a) from a birds-eye view. The in-VR member could, however, hear sound indications (e.g., water dripping, a heater turning on) and see the light from a lamp when turned on, but the sounds (in the collaborative version) did not appear to come from any specific direction (non-spatialized). In the non-collaborative version, the student wearing the headset saw arrows in a first-person view in combination with spatialized sound indications (Figure 3b).

Renewable energy sources: During the last task, the in-VR member had to build three renewable energy sources: solar panels, wind power stations, and a waterpower plant. The parts to build the different energy sources were laid out on a virtual table in a VR classroom (see Figure 4d).

There was no time limit; therefore, it was impossible to fail the subgame and hence no negative consequence scenes were shown. Instead, when all the renewable energy sources were completed, the polar bear gave some information about it, such as when a waterpower plant is most effective.

As guidance on how to build the different power stations, there were posters showing plans of how and where each part should fit together. In the collaborative version, only the out-VR members could see the poster and hence explain to the in-VR team member where each part should go. The in-VR team member could not see the poster at all. In the non-collaborative version, in-VR team members could see the poster placed behind them in the virtual environment (see Figures 4a, b, and c).

3.3.2 The Dome Tour. All students were also shown a panoramic short film about space in a dome theater adjacent to the experiment site. Some participants saw the film before, and some after, participating in the experiment. Due to a technical mishap, the voice-over was out of function, so the otherwise accompanying voice telling stories about the cosmos could not be heard. There were, however, other typical space sounds, for

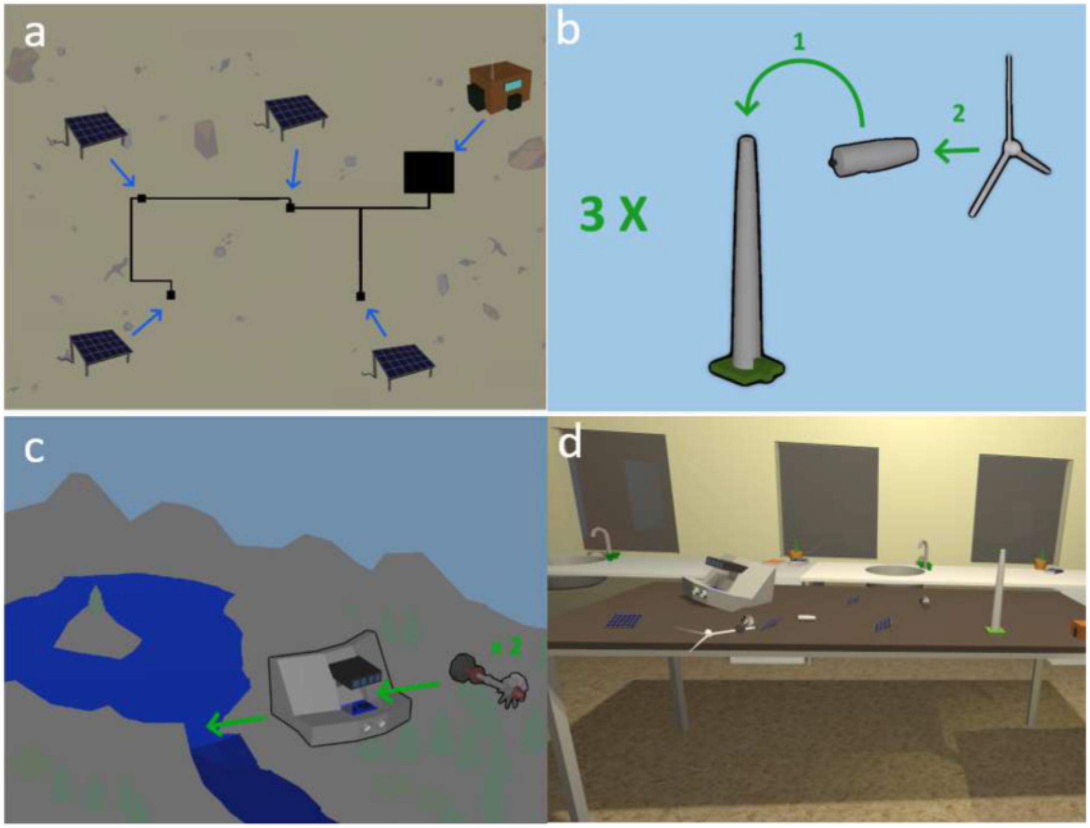

Figure 4. *The layouts of how to build the different power stations (a: solar panels, b: wind power station, and c: water power plant) in renewable energy sources subgame, displayed as a poster in the virtual environment (non-collaborative version) or on the separate screen (collaborative version). d: headset view of the table with the selectable parts of all three kinds of power stations.*

example, astronauts that could be heard at times. Despite this mishap, the museum pedagogue who guided the dome tour verified that the students found the movie interesting and exciting and that they stayed quiet and focused the whole time.

3.3.3 Performance Logs. As measures of task performance, different kinds of data were logged from each of the three subgames. During the first attempt (of two, where the difficulty level was lower for the second attempt) at subgame one (waste recycling), we logged how many items were thrown into an incorrect bin. During the first attempt at subgame two (energy conservation), it was logged how long it took between when an appliance (e.g., stove, lamp) was turned on and when the in-VR team member turned it off. For subgame three (renewable energy sources), the number of parts picked up by the in-VR team member before completing each type of power plant was logged.

3.3.4 Pre-Knowledge Test. A few days before arriving at the museum, the class was asked to complete a test. The teacher received a link to a Google-form questionnaire and the students responded individually to it via their computers. The questions concerned topics that were spoken by the polar bear during the subgames or consequence scenes, such as “why is it important to recycle plastics?” or “when can solar panels produce energy?” The questionnaire contained 21 questions; 14 of them were multiple choice questions and 6 were “fill in the blanks." These together were scored and analyzed quantitatively. Out of these 20 questions, some were related to information given specifically during one of the three subgames ("waste recycling": three questions, "energy conservation": four questions, and "renewable energy sources": nine questions).

The remaining question ("What can you do to not increase the greenhouse effect?") required a free-text answer and responses were analyzed qualitatively. For most of the multiple-choice questions, more than one of the alternatives were correct, which lowered the chance of getting a correct answer by pure guessing. Looking at the questions in more detail and analyzing them in accordance with Blooms taxonomy (Armstrong, 2010), most questions ($n = 13$) required the student to memorize facts (i.e., the "knowledge level"). These were questions such as, for example, "When can solar panels produce energy?", "You can find a rotor blade within a . . ." All fill-in-the-blanks were on the knowledge level. The other eight questions targeted the "understanding level," meaning that the student also needed to understand the memorized facts. Examples of these questions were "Why is it important to recycle plastics/paper/metal?" and "What can you do to not increase the greenhouse effect?"

3.3.5 Post-Knowledge Test. The exact same questionnaire was filled out a few days after the visit to the House of Technology and Shipping.

3.3.6 Attitude Questionnaire. On site, after the teams had completed the VR-experience, the students were also asked to complete a questionnaire regarding EnVRment and their experience. This questionnaire contained seven questions; the first two asked about the student's engagement (when someone else was wearing the headset and when wearing the headset themselves) from 1 (not at all engaged) to 5 (very engaged). The following two questions concerned how much the student listened to the polar bear (when someone else was wearing the headset and when wearing the headset themselves) from 1 (not at all) to 4 (very much). The fifth question asked whether they thought the task they had completed would be performed better in a group or individually.

3.3.7 Observer Notes. There was one researcher present at each VR-station, taking notes about what was said in the teams. Particular interest was paid to how much they talked to each other, what they were talking about, and to what extent they seemed engaged and looked at the computer screen.

3.4 Procedure

Before arriving at the museum, the class had completed the pre-knowledge test in their classroom. When on site, the class was greeted by the researchers and the pedagogues at the museum. They were briefly told what was going to happen during their visit. The class was then divided into two halves by their teacher. One half started with the VR experience and the other half went for a guided tour in the dome theater. When taking part in the VR experience, the students were further divided into teams of three; again, this was arranged by their teacher in accordance to her experience of which students were able to work together.

The first half (three teams of three students in each, and one team with two students and one teacher) used the collaborative version. The second half (three teams of three students) played the non-collaborative version. One of the students in the latter group mistakenly received the collaborative version of the game but received non-collaborative instructions.

The students were told that they could stop the VR-experience at any time if they felt motion sickness or had any other sensory issues. The VR experience took each team on average 24 minutes to complete ($M = 24$, $SD = 3$). This included the brief (approximately one minute) onboarding each student received, which consisted of orienting themselves in an intermediate space between subgames where the polar bear gave brief instructions, the experimenter entering their ID and left-/right- handedness to the application, and the student practicing a simple action with the controller (opening a door or pulling a lever).

Table 1. *Responses to Questions about Participants' Engagement in-VR as Well as Out-VR*

	How engaged were you when someone else was wearing the headset?				*How engaged were you when wearing the headset yourself?*			
	Collaborative version		Non-collaborative version		Collaborative version		Non-collaborative version	
	Nr of students	%	Nr of students	%	Nr of students	%	Nr of students	%
Very engaged	2	18%	0	0	9	82%	4	50%
Quite engaged	8	73%	5	63%	2	18%	2	25%
Neither nor	1	9%	1	13%	0	0	2	25%
Quite ungengaged	0	0	2	25%	0	0	0	0
Very Unengaged	0	0	0	0	0	0	0	0

4 Results

To explore our three research questions, we examined the attitude questionnaire, the pre- and post-knowledge tests, performance logs, and observer notes.

Data analysis was performed using *R Studio* (version 2022.07.1) with the *ordinal* (version 2022.11-16; Christensen, 2022), *lmerTest* (version 3.1-3, Kuznetsova, Brockhoff, & Christensen, 2017), and *MuMIn* (version 1.47.5; Barton, 2009) packages.

4.1 Team Engagement: RQ1

For RQ1 (whether students in the collaborative teams seemed more engaged during the VR experience), we analyzed the attitude questionnaire and the observer notes.

4.1.1 Attitude Questionnaire. Since one of the students received non-collaborative instructions but played the collaborative version of subgame one (the waste recycling game), we excluded this data from the results. This left us with 19 students taking part in the experience and answering the questionnaire.

Of particular interest was whether students in the collaborative teams did feel more engaged compared to students in non-collaborative teams. Regarding the first two questions in the attitude questionnaire, we could see that students in the collaborative condition reported higher engagement both when wearing the headset themselves but also when someone else was wearing it (see Table 1). Moreover, all students using the collaborative version reported that they were very, or quite engaged, and none of those students reported that they felt unengaged whereas two students using the non-collaborative version reported that they did.

To check for statistical differences, we analyzed the responses as a cumulative link mixed model with Laplace approximation from the *ordinal* R package. We modelled engagement as an interaction between the two fixed factors Condition (Collaborative/Non-collaborative) and Role (response to which of the two questions, in-VR / out-VR) with random factors participant ID and subgame that the participant performed (recycle, energy conservation, renewables). Supplementary Table S1 summarizes the results. Note that the model estimated that the collaborative teams rated their engagement higher ($\beta = 1.993$, std. err. = .927, $z = 2.149$, $p = .032$) and that engagement was lower when out-VR ($\beta = -2.490$, std. err. = 1.177, $z = 2.116$, $p = .034$). There was no significant interaction between the factors; that is, the collaborative teams rated higher engagement irrespective of in-VR or out-VR.

A related measure of engagement is how much attention the students paid to the polar bear while it was giving instruction and information. Listening attentively

Table 2. *Responses to Questions about How Much Participants had Listened to the Polar Bear*

	Did you listen to what the polar bear said when someone else was wearing the headset?				*Did you listen to what the polar bear said when wearing the headset yourself?*			
	Collaborative version		Non-collaborative version		Collaborative version		Non-collaborative version	
	Nr of students	%	Nr of students	%	Nr of students	%	Nr of students	%
Very much	3	27%	2	25%	7	64%	4	50%
Quite a lot	4	36%	4	50%	3	27%	4	50%
Not so much	4	36%	2	25%	0	0	0	0
Not at all	0	0	0	0	1	9%	0	0

to what the polar bear said could indicate higher engagement, compared to not listening. It appears from these answers that the inclination to listen to the polar bear both when wearing the headset or when someone else was wearing it was similar. Most students responded that they listened quite a lot or very much (see Table 2).

To check for statistical differences, we again analyzed the responses as a cumulative link mixed model with Laplace approximation from the *ordinal* package. We modelled self-reported listening as an interaction between the two fixed factors Condition (Collaborative/Non-collaborative) and Role (response to which of the two questions, in-VR/out-VR) with random factors participant ID and subgame that the participant performed (recycle, energy conservation, renewables). The regression revealed no significant effects. Supplementary Table S2 summarizes the results.

In the attitude questionnaire, students were also asked whether they would have preferred to do the task by themselves or as a team. The results indicate that students in the collaborative teams preferred to do it as a team: 82% responded "as a team" ($n = 9$) and 18% responded "alone" ($n = 2$) whereas no strong preference was observed for the non-collaborative teams: 37.5% ($n = 3$) responded that they would prefer to do it in a team and 62.5% ($n = 5$) that they would prefer to do it alone. However, a Fisher Exact Test showed that the difference in ratios between conditions was non-significant ($odds\ ratio = 6.611$, $p = .073$).

4.1.2 Observation Notes. Overall, all three notetakers who observed the students found the teams using the collaborative version to interact more among themselves. One of the game prerequisites was that out-VR members had to interact in order to help the person wearing the HMD but the interaction extended beyond this. Things like encouragement and cheering for one another was noticed much more often in the collaborative teams. In general, when comparing the notes taken, they were substantially longer for the collaborative teams. The number of utterances the notetakers had taken for the collaborative version were 158 compared to only 48 for the non-collaborative version. Of course, some utterances may have been emitted but this would more likely have a negative outcome for the collaborative groups since these groups talked more, which likely made it more challenging to find the time for notetaking. Even though there were three different notetakers, with potentially different notetaking habits, we did not see any particular differences between the amount of notes the three notetakers made. They all made approximately the same number of notes. There were, however, clear differences between the collaborative and the non-collaborative groups, apparent for all notetakers (with all of them taking notes for both types of groups). One of the three non-collaborative teams distinguished themselves from the others somewhat in that they were more talkative and to some extent interacted in more similar ways as the collaborative teams did.

Table 3. *Table Showing NR Captured Statements in the Non-Collaborative and Collaborative Groups; the +7 INDICATES 7 Comments Made by an Experiment Leader*

		Non-collab nr of utterances	Collab nr of utterances
In-VR asks for help	Waste recycling	5	10
	Energy conservation	3	7
	Renewable energy sources	12	13
	Total	20	30
Out-VR helps/instructs	Waste recycling	2	22
	Energy conservation	11 + 7	43
	Renewable energy sources	0	22
	Total	13 + 7	87
Out-VR asks in-VR member	Waste recycling	0	2
	Energy conservation	1	4
	Renewable energy sources	1	9
	Total	2	15
	Total all subgames	35 (21%)	132 (79%)

On the basis of the observational notes, the out-VR members were, in general, actively trying to help the in-VR members by instructing them on how the task was supposed to be solved, what should be done in order to succeed, or by encouraging the in-VR member to proceed and to do their best.

To study student engagement, we analyzed the observational notes in more detail by performing a thematic analysis (Braun & Clarke, 2006, 2021). Four themes reflecting level of engagement were identified. The themes were as follows: (i) *In-VR asks for help*, (ii) *Out-VR helps/instructs* (the in-VR member), (iii) *Out-VR asks in-VR member* (the out-VR member wants information regarding the experience from the in-VR member), and (iv) *Encouragement*. We suggest that one way of measuring engagement is through the amount of interaction within the groups. A higher level of interaction usually requires a higher level of engagement. During the thematic analysis, the notes from each group were categorized into these four themes. There was also a fifth theme, "other utterances," where utterances not relating to things in the game were categorized. But this category turned out to be small with only 10 utterances for the collaborative groups and 10 for the non-collaborative groups. This category is not included in the presented results. For the included categories the number of utterances is summarized in Table 3 (the category "engagement" is shown in Table 4). The number of utterances were calculated per subgame.

When it comes to encouraging the in-VR member, the directed utterances were relatively few overall (see Table 4) but there were notably more of these in the collaborative groups with 17 times of encouraging the in-VR member by utterances such as "that's good, that's good," "one more, one more," and "that's nice." Only six similar utterances occurred in the non-collaborative groups (and to be noted, all utterances in the non-collaborative group categorized as encouragement were made by one and the same group).

Looking at the different subgames in more detail, starting with the waste recycling game, students in the collaborative teams explained to the in-VR member where they should recycle what; *"What, paper boxes, where should this go?"* (in-VR) *"In the middle"* (out-VR) and also reminded them while playing *"Yes, to the left, that is nice, good"* (out-VR). In the non-collaborative

Table 4. *Table Showing Number of Captured Statements Indicating That Out-VR Members Encouraged the In-VR Member*

		Non-collab nr of utterances	Collab nr of utterances
Encouragement	Waste recycling	1	8
	Energy conservation	3	7
	Renewable energy sources	2	2
	Total	6	17

teams, the out-VR members did not interact with or help the in-VR member nearly to the same extent. For two of the non-collaborative teams, when the in-VR members asked what they were supposed to do, the researcher had to step in and help, since the out-VR members did not respond. The third non-collaborative team did, however, interact more and said things like "*Think basketball but still not basketball*" (out-VR) while observing the in-VR member trying to sort different waste into the correct bin. The collaborative teams were more instructive, interacting, encouraging, and cheering: *"You don't need any more card boxes"* (out-VR) "*Only plastic?" (in-VR)* and *"You have made it!"* (out-VR) *"Did I make it?" (in-VR), "Yes" (out-VR).*

For the energy conservation game, the collaborative teams instructed the in-VR member to a great extent, telling them where to go next to turn something off. For collaborative teams, this subgame was the one with most interaction, with out-VR members instructing the in-VR member where to go and to keep track of the time (43 utterances made in this category). For the non-collaborative teams, the interaction was not as intense; only one of these teams actually interacted at all but gave fewer instructions to where the in-VR member should go to turn something off (there were 11 utterances made in the non-collaborative group plus 7 made by the experiment leader due to lack of other interaction between group members). However, it was also apparent that all items that could be interacted with were not easy to categorize and the out-VR members said things like *"What is that, a dishwasher perhaps?" "Yes, the stove, behind you" or "It is a lamp"* (in-VR) *"Ahha* [laughter] *we could not see that"* (out-VR). In subgame two (saving energy), it was possible to talk about objects present in the virtual space, such as a dishwasher and a lamp, for non-collaborative as well as collaborative teams. However, this was not done for any of the non-collaborative teams.

During the last subgame (renewable energy sources) it was not clear for the in-VR members what they were supposed to do and some of them asked the out-VR members what the objective of the game was. Looking at the thematic analysis, 12 utterances where the in-VR member asked the out-VR members for help were made in the non-collaborative group and 13 utterances were made in the collaborative group. It seemed to be quite difficult to give instructions during this game and it was not obvious to the out-VR members what the in-VR member could see. For example, the in-VR member could ask what they were supposed to place at the squares that they saw (it should be solar panels) but after placing one or more of these panels, the out-VR members could not see how many were in the correct place, something that led to some communicative mishaps *"Should I place solar panels at all smaller squares?"* (in-VR) *"The large dot, have you placed anything there?"* (out-VR) *"What should I place there?"* (in-VR) *"A house"* (out-VR). Continuing, out-VR members guided the in-VR member through the different steps so that they could build a wind powerplant, asking for example; *"Have you placed the spinning top on the short stick?"* (out-VR) *"No, not yet"* (in-VR) *"On the SHORT stick, not the long one"* (out-VR). Even though it was somewhat difficult to get an overview over this subgame, all collaborative teams communicated well and frequently, and tried to guide the member in VR (22 utterances regarding "out-VR helps/instruct" were made in the collaborative teams compared to 0 in the non-collaborative teams). In the non-collaborative teams, the interaction was not as intense and in one team the in-VR member simply talked himself through the game (since he did not get any replies from the other team members); *"Well, this big thing maybe, or? . . . Maybe one of these?"*

Overall, most students, both in the collaborative and the non-collaborative teams seemed to pay attention to the consequence scenes and what the polar bear was saying (see 4.1.1 Attitude Questionnaire). In one non-collaborative team, it was more apparent that one person not wearing the headset did not listen; they went to the bathroom and when present, they sat quite far away from their other team members, not looking at the screen to see what was going on.

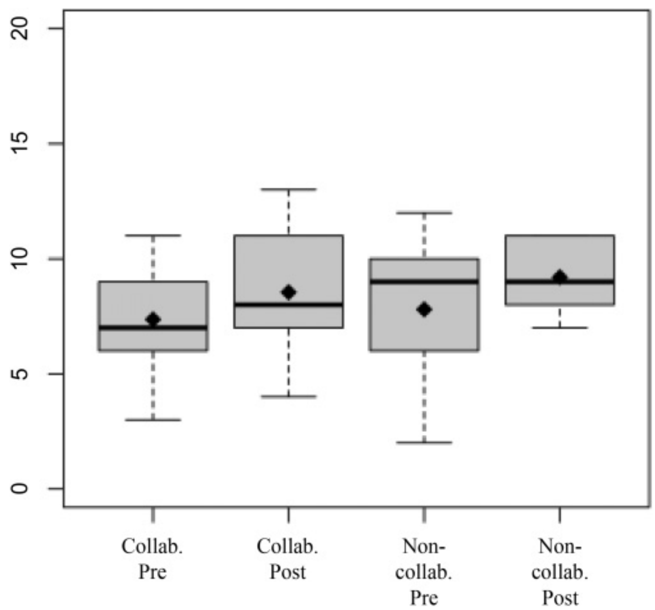

Figure 5. *Boxplot showing the distribution of total scores (out of max 20) per participant. Rhombi represent means.*

4.2 Knowledge Improvement: RQ2

Research question two (*Do students in the collaborative version learn more compared to students in the non-collaborative version?*) was explored by looking at the pre- and post-test knowledge scores. After having excluded students who did not reply to both pre- and post-tests and participated in the VR experience, we were left with 16 participants.

In total, there were 21 questions: 14 multiple choice questions, 6 fill-in-the-blank questions, and one free-text question. The free-text question was treated separately (see 4.2.1 Free-Text Question). All other questions were scored either as zero or one point for a completely correct response.

Most of the multiple-choice questions (11 out of 14) contained more than one correct answer; for example, a question could have four different alternatives with two (or three, or four) correct alternatives. Each correct alternative and none of the incorrect alternatives needed to be chosen in order to score one point on a question. For a few questions, only one answer was correct and hence only the correct answer needed to be chosen. The fill-in-the-blanks question contained 6 blanks and were scored individually. The information provided in the free-text question was analyzed, to see if they changed or added something to their answers from pre- to post-test.

Most participants showed some improvement on their scores from the pre- to post-test (7 out of 11 in the collaborative teams and 3 out of 5 in the non-collaborative teams). Figure 5 gives an overview of the scores of the pre- and post-tests per condition.

To explore the participants' information uptake from playing the game, either in-VR or out-VR, we performed an item level analysis. Apart from correct answers, we coded the subgame during which the relevant facts were given for each question. We analyzed the scores as a logistic binomial mixed regression model as a sum of the fixed factor Time (Pre/Post), the interaction term between Time and Condition (Collaborative/Non-collaborative), and interaction term between Time and Subgame-match (a factor representing if the related facts were given during the subgame the participant performed in-VR). The model also included participant ID, the team of three the participant was part of, and the specific question as random factors. The model predicts significantly higher log-odds for correct answers on the post-test (*odds-ratio* = .721, *std. err.* = .410,

Table 5. *Table Showing the Number of Students and Percent Who Mentioned a Category in Their Free-Text Responses to Question Five*

	What can you do to not increase the greenhouse effect?							
	Pre questionnarie				Post questionnaire			
	Non-collab		Collab		Non-collab		Collab	
	%	nr of stud.	%	nr of stud.	%	nr of stud.	%	nr of stud.
Fly less	3%	1	6%	2	5%	2	5%	2
Go by car less	6%	2	23%	7	5%	2	8%	3
Go by bicycle more	0%	0	10%	3	0%	0	11%	4
Less meat/vegetarian food	3%	1	6%	2	0%	0	3%	1
Reuse	3%	1	10%	3	5%	2	5%	2
Recycle	0%	0	10%	3	3%	1	16%	6
Save energy	0%	0	0	0	3%	1	16%	6
Other	10%	3	6%	2	8%	3	5%	2
No answer	3%	1	0	0	0%	0	0%	0

$z = 2.322$, $p = .020$) compared to the pre-test. None of the interaction terms were significant; however, the model estimated a non-significant positive effect of Time and Subgame-match test (odds-ratio = .351, std. err. = .326, $z = -1.884$, $p = .060$). Supplementary Table S3 summarizes the regression output. Conditional R^2 (pseudo coefficient of determination for mixed models) for the model was .450.

4.2.1 Free-Text Question. The free-text question was analyzed separately. Students' answers were categorized into similar replies by one of the authors (see Table 5). The question was "*What can you do to not increase the greenhouse effect?*" Overall, many of the replies in the pre-test suggested things like travelling less by plane and car and taking a bicycle more often. The same things were also mentioned in the post-test but, interestingly, here many replies also suggested that a good way to not increase the greenhouse effect is to recycle and to save energy. Recycling was mentioned by three students in the pre-test but mentioned seven times in the post-test. Saving energy was not at all mentioned in the pre-test but was brought up by seven students in the post-test. On both occasions, six out of the seven answers were made in the collaborative teams. The answers in the "other" category were mentioned only once, such as "*protect our oceans and forests*" and "*using an electrical car.*"

4.3 Task Performance: RQ3

To explore research question RQ3 (*Will the different conditions affect how students perform at the tasks involved in the VR experience?*), we analyzed performance data from the three subgames. The median number of throws to an incorrect bin (e.g., plastic in the cardboard bin) was very low; 1 for both participants performing subgame one (waste recycling) the collaborative version and the non-collaborative, (c.f. the 15 throws to the correct bin required to complete the task).

It took participants performing subgame two (energy conservation) in the collaborative version M = 23.1 s, (SD = 15.5 s, $n = 4$) and the non-collaborative M = 42.4 s, (SD = 22.7 s, $n = 3$) to turn off an appliance after it had been turned on, at their first attempt.

Participants performing subgame three (renewable energy sources) picked up M = 34.3 (SD = 15.3 s, $n = 4$) parts before assembling all power stations, while non-collaborative participants picked up M = 49.3 (SD = 17.2 s, $n = 2$). The low number of observations precludes us from making any meaningful statistical analysis on the performance data.

5 Discussion

The main purpose of this study was to explore the possibility of engaging students not wearing an HMD in different group tasks. We did this with simple modifications to an already existing VR-experience, making sure that the in-VR members would need varied amounts of help from the out-VR members to perform tasks in VR. This was done by giving the out-VR members access to specific information and engaging them more in the experience (collaborative version).

By comparing students who, in teams, played a collaborative version of our VR-experience (*EnVRment*) to students who, in teams, played a non-collaborative version we could explore whether engagement differed between students in the two conditions.

From our observations, we noticed that students in the collaborative teams interacted considerably more than students in the non-collaborative teams. Students in the collaborative teams also appeared to have more fun as they were laughing more and being more attentive to what the in-VR person was doing. In the collaborative teams, the out-VR members talked actively with the in-VR member, telling them what to do or, for example, where to throw things. The in-VR member did not always have to ask questions but rather the other members helped before getting stuck. Out-VR members in the collaborative teams seemed more engaged in that they were discussing more things within the game, such as what different things depicted, how they were doing in terms of time, and also just cheering the in-VR members on. It seems that giving instructions to peers led to further and more vivid interactions. One can of course discuss the fact that one of the prerequisites for the collaborative teams was that they were supposed to give instructions, but from the observations it was clear that the interaction went far beyond only necessary instructions in the collaborative teams. The latter seemed clearly more engaged than the non-collaborative teams, a result in line with Slavin (1983) and Johnson & Johnson (1999).

We could also see that the different subgames all worked in the collaborative game version. However, some subgames worked better than others. The first subgame "waste recycling" worked very well in that it forced the out-VR members to help and instruct the in-VR member. The person wearing the HMD had no way to tell where each waste should go and was rather dependent on the other team members. The in-VR member could have tried to just throw a plastic bottle in a random bin but none of them did; they preferred to receive instructions from the out-VR members. We could see that the out-VR members were more active in the beginning of this subgame, in that they were telling the in-VR members where to throw a certain waste. As the in-VR member could remember on their own where a certain waste should go, the interaction went from giving instructions to being more informative in, for example, telling the player what waste needed to be recycled and plain cheering on saying things like "*that's good, that's good*" and "*you have made it.*"

The second subgame, "energy conservation," worked well in that the out-VR members guided the person in VR to where to look in order to turn something off. The in-VR members could hear and see what could be turned off, but this was rather subtle and the sounds were not pointing in any particular direction. The out-VR members got a much more vivid view of the scene and could see the red arrows faster, indicating that something needed to be turned off. Since what is to be turned off is randomized, there were no possibilities for the in-VR members to learn where to go next and hence the support from out-VR members made the task more efficient. To be noted is that this subgame could be finished without help from out-VR members but it would take more time. This was also shown in the performance data from subgame two where collaborative teams took almost half the time to turn off an appliance compared to non-collaborative teams.

Addressing the last subgame, "renewable energy sources," it was less clear how to collaborate compared to the other subgames. The out-VR members could see a poster depicting pieces and instructions for how to, for example, build a wind power plant. The in-VR members could see the same poster, but it was placed at the back, meaning that they would have to turn around to make use of it.

What made this subgame troublesome was that the out-VR members could not see what the in-VR member was currently doing, what piece they were holding, or how many pieces and parts they had managed to put together. We could tell that this made the out-VR members uncertain; they didn't know how to instruct since they didn't know when something was in place or what piece the player was already holding. In two collaborative teams, the in-VR members asked what they were supposed to do since they didn't get spontaneous instructions and from the observation notes we could see less engagement in the form of interaction on this task.

From the attitude questionnaire, we could also see that students in the collaborative teams reported a higher engagement compared to students in the non-collaborative teams. To see whether students also learned domain-related facts from playing the VR-experience we compared pre- and post-test knowledge. The overall results show that participants in both conditions performed better on the post-test. However, we did not see that the difference in engagement between the conditions translated into better performance on the knowledge test. This contrasts with the findings of Queiroz et al., 2023, where their (symmetric) collaborative condition did yield learning gains.

A limitation here is of course that all students did not get to see all consequence scenes where useful information could be captured. In one of the collaborative teams, one member (the teacher, not included in the analysis) was successful at their first try on subgame one and hence the team did not get to see any consequence scenes at all. Furthermore, the tasks in which students collaborated did not in itself necessarily require uptake of domain facts that were part of the test. For example, it was possible to explain in what order to assemble the components of the power plants by describing them by appearance, without learning their proper names.

Returning to engagement, one could have expected that students in the collaborative teams would perform better on the post-test (compared to the non-collaborative teams) since engagement and commitment may influence learning in a positive way (Lepper, 1988). This prediction was not supported by our results. The tasks in which students collaborated did not directly involve domain learning by solving tasks together. Instead, they collaborated by instructing each other in how to manage a task (in a more game-like manner). Rather, the polar bear gave instructions that the students could use in the knowledge-tests. From the attitude questionnaire, it appears that both the collaborative and the non-collaborative teams paid much attention to what the polar bear said, and it is therefore perhaps not surprising that no difference in learning between conditions was found.

Another explanation might be that specifically the surrounding students in the collaborative teams were too engaged in managing the task and hence did not pay as much attention to instructions from the polar bear. Queiroz and colleagues (2023) actually found a negative correlation between self-reported feeling of being "part of the group" and learning, and suggest increased cognitive load in the social context as a possible explanation.

However, looking at our participants' own responses to how much they listened to the polar bear, the results are quite similar comparing the collaborative and non-collaborative teams. Overall, it should also be noted that only 5 participants from the non-collaborative teams completed both the pre- and post-questionnaires, thus severely limiting statistical power with respect to finding differences between conditions. This also precluded us from analyzing questions classified as different levels using Bloom's taxonomy separately.

The item-level analysis (logistic regression model) indicated that participants gained some factual knowledge about sustainability. While the effect of facts not being specifically covered in team members' respective in-VR subgames was non-significant, the observed trend suggests that there might be a tendency for increased attention to facts given while in-VR. To properly test this, and how it may be moderated by promoting collaboration, a larger and better-balanced sample would be necessary. A plurality of the 20 multiple-choice or fill-in-the-blank questions were related to information given during the renewable energy sources subgame, for which the collaborative version worked the least well. This may have further obscured any differences in learning between conditions.

Moreover, one might consider the potential problems with using the same pre- and post-test and how the students might be more attentive to information regarding what was mentioned in the pre-test when later using VR. However, one intention was to explore possible learning effects and therefore we wanted to make sure we measured the same thing both pre- and post-VR usage. The setup was exactly the same for all students so this would not make any difference comparing teams of students. Also, what was measured in the test was nothing the students worked with in their curricula during that time and hence the probability of them attending to or seeking out information regarding what was measured in the test is slim.

From the answers to the free-text question, we can infer that the VR-experience had some effect. Here the students were asked to suggest something they could do to not increase the greenhouse effect. In the pre-test, three students suggested recycling and in the post-test this number increased to seven students. Just as many ($n = 7$) also suggested "*saving energy*" in the post-test (something not at all mentioned in the pre-test). Thus, it appears that at least subgame one (waste recycling) and two (energy conservation) did have some effects on the students. Interestingly, these ideas were mostly mentioned in the collaborative teams. Speculating about the cause of this observation, perhaps the engagement led to better recollection for this particular question but why this would not be the case in the overall knowledge test is then left to be explained. No student suggested anything on the topic of building renewable energy sources (subgame three).

The logged data related to performance suggests that the subgames of the VR experience can be performed as well, or sometimes even more efficiently, by collaboration. For instance, the lower number of picked-up parts during the assembly of the power stations in the collaborative teams could indicate a smaller degree of trial-and-error. Again, it is important to note the small sample size and interpret the result in this light. Variation in measurement could easily be skewed by confounding factors. For example, the number of pieces picked up could also depend on the level of curiosity and acquaintance with VR. As mentioned, the low number of observations precluded us from performing any hypothesis testing statistics on the performance data.

As mentioned, the subgames themselves did not always require or involve learning of domain-related facts. It is therefore questionable how performance (collaborative or not) of the tasks translate to learning (collaborative or not) in the application under study. This could preferably be further studied by using VR applications where task performance, collaboration, and discussion are more directly integrated with domain knowledge transfer.

6 Conclusions

Most previous studies have studied collaboration in VR in other forms than this study. In this study, only one person was wearing a headset and others collaborated by added tasks for which one does not need a headset or a tablet. Instead, we started from an existing educational VR application, including tasks designed to be performed in VR without collaboration with surrounding peers. We then redesigned three subtasks by displaying different task-relevant information outside the view of the wearer of the headset.

Overall, our results indicate that raising students' engagement even when not wearing a headset themselves is possible, in the form of enriched interaction within teams, engagement with the experience itself, and enjoyment. The increased engagement did not lead to any positive learning effects looking at the scores on the pre- and post-knowledge tests. However, looking at one free-text question we could see that the responses of students—particularly having used the collaborative version—were informed by the two subgames (recycling and energy conservation) where collaboration has worked best.

The results also show that it is possible to promote engagement by enabling collaboration by quite simple measures. The cumulative development of *EnVRment* provides a blueprint for developing and evaluating educational applications first as individual experiences, and in the next step as collaborative experiences that are applicable for groups of schoolchildren. By targeting educational goals beyond the learning of simple facts, one could perhaps more readily translate increased collaboration and engagement into improved learning outcomes. However, one needs to analyze interaction and feedback loops thoroughly, making sure it is obvious to both persons in-VR and out-VR what is happening on both ends.

REFERENCES

Abulrub, A. H. G., Attridge, A. N., & Williams, M. A. (2011, April). Virtual reality in engineering education: The future of creative learning. In *2011 IEEE Global Engineering Education Conference*, pp. 751–757.

Ahn, S. J., Fox, J., Dale, K. R., & Avant, J. A. (2015). Framing virtual experiences: Effects on environmental efficacy and behavior over time. *Communication Research*, *42*(6), 839–863. 10.1177/0093650214534973

Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology*, *35*(11), 919–925.

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, 26.

Armstrong, P. (2010). Bloom's Taxonomy. Vanderbilt University Center for Teaching. https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works and what it can do*. WW Norton & Company.

Bailey, J. O., Bailenson, J. N., Flora, J., Armel, K. C., Voelker, D., & Reeves, B. (2015). The impact of vivid messages on reducing energy consumption related to hot water use. *Environment and Behavior*, *47*(5), 570–592. 10.1177/0013916514551604

Barton, K. (2009). *Mu-MIn: Multi-model inference*. R Package Version 1.47.5. http://R-Forge.R-project.org/projects/mumin/

Bennett, J. A., & Saunders, C. P. (2019). A virtual tour of the cell: Impact of virtual reality on student learning and engagement in the STEM classroom. *Journal of Microbiology & Biology Education*, *20*(2), 10–1128.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. 10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, *18*(3), 328–352. 10.1080/14780887.2020.1769238

Christensen, R.H.B. (2022). *Ordinal—Regression Models for Ordinal Data*. R package version 2022.11-16. https://CRAN.R-project.org/package=ordinal

Deakin, T. (2021, May 24). Is VR the solution to isolation or part of the problem? *MuseumNext*. https://www.museumnext.com/article/is-vr-the-solution-to-isolation-or-part-of-the-problem/

de Back, T. T., Tinga, A. M., Nguyen, P., & Louwerse, M. M. (2020). Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education*, *17*, 1–18. 10.1186/s41239-020-00228-9

De Boer, I. R., Wesselink, P. R., & Vervoorn, J. M. (2016). Student performance and appreciation using 3D vs. 2D vision in a virtual learning environment. *European Journal of Dental Education*, *20*(3), 142–147. 10.1111/eje.12152

Dillenbourg, P. (1999). *Collaborative learning: Cognitive and computational approaches. advances in learning and instruction series*. Elsevier Science.

Dinis, F. M., Guimaraes, A. S., Carvalho, B. R., & Martins, J. P. P. (2017, April). Development of virtual reality game-based interfaces for civil engineering education. In *2017 IEEE Global Engineering Education Conference*, pp. 1195–1202.

Drey, T., Albus, P., der Kinderen, S., Milo, M., Segschneider, T., Chanzab, L., . . . & Rukzio, E. (2022, April). Towards collaborative learning in virtual reality: A comparison of co-located symmetric and asymmetric pair-learning. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–19.

Hayes, A. (2022). Will virtual reality connect or isolate students? *In bridging human intelligence and artificial intelligence* (pp. 139–150). Springer International Publishing.

Hsu, W. C., Tseng, C. M., & Kang, S. C. (2018). Using exaggerated feedback in a virtual reality environment to enhance behavior intention of water-conservation. *Journal of Educational Technology & Society*, *21*(4), 187–203.

Jindal, P., Jindal, A., & Gambhir, R. (2024). Virtual reality in education: A blessing or curse? In Malik, R., Sharma, A., &

Chaudhary, P. (Eds.) *Transforming education with virtual reality*, pp. 211–228. Wiley.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory into Practice*, *38*(2), 67–73. 10.1080/00405849909543834

Kaimara, P., Oikonomou, A., & Deliyannis, I. (2022). Could virtual reality applications pose real risks to children and adolescents? A systematic review of ethical issues and concerns. *Virtual Reality*, *26*(2), 697–735. 10.1007/s10055-021-00563-w

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest Package: Tests in Linear Mixed Effects Models.* R package version 2022.11-16. https://github.com/runehaubo/lmerTestR

Lalley, J. P., Piotrowski, P. S., Battaglia, B., Brophy, K., & Chugh, K. (2010). A comparison of V-Frog [C] to physical frog dissection. *International Journal of Environmental and Science Education*, *5*(2), 189–200.

Lepper, M. R. (1988). Motivational considerations in the study of instruction. *Cognition and Instruction*, *5*(4), 289–309. 10.1207/s1532690xci0504_3

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. 10.1111/bjet.13028

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. 10.1016/j.learninstruc.2017.12.007

Maresky, H. S., Oikonomou, A., Ali, I., Ditkofsky, N., Pakkal, M., & Ballyk, B. (2019). Virtual reality and cardiac anatomy: Exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education. *Clinical Anatomy*, *32*(2), 238–243. 10.1002/ca.23292

Markowitz, D. M., & Bailenson, J. N. (2023). Virtual reality and emotion: A 5-year systematic review of empirical research (2015–2019). In R. L. Nabi & J. Gall Myrick (Eds.), *Emotions in the digital world: Exploring affective experience and expression in online interactions* (pp. 134–152). Oxford University Press. 10.1093/oso/9780197520536.003.0008

Markowitz, D. M., & Bailenson, J. N. (2021). Virtual reality and the psychology of climate change. *Current Opinion in Psychology*, *42*, 60–65. 10.1016/j.copsyc.2021.03.009

Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., & Bailenson, J. N. (2018). Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in Psychology*, *9*, 2364. 10.3389/fpsyg.2018.02364

Mayer, R. E., Makransky, G., & Parong, J. (2022). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human–Computer Interaction*, 1–10.

Merkx, C., & Nawijn, J. (2021). Virtual reality tourism experiences: Addiction and isolation. *Tourism Management*, *87*, 104394. 10.1016/j.tourman.2021.104394

Mikropoulos, T. A., Chalkidis, A., Katsikis, A., & Emvalotis, A. (1998). Students' attitudes towards educational virtual environments. *Education and Information Technologies*, *3*(2), 137–148. 10.1023/A:1009687025419

Monahan, T., McArdle, G., & Bertolotto, M. (2008). Virtual reality for collaborative e-learning. *Computers & Education*, *50*(4), 1339–1353.

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

Queiroz, A., McGivney, E., Liu, S. X., Anderson, C., Beams, B., DeVeaux, C., . . . & Bailenson, J. N. (2023, April). Collaborative tasks in immersive virtual reality increase learning. In *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning.*

Scurati, G. W., Bertoni, M., Graziosi, S., & Ferrise, F. (2021). Exploring the use of virtual reality to support environmentally sustainable behavior: A framework to design experiences. *Sustainability*, *13*(2), 943–963. 10.3390/su13020943

Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin*, *94*(3), 429–445. 10.1037/0033-2909.94.3.429

Ştefan, L. (2012). Immersive collaborative environments for teaching and learning traditional design. *Procedia-Social and Behavioral Sciences*, *51*, 1056–1060. 10.1016/j.sbspro.2012.08.287

Sugiura, Y., Ibayashi, H., Chong, T., Sakamoto, D., Miyata, N., Tada, M., . . . & Igarashi, T. (2018, December). An asymmetric collaborative system for architectural-scale space design. In *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pp. 1–6.

Tärning, B., Ternblad, E.-M., Haake, M., Gulz, A., & Nirme, J. (2022). Lessons learned from a study on distractions in virtual learning environments: Reliability, ecological validity and an elusive social component. *PRESENCE: Virtual and Augmented Reality*, *28*, 65–85. 10.1162/pres_a_00342

Weber, E. U. (2006). Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change*, *77*(1–2), 103–120. 10.1007/s10584-006-9060-3

Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, *24*(6), 1319–1333. 10.1080/10494820.2014.994533

Wen, Y. (2021). Augmented reality enhanced cognitive engagement: Designing classroom-based collaborative learning activities for young language learners. *Educational Technology Research and Development*, *69*(2), 843–860. 10.1007/s11423-020-09893-z

Yoganathan, S., Finch, D. A., Parkin, E., & Pollard, J. (2018). 360 virtual reality video for the acquisition of knot tying skills: A randomised controlled trial. *International Journal of Surgery*, *54*, 24–27. 10.1016/j.ijsu.2018.04.002

Zheng, L., Xie, T., & Liu, G. (2018, December). Affordances of virtual reality for collaborative learning. In *2018 International Joint Conference on Information, Media and Engineering*, pp. 6–10.

Supplementary Material

Table S1. *Output of Cumulative Link Mixed Model Regression of Responses about Engagement*

Cumulative Link Mixed Model fitted with the Laplace approximation								
formula: Engagement ∼ Condition * Role + (1 | Subgame/ID)								
	link	threshold	nobs	logLik	AIC	niter	max.grad	cond.H
	logit	Flexible	38	−33.35	82.70	292(595)	5.65e-05	2.2e+01
Random effects:								
Groups	Variance	Std.Dev.						
ID:Subgame(Intercept)	0.541	0.735						
Subgame(Intercept)	<.001	<.001						
Number of groups: ID:Subgame 20, Subgame 3								
Coefficients:								
	Estimate	Std.Error	z-value	p-value				
Condition (Non-Collab vs Collab)	**1.9926**	**0.9274**	**2.149**	**0.0317**				
Role (in-VR vs out-VR)	**−2.4897**	**1.1765**	**−2.116**	**0.0343**				
Condition:Role	−0.4756	1.4736	−0.323	0.7469				
Threshold coefficients:								
	Estimate	Std.Error	z-value					
Quite unengaged | Neither	−2.9096	0.9975	−2.917					
Neither | Quite engaged	−1.4302	0.7110	−2.012					
Quite | Very engaged	1.7901	0.7840	2.283					

Table S2. *Output of Cumulative Link Mixed Model Regression of RESPONSES about How Much They Had Listened to the Polar Bear*

Cumulative Link Mixed Model fitted with the Laplace approximation								
formula: Listening ~ Condition * Role + (1 | Subgame/ID)								
	link	threshold	nobs	logLik	AIC	niter	max.grad	cond.H
	logit	flexible	38	−33.70	83.39	359(1411)	1.56e-05	9.6e+01
Random effects:								
Groups	Variance	Std.Dev.						
ID:Subgame(Intercept)	20.130	4.487						
Subgame(Intercept)	<.001	<.001						
Number of groups: ID:Subgame 20, Subgame 3								
Coefficients:								
	Estimate	Std.Error	z-value	p-value				
Condition (Non-Collab vs Collab)	−3.71535	2.22668	−1.669	0.0952				
Role (in-VR vs out-VR)	−0.79946	1.90084	−0.421	0.6741				
Condition:Role	−0.09511	2.37385	−0.040	0.9680				
Threshold coefficients:								
	Estimate	Std.Error	z-value					
Not at all | Not so much	−9.9782	4.6335	−2.153					
Not so much | Quite a lot	−5.1016	2.8817	−1.770					
Quite a lot | Very much	0.9479	1.8711	0.50					

Table S3. *Output of the Logistic Binomial Mixed Regression Model of Scores (Correct/Incorrect) on Multiple Choice and Fill-in-the-Blank Questions*

Formula: Score ~ Time + Time:Condition + Time: Subgame.match + (1 | ID) + (1 | Question)					
	AIC	BIC	logLik	deviance	df.resid
	696.3	723.0	−342.1	684.3	634
Scaled residuals:					
	Min	1Q	Median	3Q	Max
	−2.9955	−0.6651	−0.2125	0.6822	4.3459
Random effects:					
Groups	Name	Variance	Std.Dev.		
ID	(Intercept)	0.3684	0.6069		
Question	(Intercept)	2.9810	1.7265		
Number of obs: 640, groups: Question, 20; ID, 16					
Fixed effects:					
	Estimate	Std. Error	z-value	p-value	
(Intercept)	−0.8773	0.4441	−1.976	0.0482	
Time	**0.9522**	**0.4101**	**2.322**	**0.0203**	
Time:GroupCOLLAB	−0.1735	0.3717	−0.467	0.6407	
Time:Subgame.matchFALSE	−0.6136	0.3257	−1.884	0.0596	