Journal of Universal Computer Science, vol. 24, no. 2 (2018), 220-245
submitted: 2/1/18, accepted: 19/1/18, appeared: 28/2/18 © J.UCS

Comparing Learning in Virtual Reality with Learning on a 2D Screen Using Electrostatics Activities

Scott W. Greenwald
(MIT Media Lab, Cambridge, MA, USA
scottgwald@media.mit.edu)

Wiley Corning
(MIT Media Lab, Cambridge, MA, USA
wjlc@media.mit.edu)

Markus Funk
(Technische Universität Darmstadt, Germany
funk@tk.tu-darmstadt.de)

Pattie Maes
(MIT Media Lab, Cambridge, MA, USA
pattie@media.mit.edu)

Abstract: Virtual Reality (VR) has the potential to overcome natural constraints and present things that would not be visible in the physical world. This makes the medium of VR a powerful tool for learning that allows users to become highly immersed in complex topics. In this paper, we compare a VR learning environment with a traditional 2D learning environment. To investigate the differences between VR and 2D learning environments, we designed two activities that help learners gain an intuitive understanding of concepts from electricity and magnetism. We performed an experiment comparing the learning that took place using these two variant. Although our quantitative learning measures did not reveal a significant difference between 2D and VR, VR was perceived by learners to have advantages. We did find significant quantitative differences in learners' completion times. We share findings, based on the quantitative and qualitative feedback received, about what makes VR environments beneficial for learning about complex spatial topics, and propose corresponding design guidelines.

Key Words: Virtual Reality (6DoF), Learning Environments, Mixed Methods, Varied Practice, 2D vs. VR, Physics Education Research, Simulation

Category: H.5.m, L.3.6, L.5.0

1 Introduction

The potential for Virtual Reality (VR) to provide novel learning experiences has long been recognized. VR can allow learners to go places they can't go physically, shrink or grow to see the world at different scales, control the passage of time, and hear or see information right when and where it is needed. Related work has found that VR can be beneficial for learning [Dede et al. 1996], but the benefit is only seen when the unique attributes of VR are leveraged. The relevant studies have focused on proving out the viability of using VR for certain learning approaches and content, but have not attempted to isolate the specific attributes of the medium that contribute to the benefits obtained. Three highly salient categories of research on the use of VR for learning moving forward are: (1) Identifying and proving out the fundamental advantages of the VR medium. Studying the consequences of linking sensorimotor activities directly with dynamic simulations is one such avenue of research to which the current paper contributes. There may also be advantages to accessing existing materials in a way that is more intuitive, eliminates distractions, or decreases extraneous cognitive load [Regian et al. 1992]. (2) Exploring how to apply existing instructional strategies, or inventing new ones, that uniquely leverage the combination of deep similarity to the real world (human-scale or otherwise) and the great freedom to design information presentation afforded by VR [Dede et al. 1996, Roussou et al. 2006]. This includes learner pacing, smart tutoring that is contingent on attention [Hubbard et al. 2017], merging spatial representations [Rau et al. 2015, Ainsworth 2006], and facilitating novel forms of collaboration with peers and experts [Karutz and Bailenson 2015, Greenwald et al. 2017]. (3) Applied design knowledge. Field research based on the combination of the two previous categories should give rise to design guidelines and specific knowledge about subject areas and usage scenarios where the advantages of VR are the greatest [Pantelidis 2010, Seidel and Chatelier 2013].

In this paper, we conduct a study making a side-by-side comparison of the learning that takes place using a standard VR system compared with a standard 2D system. For the best comparison between the two modalities, the interactions and the content are designed to be as similar as possible. We contribute to the field of learning in VR by reporting the findings of our comparative study and by discussing the implications of the findings for VR learning experiences. To position the question of the possible implications of this study for teaching practice, the intention is to identify the characteristics of subject matter where VR is advantageous, through the study of one specific example. If such advantages are found, there are many ways that VR could be integrated into instructional strategies and teaching practices. On one end of the spectrum, it is possible that a great deal of teaching and learning could take place in VR. This would be the case if some intrinsic advantages of VR, such as increased engagement and eliminating distraction, were to make it VR advantageous even when not making use of the 3D aspects. On the other end of the spectrum, VR could be an occasional supplement to existing techniques, which would help to make particular concepts more accessible and intuitive. In the context of the work presented here, we are agnostic to this question of real-world integration; focusing instead specifically on the question of where learning advantages might be found.

2 Related Work

We will first give a brief history of the idea of immersive, simulation-based learning environments, tracing it through earlier applications of computers to learning including exploratory learning and virtual labs. Then we will discuss research on learning procedural and spatial knowledge in VR. Finally, we frame the current study in terms of the learning theory of varied practice.

As early as 1959, some began to take seriously the idea that computers could transform education [Alpert and Bitzer 1970, Kay 1972, Smith and Sherwood 1976], calling it "science fiction" while stating that the trajectory of technology would "almost guarantee" it to happen [Kay 1972]. It seems that the idea of supporting highly non-linear learning processes only emerged later, initially to solve the problem of learning complex computer interfaces. The term "exploratory learning" [Rieman 1996] originally referred to an alternative method for this application that was investigated starting in the the early 1980's. This work demonstrated computer systems explicitly designed to encourage exploration, using fun and pleasure to assist in onerous learning tasks [Carroll 1982, Malone 1982, Shneiderman 1983]. The term gradually broadened in scope to eventually refer to learning any topic in a learning environment that gives the learner the autonomy to choose what to interact with and how to interact with it, although it does always seem to refer to settings that are computer-mediated [Mavrikis et al. 2013]. A key idea for applying this idea to learning about science–not just computer interfaces–is to use interactive simulations. One of the first examples of research on learning from simulations in higher education used the domain of control theory [Njoo and De Jong 1993]. Students were more successful at learning from this process when they were given more guidance (in this case, hypotheses to test). A recent review including this and much subsequent work on virtual labs indicated proper guidance as the single most important factor in their successful application [de Jong et al. 2013], consistent with work on exploratory learning outside of virtual labs [Hmelo-Silver et al. 2007, Kirschner et al. 2006]. More recently, multiple researchers have used online platforms such as Second Life to create multiuser virtual learning environments, albeit without the immersive characteristics of VR
[Thomas and Mead 2008, McCaffery et al. 2014].

By 1995, the vision of immersive, simulation-based learning environments that would allow learners to interact with local and remote others in real-time was clearly articulated [Dede 1995]. The ensuing research focused in particular on the ability of these kinds of environments to facilitate constructivist learning [Dede et al. 1996, Roussou et al. 2006]. A review of this work from 2009 concludes that the most salient strengths of immersive virtual environments for learning are their ability to (1) show multiple perspectives, (2) facilitate situated learning, and (3) transfer to the real world [Dede 2009].

Next, prior work on the acquisition of procedural and spatial knowledge in VR probes the learning properties of the medium in a more focused way. A sequence of studies conducted by the US Air Force, published starting in 1992, explored the use of VR simulation-based training. The first pair explored learning navigational and procedural knowledge of small-scale and large-scale spaces, and compared knowledge of real-world spaces with knowledge of virtual spaces acquired during an experiment. Both studies indicated that the use of VR was successful and the knowledge acquired was comparable to that acquired in the real world [Regian et al. 1992, Regian and Yadrick 1994]. A third study, however, compared the transfer from 2D to the real world with that from VR to the real world and found no significant difference [Regian 1997]. The tasks were procedural and navigational. It could be argued that neither took real advantage of the 3D aspect of the environment. The procedural task involved operating a console, which entailed memorizing a sequence of button presses and knob turns. The other involved navigating around a building with two levels, which is technically 3D, if only in a rudimentary way. A later study introduced a more sophisticated 3D navigational task–navigating a system of tunnels–and did find that VR was advantageous [Schuchardt and Bowman 2007].

In our study, we propose a new framing for the potential advantages of VR over 2D, using the concept of varied practice. Early work in this area showed that varied practice outperforms specific practice in the case of physical skills [Kerr and Booth 1978]. A seminal experiment showed that children who trained at the task of throwing items into buckets placed at two feet and four feet away outperformed children trained only on buckets three feet away when tested on the task of throwing items into buckets three feet away [Kerr and Booth 1978]. More recent research has shown that the advantage of varied practice applies to cognitive skills as well; one study used anagrams as its experimental activity with an experimental design similar to the bucket study [Goode et al. 2008]. An overview of related research is provided by Brown, et al. [Brown et al. 2014].

Relating this to the present study, we design a VR activity with a corresponding 2D activity and propose viewing the 2D activity as a specific subset of the VR activity. For any 3D activity that can be constrained to a plane, this framing is logically and mathematically valid. However, the question of its empirical utility is separate one. One advantage is that it provides a prediction, grounded in cognitive science, that 2D performance can be improved through training a skill in VR. At present, the most widely accepted way to prove the efficacy of a new instructional technology is to improve learners' performance on existing assessments. Second, it provides a conceptual method for constructing examples that allow for side-by-side comparison of 2D with VR interfaces, which will help the further study of this area. This is only one of many approaches, though, since it is less applicable to the equally interesting and important question of whether it is advantageous to accessing the same 3D content through a VR headset versus a mobile device. This cannot be considered varied practice, since the use of a 2D touch screen interface to access a 3D space is qualitatively different from the use of a six degree-of-freedom input device for the same purpose – as such it is something different, rather than a subset. The fact that a VR activity focused on a cognitive task engages both motor skills and abstract cognition would seem to suggest that the effect should be present.

VR has the potential to make difficult spatial concepts more accessible, and to deepen the learning that takes place. By eliminating extraneous mental translation steps between 2D and 3D and between spatially separate representations of the same content, learners can focus more of their energy on conceptual understanding. By facilitating direct spatial interaction, VR also eliminates extraneous cognitive load associated with operating the user interface. In this paper, we provide a method for quantifying these advantages as an aggregate through analogous 2D and VR activities.

3 Design of Activities for VR and 2D

In order to avoid a complex experiment design, we chose to design activities that were intended to help users develop intuition for spatial phenomena, as opposed to activities that would help them remember formulas or solve algebraic problems. Our chosen activities permitted us to quantify performance (in terms of completion speed and number of attempts) and to create a multiple-choice test based on the activities themselves. In addition, these activities needed to be easily adaptable to both VR and 2D interfaces in a way that would keep the scale and nature of the interaction highly analogous. Furthermore, although collaborative applications are relevant to our research interests, we chose to design single-user activities here in order to maintain a tighter scope for the experiment. We designed two activities that fit these criteria.

3.1 Activity Subject Matter: Electrostatics

Electrostatics is a subject that is introduced in most high school physics classes, and treated in further detail in introductory university physics courses on electricity and magnetism. Among the basic principles of electrostatics, we incorporated the ideas that there are positive and negative charges; that like charges repel, while opposite ones attract (e.g. positive charges repel one another); that charges have numerical magnitudes; that the strength of attraction or repulsion is proportional to the product of the two charge magnitudes, and inversely proportional to the square of the distance between them; and that diagrams using electric field lines can illuminate how a configuration of charges in space will evolve over time [Purcell 2013].

These principles are easy to state, but gaining an intuition for them can be challenging. How does one imagine something that is proportional to the inverse of the square of a distance? What kinds of insights should a particular field line diagram reveal? We made these kinds of intuition the subject of our activities because (1) the domain is abstract and the imagery unambiguously spatial; and (2) exploring and showing comprehension only require using direct spatial interaction, to position charges in 2D or 3D space. This way there are no irrelevant effects to confound our results, like changing modes of cognition or interaction (e.g. changing mode to solve an equation or perform a numerical evaluation).

3.2 Activity I: Target Hitting

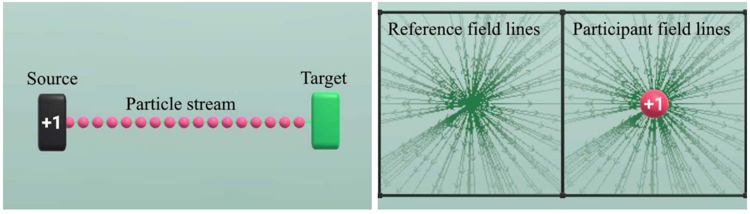
In the target hitting activity, a beam of charges must be redirected to intersect a given target. The dynamics of the particle beam follow a physical model of electrostatic interaction in which like charges repel, opposites attract, and the strength of these forces is proportional to the inverse of the square of the distance (as described above). The participant is given a certain set of charges with different positive and negative magnitudes to place in space in order to accomplish this goal. There are given positions where they may be placed, but the beam trajectory responds smoothly and continuously as they move through space. A diagram of the activity is shown in Figure 1.

Our reasons for constraining the user to choose from predefined positions–rather than being able to place charges anywhere in the continuous 3D space–were twofold. First, we wanted to encourage planning and reflection, which we found to be critical to learning. In piloting the activity without predefined positions, we observed that participants would move particles randomly around the space without developing a strategy. Sometimes this led to frustration, while at other times they would eventually discover the need to plan and reflect. With predefined positions, it seemed that participants would more quickly discover a reflective approach. After initially moving particles quickly and randomly from position to position, they would realize the combinatorial complexity of the set of possibilities (e.g. that there are 42 ways to place two different particles in seven possible locations) and resort to a more thoughtful strategy. Second, the use of predefined positions supports useful metrics for analysis, allowing us to examine the quantity and timing of the attempts made by each user while completing the activity.

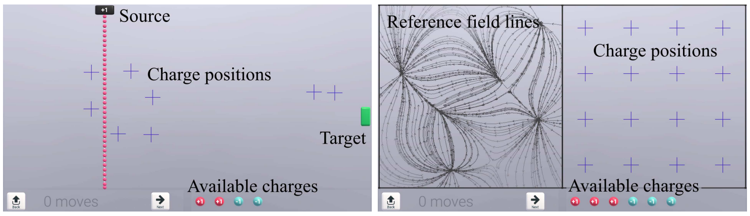
The design of individual exercises posed an interesting problem that was also solved through iterative design. Each exercise is parametrized by a set of charges to place and a set of fixed positions at which those charges may be placed. We initially experimented with a large number of given charges (e.g. five) and a large number of predefined positions (e.g. 20). We observed that such activities would take a long time to solve, and would induce a great amount of frustration. This also made the time to completion and number of attempts highly subject to individual variation. We also tried introducing "extra" charges that would be left over and not used in the solution, but this exacerbated the problem of difficulty. With these considerations in mind, we settled on exercises with one to four charges and two to ten allowed positions, in which all the charges provided were required to complete the activity. Figure 2 illustrates these parameters in a typical layout.

3.3 Activity II: Field Matching

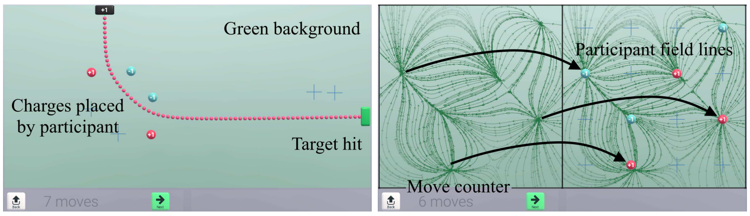
The field matching activity is similarly based on the idea of moving particles to predefined positions in space in order to accomplish a goal. In field matching, the goal is to generate a particular configuration of electric field lines by placing particles in the correct positions, as shown in Figure 1. This was the second activity to be designed; we



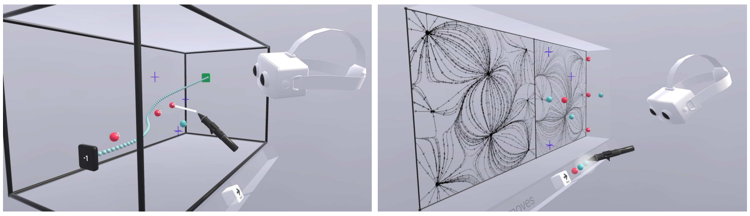
(a) Goal of activity



(b) Initial state of an exercise



(c) Completed state of an exercise



(d) VR interface with avatar interacting

Figure 1: The activity shown in the left column is Target Hitting, and in the right column is Field Matching. These are the basis for our learning study comparing 2D with VR. From top to bottom, the subfigures illustrate the goal of each activity, the initial and completed states of a specific exercise, and the corresponding VR interface.

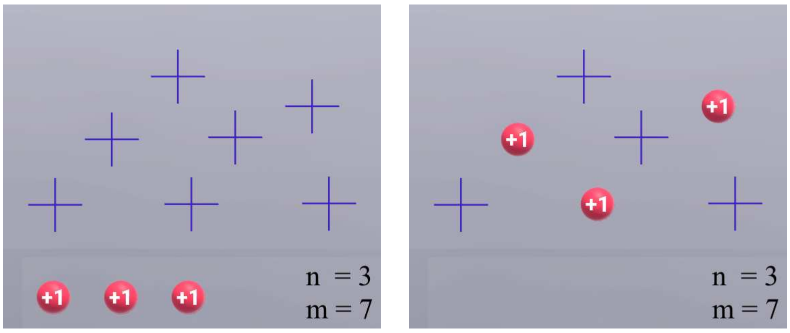


Figure 2: Exercise Parameters: n = number of charges to place, m = number of predefined positions. Exercise shown before (left) and after (right) placement.

used sets of predefined positions for the same reasons cited above in the discussion of the target hitting game.

The set of parameters used in designing exercises for the field matching activity were different from those of the target hitting game. In particular, the field matching activity is generally easier since the field line projections are more likely (although not guaranteed) to reveal where particles must be located. Therefore, the challenge was based more on recognizing attributes of like versus opposite charges that are neighboring, and, in more advanced configurations, recognizing the ways in which adjacent particles can lead to a field configuration where it is *not* obvious in which positions particles must be located. Therefore we increased the number of charges the user needed to place, but not as much the number of possible positions, settling on the range of two to nine particles to be placed in four to twelve positions.

3.4 Activity Interfaces in VR vs. 2D

Here we describe how the VR and 2D activity interfaces compare, enumerating the attributes of VR and 2D which are the same or different. These are summarized in Table 1.

Beginning with the differences, the VR interface supports the following attributes, while the 2D interface does not: visual and auditory immersion, the use of stereoscopic 3D display, and the use of head movement as an input. Moving on to the similarities, both activities are based on the task of moving charges to positions in space, so we were able to use the same mapping between input methods. The VR interaction requires using the controller's trigger button, while the 2D interaction uses the pen down and pen up actions to initiate and complete charge movement. The VR controller was physically larger, which we were not able to adjust for, but we did make the visual representation analogous. Lastly, we made the interaction space the same physical scale in both VR and 2D, with the cross-sectional area of the VR interaction space matching that of the

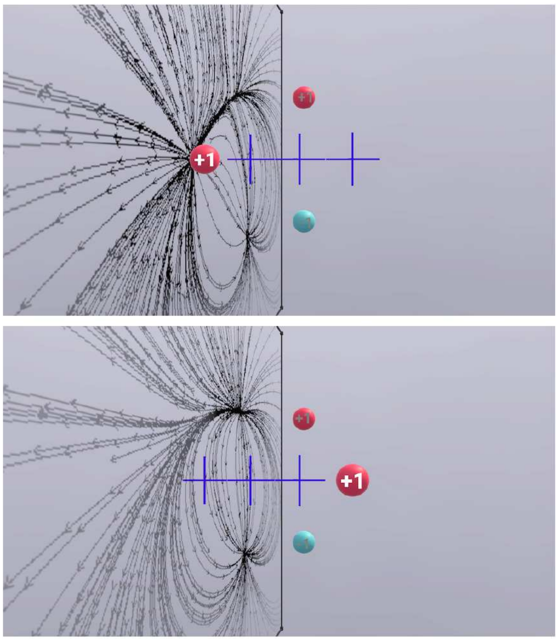


Figure 3: Effect of differing z-axis placement in field matching activity. A +1 charge is placed closer (top) or further (bottom) from the field line plane, leading to differences in the pattern of field lines.

Attribute	VR	2D
Immersion	Yes	No
Stereo 3D display	Yes	No
Head movement input	Yes	No
Direct manipulation	Yes (6DoF)	Yes (2D)
Primary input	Trigger, move	Pen down, drag
Interaction area/volume	Size of tablet	Size of tablet

Table 1: An overview of the attributes of both VR and 2D interfaces, which we were using in our experiment.

2D interaction space.

4 Experiment Comparing Learning in VR and 2D

We sought to compare the effect of 2D vs. VR interaction on learning. In our experiment design, each participant completes each of the two activities, one in 2D and the other in VR. We counterbalanced the order in which the VR and 2D activities were performed,

Figure 4: We designed the 2D and VR systems in a way that they also have similar input affordances for both modalities.

Group	ACTIVITY 1	ACTIVITY 2
A	TH-2D	FM-VR
B	TH-VR	FM-2D
C	FM-2D	TH-VR
D	FM-VR	TH-2D

LEGEND: TH=Target Hitting, FM=Field Matching

Table 2: Order of Activities

giving us a total of four participant groups as shown in Table 2. A two-part multiple choice pre/post test assesses participants' competencies associated with each of the activities.

4.1 Method

4.1.1 Hypothesis

We hypothesized that learners would do better on the multiple choice test and complete activities faster and with fewer moves when trained in VR, both (i) immediately after VR training and (ii) two weeks after VR training.

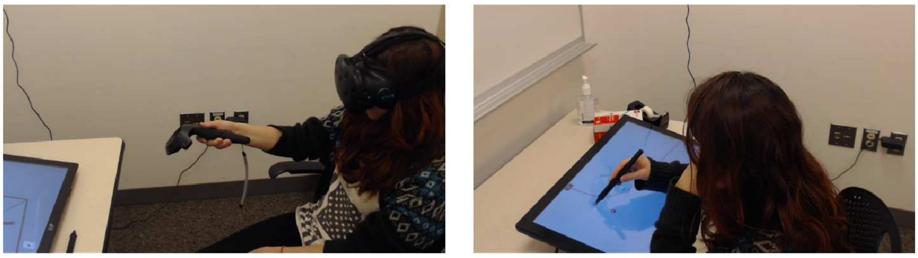


Figure 5: We conducted both conditions in the same physical space while sitting in a chair. In the VR condition (left), the chair was moved away from the table. In the 2D condition (right), the chair could be positioned closer to the 2D display by the participants.

4.1.2 Independent Variables

The independent variables in the experiment design are the modality (2D or VR) and the activity (target-hitting or field matching).

4.1.3 Dependent Variables

In order to measure the effect of the independent variables, we measured the following dependent variables:

- Completion times and number of moves required during testing activities, during first and second sessions.
- Two-part multiple choice test, performed three times: as a pre/post test for the first session, and at the beginning of the second session. Each of the two separate parts were relevant to one of the activities. The identical test is given all three times.
- Raw NASA-TLX [Hart 2006] (perceived cognitive load) was measured for every 2D or VR activity session.
- Text-based questionnaire comparing activities and 2D/VR

4.2 Apparatus

The 2D activities are completed on a 22-in drawing tablet (Monoprice 114481). Moves were performed using a touch-and-drag interaction with the stylus. The VR activities run on the HTC Vive and use the controller's trigger for clicking and dragging charges in space. Participants are shown in the two conditions in Figure 5, and the 2D and VR designs for the two activities are shown in Figure 1.

4.2.1 Multiple Choice Test

The multiple choice test contained static graphical representations of systems that corresponded to the dynamic interactive activities used in both style and content. One example is shown in Figure 6. Half of the questions were relevant to the target hitting game, and the other half were relevant to the field matching game. These questions were appropriate for use in a pre-test, since the basic representations (charged particles, electric fields) are standard.

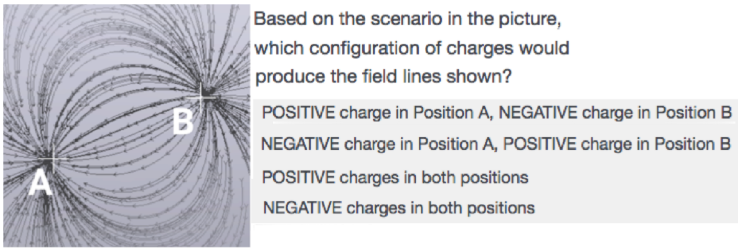


Figure 6: An example of a multiple choice question as it was used in our evaluation. Each multiple choice question contains a picture and four answers from which the participants need to chose one.

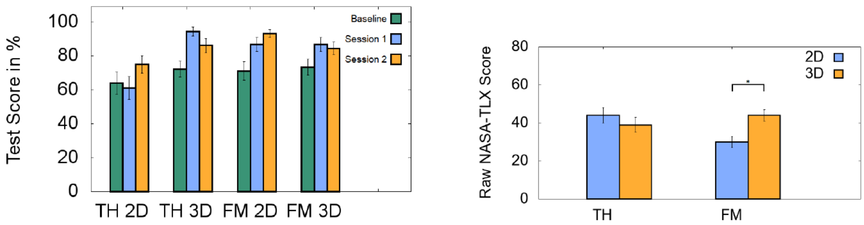


Figure 7: The test scores in % according to the two modalities (2D and VR) and the different games (TH = Target Hitting, FM = Field Matching) and the Raw NASA-TLX scores according to the different modalities and activities.

4.3 Participants

We invited 20 participants (aged between 18 and 22 years) to take part in this experiment. They had all taken our university's electricity and magnetism course in the past three years, so they had a familiarity with the subject matter of the activities. Participants were split into 4 groups according to the counterbalanced orders of the activities

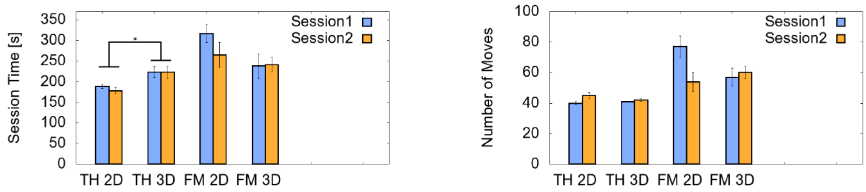


*Figure 8: The session times in seconds and the number of moves according to the two modalities (2D and VR) and the different games (TH = Target Hitting, FM = Field Matching). Differences that are marked with * indicate statistical significance.*

and modalities. Due to scheduling issues, two of the participants did not complete the experiment and were excluded.

4.4 Procedure

Participants were required to attend two sessions, separated by 12 to 16 days. The sequence of activities in the first session was as follows: they were given the multiple-choice pretest, completed each of the two activities in the modality appropriate to their group with training (modality varied) and testing (always 2D) portions, completing the TLX questionnaire after each testing segment, and finally they were given the multiple-choice post-test (identical to the pre-test). Participants were told they were being timed at the beginning of the overall session, but they were not explicitly reminded between activities.

In the second session, the sequence of activities was as follows: participants completed the multiple-choice test, and testing sets for each game, completing a TLX questionnaire directly after each. At this point, all of the tasks relevant to our quantitative analysis were complete, and some additional tasks were performed to gather further qualitative feedback. In particular, we wanted participants to be able to contrast the two activities and the two modalities. Recall that they had each tried only one of the two activities in VR during the earlier portion of the study. Therefore, participants were next given exercises for both activities in VR (one of which they had not seen before). These exercise sequences were shorter than the the ones during the earlier part of the study, since the idea was for them to gain experience with both activities in VR, not to assess their performance as before.

Finally, they were given a free-response, text-based questionnaire about the differences between the activities and the VR vs. 2D experiences. The complete list of questions is given in Table 4. A complete list of activities performed (as an example) by a participant in Group A is shown in Table 3.

Sess	#	Activity
1	1	Inbound survey
1	2	Multiple choice pre-test
1	3	TH-2D training
1	4	TLX
1	5	TH-2D testing
1	6	TLX
1	7	FM-VR training
1	8	TLX
1	9	FM-2D testing
1	10	TLX
1	11	Multiple choice post-test
2	1	Inbound survey
2	2	Multiple choice post-test
2	3	TLX
2	4	TH-2D testing
2	5	TLX
2	6	FM-2D testing
2	7	TLX
2	8	TH-VR
2	9	FM-VR
2	10	Text-based questionnaire

LEGEND: TH=Target Hitting, FM=Field Matching

Table 3: List of Activities for Group A Participants

4.5 Results

4.5.1 Quantitative Performance Results

We used an independent t-test to compare the session times and number of moves between groups. For comparing the NASA-TLX scores, we used a one-way ANOVA. We applied a Bonferroni correction for all post-hoc tests. We filtered the results for outliers by excluding data points with $\mu > 3 \times SD$.

4.5.2 Target Hitting

First, we analyzed the performance of the participants considering the multiple choice questions in the target hitting game. For the 2D modality, the participants scored 63.89% ($SD = 28.26\%$) for the baseline, 61.11% ($SD = 28.26\%$) for after the first game, and 75.00% ($SD = 21.65\%$) after two weeks. Considering the VR modality, the participants scored 72.22% ($SD = 19.54\%$) for the baseline, 94.44% ($SD = 11.24\%$) for after the first game, and 86.11% ($SD = 18.16\%$) after two weeks. An independent t-test revealed no significant difference between 2D and VR conditions, when comparing the percentage of improvement over the baseline for each session.

We compared the session times between the two sessions for the target hitting game. For the 2D modality, the times after the first session ($M = 189.56$s, $SD = 18.80$s) were higher than the times of the second session two weeks after ($M = 178.44$s, $SD = 26.06$s). Considering the VR modality, the participants took an average of 223.47s ($SD = 40.37$s) for the first session and an average of 223.80s ($SD = 43.76$s) after two weeks. Using VR led to statistically significant longer task completion times ($M = 223.8, SD = 40.38$) compared to using 2D ($M = 189.6, SD = 18.81$), $t(16) = 2.284, p = .005$ in Session I. In Session II, however, no statistically significant difference between VR ($M = 223.8, SD = 43.76$) vs. 2D ($M = 178.4, SD = 26.06$) could be detected, $t(16) = 2.672, p = .491$.

Further, we compared the number of moves that the participants made in the activities between the two sessions for the target hitting game. For the 2D modality, the number of moves in the first session ($M = 40.44$, $SD = 5.91$) were fewer than in the second session two weeks after ($M = 45.89$, $SD = 6.12$). Considering the VR modality, the participants made slightly fewer moves in the first session ($M = 41.33$, $SD = 2.44$) than in the session after two weeks ($M = 42.67$, $SD = 5.33$). An independent t-test revealed no statistically significant difference in the number of moves between using VR ($M = 34.3, SD = 2.45$) vs. 2D ($M = 40.4, SD = 5.92$) in Session I, $t(16) = .416, p = .174$. Similarly, for Session II, there was no statistically significant difference between using VR ($M = 42.7 SD = 5.34$) vs. 2d ($M = 45.4, SD = 6.13$), $t(16) = -1.026, p = .916$.

Finally, we compared the perceived cognitive load between the VR and the 2D variant of the target hitting activity using the raw NASA-TLX score. The 2D variant of the training activity ($M = 44.56$, $SD = 12.79$) was perceived as a little more cognitively demanding compared to the VR variant ($M = 39.11$, $SD = 13.57$) of the game. However, a one-way ANOVA test could not reveal a significant difference.

4.5.3 Field Matching

Also for the field matching game, we analyzed the participants' performance according to the modality and the trial. For the 2D modality, the participants scored 71.11% ($SD = 22.60\%$) for the baseline, 86.67% ($SD = 17.32\%$) after the first activity, and 93.33%

(SD = 10.00%) after two weeks. Considering the VR modality, the participants scored 73.33% (SD = 20.00%) for the baseline, 86.67% (SD = 17.32%) for after the first activity, and 84.44% (SD = 16.17%) after two weeks. An independent t-test revealed no statistically significant difference between 2D and VR conditions, when comparing the percentage of improvement over the baseline for each session.

We compared the session times between the two sessions for the field matching activity. For the 2D modality, the times after the first session (M = 317.21s, SD = 66.82s) were higher than in the second session two weeks after (M = 265.57s, SD = 90.91s). Considering the VR modality, the participants were faster in the first session (M = 238.62s, SD = 92.53s) than in the session after two weeks (M = 241.46s, SD = 54.72s). An independent t-test revealed no statistically significant difference in the task completion time between using VR ($M = 238.6, SD = 92.53$) vs. 2D ($M = 317.2, SD = 66.82$) in Session I, $t(16) = -2.065, p = .423$. Similarly, for Session II, there was no statistically significant difference between using VR ($M = 241.5, SD = 54.72$) vs. 2d ($M = 265.6, SD = 90.91$), $t(16) = -.681, p = .093$.

Further, we compared the number of moves that the participants made in the activities between the two sessions for the field matching activity. For the 2D modality, the number of moves in the first session (M = 77.44, SD = 21.57) were higher than in the second session two weeks after (M = 54.89, SD = 19.98). Considering the VR modality, the participants made slightly fewer moves in the first session (M = 57.22, SD = 20.02) than in the session after two weeks (M = 60.22, SD = 14.98). An independent t-test revealed no statistically significant difference in the number of moves between using VR ($M = 57.2, SD = 20.02$) vs. 2D ($M = 77.3, SD = 21.58$) in Session I, $t(16) = -2.061, p = .88$. Similarly, for Session II, there was no statistically significant difference between using VR ($M = 60.22, SD = 14.99$) vs. 2d ($M = 54.9, SD = 19.98$), $t(16) = .64, p = .705$.

Finally, we compared the perceived cognitive load using the raw NASA-TLX score for the field matching activity. The 2D variant of the activity (M = 30.22, SD = 10.12) was perceived as less cognitively demanding compared to the VR variant of the activity (M = 44.78, SD = 10.50). A one way ANOVA test revealed a significant difference between the 2D and VR variants, $F(1,8) = 17.135$, $p = .003$. The effect size estimate shows a large effect ($\eta^2 = .682$).

4.5.4 Free Response Questionnaire Results

In this section, we take a factual approach to reporting the results before bringing them into the context of an overall discussion in the next section. Three of the free-response questions were concerned with the comparison between the two activities, while the other three were concerned with the comparison between 2D and VR interaction; these sets are discussed separately in this section. The text of the questions is shown in Table 4.

	Category	Question
Q1	Activities	Was one activity easier than the other? If so, why?
Q2	Activities	Do you feel like you learned more from one activity than the other? If so, why?
Q3	Activities	Was one activity more engaging or fun than the other? If so, why?
Q4	2D/VR	How did your experience doing TH (the activity with the particle stream) differ between the 2D version and the VR version?
Q5	2D/VR	How did your experience doing FLM (the activity with the field lines) differ between the 2D version and the VR version?
Q6	2D/VR	In general, how did the VR interface compare to the 2D interface?

Table 4: Free Response Questions from Second Session

Answer	Q1 (easier)	Q2 (learned more)	Q3 (more engaging)
TH	0	8	13
FM	18	4	2
Neither/same	0	6	3

Table 5: Free Response Results for Questions Comparing the Two Activities

4.5.4.1 Comparing Target Hitting with Field Matching

The three free-response questions concerning the comparison between activities inquired respectively about comparative (Q1) difficulty (which was easier), (Q2) amount of learning, (Q3) engagement. Since these questions asked the participant to choose one activity or the other, the responses could be coded according to their choices, and additional nuance obtained from their elaborations. Table 5 shows participants' coded responses.

There was a complete consensus among participants that the field matching activity was the easier of the two (Q1). There was a mixed response about whether they learned more from either activity (Q2), or equally from both. The largest group said that they learned more from the harder activity, but this group constituted less than half of the respondents. The questions about engagement (Q3) exposed participants' ambivalence about being challenged. Specifically, 72% of respondents said that the target

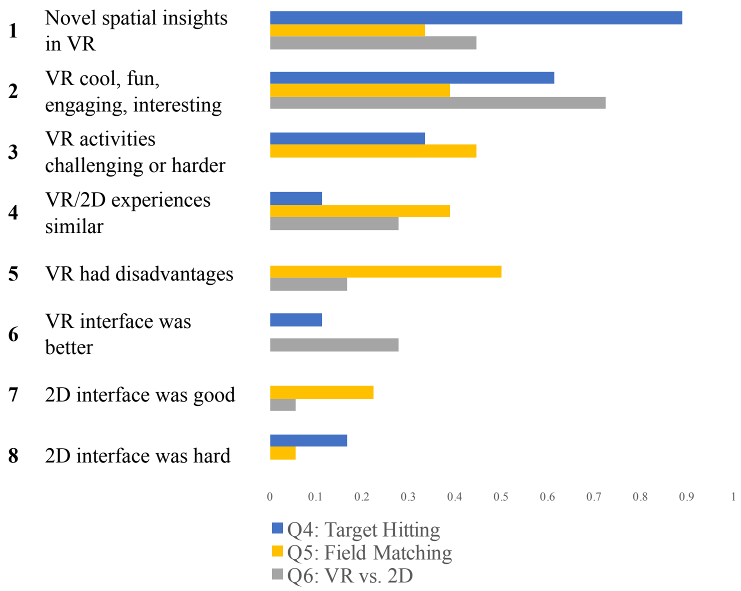


Figure 9: Free Response Codes. Horizontal axis shows percentage of respondents who mentioned these in their responses

hitting activity was more engaging, but 33% explicitly (6 participants) mentioned the tension between feeling satisfaction while doing well at the easier activity, and the rewarding but also frustrating experience of doing something more novel and difficult while doing the harder activity.

4.5.4.2 Comparing the 2D and VR experiences

The remaining three of the free-response questions aimed to uncover differences between the 2D and VR experiences. Unlike the first three, these were formulated as "how" questions (see Table 4). To analyze the results, we applied a two-cycle, simultaneous coding procedure. The first cycle established an initial, in vivo coding. The second cycle used a holistic pattern coding method in order to provide both a summary of the content and a point of reference for the discussion of specific quotes [Saldana 2015]. The second cycle codes and results are shown in Figure 9. As a preface to the discussion, we comment on how these results should be technically interpreted. The key point is that every code reported was a spontaneous response to a broad question. For example, one third of respondents mentioned that the *VR activity was challenging or harder* (code 3) in response to *Q4: How did your experience doing TH differ between the 2D version and the VR version?*. The fact that six different people spontaneously made the same point is very strong evidence in favor of the point. This is in contrast with the similar-sounding result "one third of respondents agreed with the statement" which might result from a Likert-scale questionnaire and lead to the opposite conclusion.

To begin, we'll discuss *Q4*, which concerned the 2D vs VR experience of the Target Hitting activity. The *Code 1* rate of 88% indicates that all but two participants reported novel spatial insights in VR. Some examples are: *"The VR version's 3D aspect made the game more interesting, and also made the game feel more representative of a physical scenario."* and *"Having the particles streams in space allowed me to better visual [sic] the effect the added particles would have on them."* Next, 62% of respondents reported *Code 2*, that VR was fun, engaging, or interest. Some example statements: *"It looked super realistic and was more fun to play and experiment around with."* and *"It was more interactive which made it more fun."* [referring to VR].

The statements for *Code 3* (33% of participants) indicated that the VR activity was more challenging. Respondents referred to the inherent additional complexity of solving a 3D challenge compared with a 2D puzzle: *"Adding a 3D component made in [sic] much more challenging." and "The VR version is definitely harder because you have another dimension to take into account."* However, 16.7% of participants made a contrasting statement that the 2D version was harder. A closer look at these responses reveals that these participants were referring to the difficulty of the specific exercises they were presented with, as opposed to the inherent properties of the activity: for example *"I think some of the 2D puzzles were either harder or harder to visualize."* This contrast was made explicit in a single response which made both points *"The 2D version I thought was harder than the VR version ... I thought the VR version was more fun because there was a 3rd dimension that made it a little trickier."* The questionnaire was administered in Part 2 of the experiment (the second session, two weeks after the first), where the VR sessions were actually shorter and easier than the 2D versions. These sessions were included after the conclusion of the activities measuring performance, to give the participants a taste of both activities in VR, one of which they had not done in VR previously (as dictated by their experimental group).

Next we'll move on to *Q5*, which concerned the contrast between Field Line Matching in VR vs. 2D. Participants were not as enthusiastic about the value of this activity in VR. One third reported novel spatial insights–comparatively fewer than for Target Hitting. Two attributes of the field-line activity seem to account for this. Firstly, the activity was seen as being similar between 2D and VR versions, in particular because the field line plane was in 2D in both 2D and VR activities. Furthermore, the remaining 3D aspect–that the field lines in the field line plane result from charges outside the plane–was considered interesting by some, but confusing by others. The former contingent made statements such as *"I was really surprised because even after learning physics E&M in both high school and at our university, it never occurred to me that electron FLM vary in 3D, and it was just really interesting to realize."* and *"the 3D part of the VR version seems quite useful for getting an intuition of how field lines look in 3D."*, while the latter made statements such as *"the building of a 2 dimensional field line diagram in 3d space is a little confusing."*, *"The VR version was confusing at* *first."* and *"we still only had a 2D projection of the field lines, which was weird."*. Multiple participants highlighted that it was difficult to distinguish the influence of depth (Code 5): *"It was hard to figure out how far the charges are affected the 2D drawing."* and *"The effect of the charges on field lines in space was less clear."* [referring to VR version], and a third participant went further described how this led to a trial-and-error strategy *"it was tough to distinguish between the field lines when a charge is placed in two similar locations (that have a different depth). Because of this it required some trial and error."*

Finally, in *Q6*, we asked participants to reflect on the general difference between VR and 2D, with a focus on the interface. The most common response focused on the fun, engaging, and interesting aspects of the VR experience. *"It was more mentally stimulating and more fun. Whereas I got frustrated more quickly doing the 2D puzzles when I couldn't solve something, I was entertained when I couldn't solve the VR puzzles." "VR made the games more interesting and interactive, which made them more engaging."*

Comments on the advantages of the VR interface (Code 6) described it being easy, smooth, or natural: *"The VR interface felt a lot more natural and engaging than the 2D one. I liked being able to place the charges in a real way that made it easier for me to understand what my placement would do." "Overall I like the VR interface better than the 2D interface. It is easier to navigate and move around the particles." "I think it was also easier to move particles around in the VR version than 2D version."*, *"The VR interface felt very smooth and elegant."* An equal number of comments indicated that the 2D and VR experiences were similar (Code 4): *"Very similar interfaces. VR seemed sort of unnecessary - it was cool to have, but ultimately the same goal could be reached on the 2D interface."*, *"The VR interface and 2D interface were about the same in terms of being an intuitive / easy to use interface."*. Two participants distinguished explicitly between the activities, making both points, each in reference to one of the activities: *"I really enjoyed the VR interface for the particle stream game. I think having that extra dimension is really exciting and adds another way that the particles you place can mess up your streams. The field line matching one felt the same regardless of interface, but it'd be interesting to see more use of the fact that's in 3D otherwise I feel that the 2D interface accomplishes what the 3D one does in its current iteration without the need for the fancy setup." "The VR interface allowed for added complexity with respect to how charges interact. It helped with rounding out understanding for the beam game [target hitting], but may have added confusion to the FLM game."*

5 Discussion

Some of our results were expected, while others proved surprising. Taking all of our data into consideration, we are able to posit a coherent explanation for most of our unexpected results.

To reiterate two important aspects of the experimental setup, we will refer below to participants' "completion times" and "number of moves" during the testing phase, in which participants from both the VR and the 2D training groups used the 2D modality to complete the same set of activities. That is, the training phase takes place in different modalities, while the testing phase is identical in every way across all groups.

For the target-hitting activity, participants who trained in VR performed better – but not significantly better– on the multiple choice test: the VR group moved from 72% to 94% in their first session, while the 2D group stayed the same, scoring 64% and then 61%, as shown in Figure 7. At the same time, the VR-trained group took significantly longer to complete the 2D testing activities during the first session, while the number of moves tried did not differ significantly in either session. This implies a greater time between moves on average, painting a picture of a more pensive approach. Recall that there was no difference in the instructions each group received with respect to the timer or the importance of completing the activities quickly– which makes the difference in completion time more surprising. It might be possible to explain this difference as a consequence of learners being less familiar with the interface, but the reports that both interfaces were extremely easy to use would seem to contradict this interpretation. Future work should investigate how presenting a more immersive interfaces and more challenging activities might contribute to self-regulation and metacognition.

Participants did not perform better in VR for the field matching activity, and this is illuminated by the qualitative feedback provided by participants. In particular, the skill being trained (arranging charges to create a pattern of field lines) seems intuitive in 2D, but confusing and unfamiliar in VR. Participants mentioned never been introduced in their academic coursework to the concept of projecting field lines onto a plane. In hindsight, while this is a valid concept, it is not clear when or how it would be used practically. A more meaningful and engaging use of this concept would require a different activity design (e.g. trying to influence a particle constrained between two surfaces)– in the field matching activity used in this experiment, it was not sufficiently justified. In addition to that shortcoming, participants reported that the effect of depth in the VR version was too hard to discern, making its underlying meaning less interpretable. It is worth noting that, despite this, the VR-trained participants' approach in the testing phase was not significantly different from the 2D-trained participants', nor was their performance on the multiple choice test.

Our results here are consistent with the most closely related piece of prior work. A similar experiment was conducted in [Salzman et al. 1999], which compared a custom VR learning application with an established 2D one. While both of the groups in that study performed significantly better than baseline on a series of tests after training, neither group performed significantly better than the other. This supports the claim that learners are benefiting from the VR content, but not so much that a learning difference compared with the 2D interface was measurable an significant. Like the authors of that piece of work, we conclude that the perceived advantages that were strongly pointed to by participant feedback will be successfully quantified when (1) more is known about the design of interfaces and activities in VR, and hence better examples will be available to test, and (2) assessments can be devised that measure more difficult but more important aspects of conceptual learning through alternative means.

Taking a step back, we observe the noteworthy result that learners were equally well prepared for a 2D test activity after practicing an analogous VR activity than they were after practicing a 2D activity identical to the test. Participants were able to overcome the barrier of skill transfer that impedes knowledge acquired in one setting from being applied to a different setting. There was a correlated measurable difference in approach, as indicated by the fact that the VR-trained learners took more time to complete their 2D test activities. We believe that at least two effects are present here: the advantage of engaging with a 3D system in a visual and sensorimotor context, and the advantage of taking a more pensive and reflective approach to the learning process.

The difference in learning approach was not something that we predicted prior to the study; on the contrary, we had hypothesized that the VR learners would be faster and use fewer moves. In hindsight, however, it seems clear that the learning result assessed by the final multiple choice test need not correlate with the speed of completing the 2D test activities. As a general takeaway, note that although behavioral metrics in completing exercises can in fact give insight into the learning that is taking place, naïve interpretations of these metrics could lead to exactly the wrong conclusions. Had we not included the multiple choice assessment in our study, we may have concluded that the VR interface had been inferior at preparing students for the 2D activity, since they took longer to complete the test activities.

6 Guidelines for VR learning experiences

Through developing VR learning activities and through both quantitative and qualitative results of our evaluation, we identified two guidelines for designing VR learning experiences. Based on our study results, we conclude that following these guidelines can help to achieve learning benefits. Below, we state and explain these two guidelines. As written, they are most applicable to settings involving interactive spatial simulations and give special attention to cases where a design is being translated between 2D and VR. Even so, they may prove relevant when interpreted more broadly.

I The VR activity should use representations that are not possible in 2D – in particular, those that require a method of changing perspective (in VR, head movement) to be fully interpretable.

II The VR activity should use all relevant representations. When complexity is added to the activity through translation into VR, there must be corresponding additional representational support.

The rationale for Guideline I is two-fold. The first is that 2D representations are at a disadvantage from the perspective of input methods when changing perspective is necessary. Using the hands instead of the head for this induces extraneous cognitive load since it is less natural and requires learning. The second is that, according to the varied practice framing, there will only be a cognitive dimension of "variation" tied to VR if this is the case.

Regarding Guideline II, we hypothesize that the main failing of the field matching activity was that it provided too little support for the learner to connect the 3D field of the charge system with the 2D projection. The task itself involves a translation from 3D to 2D, but our design failed to show a relevant 3D representation. One of the most important advantages of VR is that representations that would be difficult to combine in 2D due to clutter can work well in VR since head movement can be used to change perspective. Overlapping or overlaid representations can be mentally separated without moving them out of place. In 2D, the 2D field line diagram that is included in the activity is a complete representation of the system in reference to the activity. Going from 2D to VR, we added complexity, but did not add corresponding support to deal with that complexity. It would be possible to add such support by displaying 3D field lines in place to assist in the translation process. For target hitting, in contrast, the additional representational support we provided was the display of the particle trajectory in 3D. In this case, extending the 2D activity to VR was straightforward and did not require special consideration.

7 Conclusion

In this paper, we presented several different contributions related to the theory and practice of learning in VR. We showed that people perceived many advantages of a VR interface over an analogous 2D interface, although the particular activities and assessment we applied did not demonstrate a quantitative difference in learning outcomes. This result is positioned at the intersection of HCI and the learning sciences. We made a methodological contribution in providing a varied practice framing for comparing learning between 2D and VR interfaces. As another methodological contribution, we made the observation that behavioral metrics related to activity performance provide important insights, but should not be assumed to measure conceptual understanding. We presented an experimental design in which a non-interactive 2D test (e.g. multiple choice) placed 2D and VR activities on equal footing for comparison. Finally, we provided design guidelines for VR learning activities, with a focus on interactive spatial simulations. Our key insight in defining these guidelines is that all relevant visual representations must be accounted for and included; those that are essential to the definition of the activity are not at risk of being omitted, so special attention must be paid to include all representations that could provide cognitive or conceptual support in completing the activity.

To conclude, we remark that the VR learning advantage demonstrated here may be the tip of a very large iceberg, one that others mentioned in the *Related Work* have begun to uncover as well. The set of experimental activities constituted less than half an hour of learning, and we limited the use of immersion, 3D, head and hand-input in a way that could be easily mapped to a 2D analogue. It is conceivable that the gains we measured would be compounded through longer and less constrained use of the affordances of VR. Our design guidelines make an initial effort at characterizing the extent of the iceberg, i.e., which topics stand to gain the most from being learned in VR. Further developing such knowledge and guidelines is surely a fruitful and exciting area for future work.

Acknowledgements

A very special thanks to John Belcher for his enthusiasm and dedication since the very beginning of this project; Gabriel Fields for his hard work on the predecessor system; Peter Dourmashkin, Michelle Tomasik, Saif Rayyan, Ted Selker, Lei Xia, Misha Jamy, and Tilman Dingler for their valuable input and contributions. Many thanks to MIT Open Learning for their generous support.

References

[Ainsworth 2006] Ainsworth, S.: Deft: A conceptual framework for considering learning with multiple representations; *Learning and Instruction* (2006), 16(3):183 – 198.

[Alpert and Bitzer 1970] Alpert, D. and Bitzer, D. L.: Advances in computer-based education; *Science* (1970), 167(3925):1582–1590.

[Brown et al. 2014] Brown, P., Roediger, H. and McDaniel, M.: *Make It Stick*; The Belknap Press (2014)

[Carroll 1982] Carroll, J.M.: The adventure of getting to know a computer; *Computer* (1982), 15(11):49–58.

[de Jong et al. 2013] de Jong, T., Linn, M. C. and Zacharia, Z. C.: Physical and virtual laboratories in science and engineering education; *Science* (2013), 340(6130):305–308.

[Dede 1995] Dede, C.: The evolution of constructivist learning environments: Immersion in distributed, virtual worlds; *Educational technology* (1995), 35(5):46–52.

[Dede 2009] Dede, C.: Immersive interfaces for engagement and learning; *Science* (2009), 323(5910):66–69.

[Dede et al. 1996] Dede, C. Salzman, M. C. and Loftin, R. B.: Sciencespace: virtual realities for learning complex and abstract scientific concepts. In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium* (1996), 246–252, 271.

[Goode et al. 2008] Goode, M. K., Geraci, L. and Roediger, H. L.: Superiority of variable to repeated practice in transfer on anagram solution; *Psychonomic Bulletin & Review* (2008), 15(3):662–666.

[Greenwald et al. 2017] Greenwald, S. W., Wang, Z., Funk, M. and Maes, P.: Investigating social presence and communication with embodied avatars in room-scale virtual reality; In *International Conference on Immersive Learning*, Springer (2017), 75–90.

[Hart 2006] Hart, S. G.: Nasa-task load index (nasa-tlx); 20 years later; In *Proceedings of the human factors and ergonomics society annual meeting*, Sage Publications Sage CA: Los Angeles, CA. (2006), 50:904–908.

[Hmelo-Silver et al. 2007] Hmelo-Silver, C. E., Duncan, R. G. and Chinn, C. A.: Scaffolding and achievement in problem-based and inquiry learning: A response to kirschner, sweller, and. *Educational Psychologist* (2007), 42(2):99–107.

[Hubbard et al. 2017] Hubbard, R., Sipolins, A. and Zhou, L.: Enhancing learning through virtual reality and neurofeedback: A first step; In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, New York, NY, USA. ACM (2017), 398–403.

[Karutz and Bailenson 2015] Karutz, C. O. and Bailenson, J. N.: *Immersive Virtual Environments and the Classrooms of Tomorrow*; John Wiley & Sons, Ltd. (2015), 290–310.

[Kay 1972] Kay, A. C.: A personal computer for children of all ages; In *Proceedings of the ACM Annual Conference - Volume 1*, ACM '72, New York, NY, USA (1972)

[Kerr and Booth 1978] Kerr, R. and Booth, B.: Specific and varied practice of motor skill; *Perceptual and Motor Skills* (1978), 46(2):395–401.

[Kirschner et al. 2006] Kirschner, P. A., Sweller, J. and Clark, R. E.: Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching; *Educational Psychologist* (2006), 41(2):75–86.

[Malone 1982] Malone, T. W.: Heuristics for designing enjoyable user interfaces: Lessons from computer games; In *Proceedings of the 1982 conference on Human factors in computing systems - CHI '82*, ACM Press, New York, New York, USA (1982), 63–68

[Mavrikis et al. 2013] Mavrikis, M., Gutierrez-Santos, S., Geraniou, E. and Noss, R.: Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments; *Pers Ubiquitous Comput* (2013), 17(8):1605–1620.

[McCaffery et al. 2014] McCaffery, J., Miller, A., Oliver, I. and Allison, C.: Augmented learning roads for internet routing; In *Frontiers in Education Conference (FIE), 2014 IEEE* (2014), 1–9.

[Njoo and De Jong 1993] Njoo, M. and De Jong, T.: Exploratory learning with a computer simulation for control theory: Learning processes and instructional support; *J Res Sci Teach* (1993), 30(8):821–844.

[Pantelidis 2010] Pantelidis, V. S.: Reasons to use virtual reality in education and training courses and a model to determine when to use virtual reality; *Themes in Science and Technology Education* (2010)

[Purcell 2013] Purcell, E.: *Electricity and Magnetism*; Electricity and Magnetism. Cambridge University Press (2013)

[Rau et al. 2015] Rau, M. A., Michaelis, J. E. and Fay, N.: Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry; *Computers & Education* (2015), 82:460 – 485.

[Regian 1997] Regian, J. W.: *Virtual Reality for Training: Evaluating Transfer*, Springer US, Boston, MA (1997), 31–40.

[Regian et al. 1992] Regian, J. W. Shebilske, W. L. and Monk, J. M.: Virtual reality: An instructional medium for visual-spatial tasks; *Journal of Communication* (1992), 42(4):136–149.

[Regian and Yadrick 1994] Regian, J. W. and Yadrick, R. M.: Assessment of configurational knowledge of naturally- and artificially-acquired large-scale space; *Journal of Environmental Psychology* (1994), 14(3):211 – 223.

[Rieman 1996] Rieman, J.: A field study of exploratory learning strategies; *ACM Trans. Comput.-Hum. Interact.* (1996), 3(3):189–218.

[Roussou et al. 2006] Roussou, M. Oliver, M. and Slater, M.: The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning; *Virtual Reality* (2006), 10(3):227–240.

[Saldana 2015] Saldana, J.: *The Coding Manual for Qualitative Researchers*; SAGE Publications (2015)

[Salzman et al. 1999] Salzman, M.C., Dede, C., Loftin, R.B. and Chen, J.: A Model for Understanding How Virtual Reality Aids Complex Conceptual Learning; Presence: Teleoperators and Virtual Environments (1999), 8(3):293-316.

[Schuchardt and Bowman 2007] Schuchardt, P. and Bowman, D. A.: The benefits of immersion for spatial understanding of complex underground cave systems; In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*, VRST '07, New York, NY, USA. ACM (2007), 121–124.

[Seidel and Chatelier 2013] Seidel, R. and Chatelier, P.: *Virtual Reality, Training's Future?: Perspectives on Virtual Reality and Related Emerging Technologies*. Defense Research Series. Springer US (2013)

[Shneiderman 1983] Shneiderman, B.: Direct manipulation: A step beyond programming languages; *Computer* (1983), 16(8):57–69.

[Smith and Sherwood 1976] Smith, S. and Sherwood, B.: Educational uses of the plato computer system; *Science* (1976), 192(4237):344–352.

[Thomas and Mead 2008] Thomas, L. D. and Mead, P.: Work in progress-implementation of second life in electromagnetic theory course; In *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual*, (2008). F3D–3.