ORIGINAL RESEARCH ARTICLE

Learning in virtual reality: Effects on performance, emotion and engagement

Devon Allcoat* and Adrian von Mühlenen

Department of Psychology, University of Warwick, Coventry, UK

(Received 12 June 2018; final version received 23 October 2018)

Recent advances in virtual reality (VR) technology allow for potential learning and education applications. For this study, 99 participants were assigned to one of three learning conditions: traditional (textbook style), VR and video (a passive control). The learning materials used the same text and 3D model for all conditions. Each participant was given a knowledge test before and after learning. Participants in the traditional and VR conditions had improved overall performance (i.e. learning, including knowledge acquisition and understanding) compared to those in the video condition. Participants in the VR condition also showed better performance for ‘remembering’ than those in the traditional and the video conditions. Emotion self-ratings before and after the learning phase showed an increase in positive emotions and a decrease in negative emotions for the VR condition. Conversely there was a decrease in positive emotions in both the traditional and video conditions. The Web-based learning tools evaluation scale also found that participants in the VR condition reported higher engagement than those in the other conditions. Overall, VR displayed an improved learning experience when compared to traditional and video learning methods.

Keywords: VR; education; experience; mood

Introduction

Interactive technology is progressing at an incredibly fast rate, and advances in virtual reality (VR) technology have led to many potential new applications. Commercial VR headsets are widely used for entertainment purposes, with many individuals’ experiences of VR being from video games and other widely distributed media, as these media are widely advertised and well known, leading to higher popularity. However, VR has broader application possibilities, thanks to significant advances in the technology, including the technology now available in a mobile format.

VR technologies allow the user to see and interact with virtual environments and objects. Modern VR is delivered through a headset, which allows the user to see – and in some cases, hear – the 3D environment. In this way the user is totally immersed in the virtual environment, as it replaces the physical environment around them. Immersion and engagement can be considered intrinsically linked in virtual environments.
(McMahan 2003). Mount et al. (2009) discussed the relationship between immersion, presence and engagement. They explored what it means for a learner to be immersed and considered immersion and engagement in 3D virtual environments, to outline how 3D virtual environments can be used to enhance learner engagement.

VR boasts a number of features that could be useful for education: it presents environments in 3D, it is interactive and it is able to give audio, visual and even haptic feedback. Presenting learning materials in 3D can be especially beneficial for teaching subjects where it is important to visualise the learning materials (e.g. in chemistry or in engineering). Though visualising is one of the most obvious benefits of VR, this could also be accomplished with simple video. However, videos are passive learning objects, whereas VR allows for a direct interaction with the environment. Interactivity and feedback can be valuable for all subjects, as there are specific benefits of interactive learning because it promotes active learning instead of passive learning.

The usefulness of VR in education might also depend on the type of learning. Learning styles theories suggest that there are various ways to learn, and some individuals learn better with some methods than others, as they have different approaches to information processing. The well-known visual–auditory–kinaesthetic learning styles model (Barbe, Swassing, and Milone 1979) suggests there are three types of learning styles: visual, auditory and kinaesthetic. VR allows all three of these learning styles to be targeted in one application, as VR headsets allow for complex visual renderings, audio and movement tracking. Though there has been much contention over learning styles theories, as discussed in the following, having one learning environment that can encompass multiple learning styles could be very beneficial as it would be suitable for a much wider range of individuals.

Other learning styles models include the importance of learning through different perceptual modalities, many of which are able to be targeted in VR (for an overview of various learning styles theories, see Cassidy 2004). It has been suggested that having a variety of learning methods is valuable; Gaytan and McEwen (2007) concluded that it is beneficial to use a variety of instructional methods to appeal to students’ learning preferences. VR activities could be designed to include multiple learning methods, so learners can choose to engage with the learning materials in the manner that interests them the most, as students have a preference for multiple modes of information presentation (Lujan and DiCarlo 2006).

Scholars now are more critical of learning styles theories (e.g. Pashler et al. 2008; Riener and Willingham 2010), stating that, though there are many theories, there is little empirical evidence for learning styles. However, others still consider it important to be aware of varying sensory modalities and learning approaches because of students’ differing learning habits and preferences (e.g. Hawk and Shah 2007; Kharb et al. 2013). The impact of learning styles on e-learning is also debated (Truong 2016), including how best to design adaptive virtual learning environments whilst considering learning styles (Kanninen 2008). There are potential benefits of targeting multiple methods of learning within VR, to allow for different information processing. This could be not only a result of learning methods and individuals’ preferences but also an illustration of how different types of information may be better presented in some formats than others (e.g. language may be best learnt with audio, whereas engineering may be better suited to visualisation).

VR is not necessarily equally suitable for all subject areas; benefits of visualising are more significant in some subjects than others. As such, VR applications may be more suited to some areas of education than others. The revised Bloom’s
taxonomy (Bloom et al. 1956; see also Anderson et al. 2001) suggests that there is not simply one way in which information is processed and learnt; instead it presents learning as a hierarchy of learning, consisting of six stages that involve cognitive processes from simplest to most complex (from remember, understand, apply, analyse and evaluate to create). It is suggested that these different types of learning can be processed differently; some methods of study that are used in education are only applicable to some subjects. Debates, for example, are often good at engaging students with material that requires critical thinking (Camp and Schnader 2010; Scott 2008) but are less suited to learning more concrete information, such as for sciences like physics or chemistry. VR, for example, may be less beneficial for learning to play a musical instrument that requires tactile feedback, such as a guitar, but may be particularly useful for topics where spatial arrangement is important or there are dynamic changes.

Though not many empirical studies have yet been conducted, VR has been compared to traditional learning in some areas. In one study a group of military students were taught with either the lecture-based teaching methods that are traditionally used for the subject material (corrosion prevention and control) or with an immersive VR-based teaching method (Webster 2015). They found that whereas the traditional learning group had an improvement of 11%, the VR group had a higher improvement of 26%.

Bellamy and Warren (2011) conducted a case study using simple online interactive simulations which mimicked real experiments. Eighty-three per cent of their students reported that they found these online simulations helpful or very helpful, and their demonstrators stated that the students seemed much better prepared and more willing to answer questions when they had done the online simulations. These and other examples promote for learning the usefulness of simulated environments as alternatives to real-life scenarios.

Creating educational applications for VR could be a laborious and costly endeavour, so it is important to investigate whether these applications are useful for learning or not. Therefore, explorative research can help answer whether the development of educational applications for this type of hardware is worth pursuing. As VR technology has only recently become more accessible and affordable, research in the past using VR in educational and pedagogic settings has typically used smaller sample sizes with less rigorous methodologies. This study looks to address that, considering not only test performance (used as a measure of learning) but also other outcomes of using VR for learning, such as effects on emotion and engagement.

Method
Participants

All participants were first-year Psychology students at the University of Warwick (UK), who completed the study for course credit. A total of 99 participants (84 females, 15 males) who were 19 years of age on average were assigned randomly to one of three learning conditions: traditional (textbook style), VR and video. All participants reported normal or corrected-to-normal vision. The study was approved by the university’s Humanities and Social Sciences Research Ethics Committee, and all participants gave informed written consent and were aware of their right to withdraw at any time.
Apparatus
The questionnaires and learning materials were presented on a 19" LCD computer screen (1920 × 1080 pixels, 60 Hz) using Microsoft Word and Qualtrics. Responses were collected through mouse and keyboard. A HTC Vive (Xindian, New Taipei, Taiwan) (Figure 1) was used for the VR condition. The headset weighs 550 g and displays a 3D environment via two OLED displays (1080 × 1200 pixels per eye, 90 Hz) with a field of view of 100 × 110 degrees. Participants controlled the VR environment with the standard handheld HTC Vive controller.

Learning materials
The learning materials used the same text and 3D model of a plant cell for all three conditions. The VR condition presented the model from the application ‘Lifeliqe Museum’ on the HTC Vive headset, allowing the participants to see and interact with a 3D model, with accompanying descriptive text (Figure 1). The 3D plant cell model was fully interactive, allowing participants to highlight individual cell parts, change the size of the cell and rotate it. They could also teleport around the virtual room, with the plant cell appearing as a floating object in the room with them, which they could navigate around. A menu was available, virtually attached to one of the controllers, showing names of each part of the plant cell. Participants could select one of these parts from the menu (e.g. the Golgi apparatus) and it would highlight the part on the 3D model. This could also be done the opposite way, by selecting the part on the model, which would highlight the name on the menu. A written explanation of the purpose of each part of the plant cell was also available on this menu. The option of a narrator was disabled for this study, in order to remove audio learning as a confounding variable.

Figure 1. The HTC Vive headset and examples of the 3D model used as learning material for all conditions from the Lifeliqe Museum virtual reality environment.
The video condition used a 2D recording from the HTC Vive, matched from participants in the VR condition and presented on a computer screen. Participants were informed that they could navigate this video at will (play or pause, fast forward, rewind), as they would in a distance learning scenario. This was a control to the VR condition as it presented the same visual information, with the same graphics, but did not have other VR features, such as interactivity and immersive 3D display. As such, this condition acted as a carefully matched control condition.

The textbook condition used screenshots of the 3D model with the same accompanying text and presented them on a computer screen as a PDF file (Figure 2). This ensured that all three groups had the same information and visuals to learn with, with the only difference being the format in which these materials were presented.

Rating scales
An adapted version of the Differential Emotions Scale (DES, Izard et al. 1974), with nine emotion categories (interest, amusement, sadness, anger, fear, anxiety, contempt, surprise and elatedness), was used to measure participants’ mood before and after the learning phase. Participants were asked to rate to which extent the emotional adjectives, each represented with three words (e.g. surprised, amazed, astonished), applied to them on a scale from 1 (not at all) to 5 (very strongly). Five of the categories related to negative emotions, and four related to positive emotions.

The Web-based Learning Tools (WBLT) Evaluation Scale questionnaire (Kay 2011) was used to measure engagement. The WBLT Evaluation Scale asks participants to rate what they thought about the learning tools across 13 questions on a scale of 1 (strongly disagree) to 5 (strongly agree). The questions included items such as ‘the learning object helped teach me a new concept’ and ‘I would like to use the learning object again’. The questions can be grouped into the three categories ‘learning’, ‘design’ and ‘engagement’.

Procedure
The procedure was the same for each participant, starting with a pretest and the DES, followed by the learning phase. For the learning phase participants were instructed to learn as much as they could from the learning materials, and all conditions were given the same amount of time (7 min). After the learning phase, participants completed a post-test consisting of the same questions as the pretest, the DES, the WBLT and one
question that allowed for qualitative feedback. The improvement from pretest to post-test was used as the main measure of learning performance. This method was used in order to account for any participants with prior knowledge of the subject (plant cells). Questions used for the test were either sourced directly from a British AQA Biology A-Level exam or were in the same style as these questions.

Results

Learning

The 17 biology knowledge questions were marked as correct or incorrect and used in the calculation of an overall percentage correct, separately for each participant. The top half of Table 1 shows the average knowledge scores in the pretest and in the post-test, together with the difference scores, as an indicator for learning. Here the overall difference between pretest and post-test is referred to as ‘performance’ to differentiate it from the ‘learning’ scores of the WBLT Evaluation Scale. The corresponding average confidence ratings are given in the bottom half of the table.

The knowledge scores were analysed with a mixed-design ANOVA with the between-subject factor condition (textbook, video, virtual) and the within-subject factor test (pre-, post). The ANOVA revealed a significant main effect for test, $F(1,96) = 273.25, p < 0.001, \eta^2_p = 0.740$, indicating that knowledge improved overall by 23.2% from pretest to post-test, and a significant test × condition interaction, $F(2,96) = 6.80, p = 0.002, \eta^2_p = 0.124$. The significant interaction was further analysed with two split-up ANOVAs, separately for pretest and for post-test. The ANOVA on the post-test data revealed a significant condition effect, $F(2,96) = 3.51, p = 0.034, \eta^2_p = 0.068$. Post-hoc least significant difference (LSD) showed that participants in the VR condition scored significantly higher than participants in the video condition (56.5% vs. 43.9%, respectively; $p = 0.009$). The pretest ANOVA showed no significant effect ($p = 0.793$).

The confidence ratings showed a similar pattern of results as the knowledge data (see bottom half of Table 1). The equivalent mixed-design ANOVA revealed a significant effect for test, $F(1,96) = 266.96, p < 0.001, \eta^2_p = 0.736$, as a result of participants being more confident in the post-test than in the pretest (3.24 vs. 2.24, respectively), as well as a significant test × condition interaction, $F(2,96) = 5.80, p = 0.004, \eta^2_p = 0.108$, because of less confidence gain in the video than in the VR or textbook condition (0.71 vs. 1.12 and 1.18, respectively).

Table 1. Number of participants (N), knowledge scores (percentage correct) and confidence ratings (1–5) in the pretest and post-test separately for the three conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Pretest</th>
<th>Post-test</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual</td>
<td>34</td>
<td>28.1%</td>
<td>56.5%</td>
<td>28.5%</td>
</tr>
<tr>
<td>Video</td>
<td>34</td>
<td>27.9%</td>
<td>43.9%</td>
<td>16.1%</td>
</tr>
<tr>
<td>Textbook</td>
<td>31</td>
<td>25.3%</td>
<td>50.2%</td>
<td>24.9%</td>
</tr>
<tr>
<td>Confidence ratings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual</td>
<td>34</td>
<td>2.24</td>
<td>3.35</td>
<td>1.12</td>
</tr>
<tr>
<td>Video</td>
<td>34</td>
<td>2.33</td>
<td>3.04</td>
<td>0.71</td>
</tr>
<tr>
<td>Textbook</td>
<td>31</td>
<td>2.14</td>
<td>3.32</td>
<td>1.18</td>
</tr>
</tbody>
</table>
The knowledge questionnaire data was further analysed by splitting the questions into two categories on the basis of Bloom’s taxonomy (Bloom et al. 1956). The first group (12 questions) related to the remembering of information, whereas the second group (5 questions) was more concerned with the understanding of information. The overall percentage correct in each category is shown in Figure 3. A 3 × 2-way ANOVA on the remembering scores showed a significant test × condition interaction, $F(2,96) = 6.28, p = 0.003, \eta^2_p = 0.116$. Further split-up ANOVAs and LSD tests revealed that in the post-test participants scored significantly higher in the VR than in the video and the textbook condition (53.1% vs. 40.6% and 43.6; $p = 0.008$ and $p = 0.041$, respectively). The corresponding analysis of the understanding scores also revealed a significant interaction, $F(2,96) = 3.15, p = 0.047, \eta^2_p = 0.062$; however, further tests showed no difference between VR and textbook, but scores in the video condition were lower than scores in the VR and textbook conditions (50.2% vs. 60.2% and 62.3%; $p = 0.071$ and $p = 0.79$, respectively). In summary, participants in the VR group showed better remembering than participants in the textbook group, but there was no difference between the two groups in terms of understanding.

**Emotional response**

DES ratings were split into the two categories: positive emotions (interest, amusement, surprise and elatedness) and negative emotions (sadness, anger, fear, anxiety and disgust), and average ratings are shown in Figure 4. A 3 × 2-way ANOVA with the factors condition and test on the positive emotions revealed a significant main effect of condition, $F(2,96) = 13.24, p < 0.001, \eta^2_p = 0.216$, and a significant interaction effect, $F(2,96) = 31.40, p < 0.001, \eta^2_p = 0.395$. The significant interaction was further analysed with three split-up $t$-tests, to see whether ratings changed from pre- to post-test. Positive emotion significantly increased from 3.2 to 3.8 in the VR condition, $t(30) = 4.73, p < 0.001$, and significantly decreased in the video condition, $t(33) = 4.92, p < 0.001$, and in the textbook condition, $t(30) = 4.37, p < 0.001$. The corresponding ANOVA on the negative emotions also revealed a significant interaction effect, $F(2,96) = 4.37, p = 0.015, \eta^2_p = 0.084$, which was a result of a significant decrease in negative emotion.
(from 1.7 to 1.3) in the VR condition, $t(30) = 4.20, p < 0.001$, and no change in the video or textbook condition (both $p$'s > 0.50).

**Learning experience**

Average WBLT ratings were grouped into the three categories ‘learning’, ‘design’ and ‘engagement’ and calculated separately for each category (see Figure 5). Three separate one-way ANOVAs revealed a significant effect of condition for each of the three subscales (all $p < 0.001$). Post-hoc LSD tests showed that both learning and engagement ratings were significantly higher in the VR than in the textbook condition ($p = 0.005$ and $p < 0.001$, respectively), and they were significantly higher in...
the textbook than in the video condition ($p < 0.001$ and $p = 0.016$, respectively). For design, ratings were significantly higher in the VR and textbook conditions than in the video condition (both $p < 0.001$), but there was no difference between the VR and the textbook condition.

**Qualitative feedback**

Qualitative data was also gathered; participants were asked as part of their online questionnaire: ‘What did you think of the format of the learning materials/the equipment used?’ The question was optional, and about half of the participants ($n = 52$) gave some written feedback. Each participant who responded with qualitative feedback was grouped into positive, negative and mixed feedback, and the overall counts for each category and condition are given in Table 2.

Multiple participants reported that the video learning material was ‘confusing’, with one participant stating that it was ‘engaging but confusing’ and another saying it was ‘difficult to navigate’. Participants described the textbook-style learning materials as ‘basic’, ‘boring’ and ‘bland’. There were discrepancies in reports, with some participants stating the materials were ‘clear’ and ‘easy to learn from’ but others expressing that the materials were ‘unclear’ and the diagrams ‘weren’t very helpful’. On the other hand, participants found that the VR was ‘difficult’ to use, often clarifying ‘at first’, but found it more ‘engaging’, with one participant stating that it ‘made learning more exciting’ and another stating that it was ‘very useful and immersive’.

The qualitative feedback suggests that because of the difficulty using the equipment ‘at first’, future studies may benefit from giving VR participants a trial period with the equipment first, to familiarise themselves with controls. Similarly, video recording of VR would not be suitable as a primary learning condition (as opposed to a control condition as in this study) because of the jarring and confusing nature of the movement and interaction when they are not controlling it, with one participant stating that the ‘video felt all over the place’.

**Discussion**

The aim of this study was to consider the effects of using VR headsets for learning. Overall, participants in both the VR and the textbook-style conditions showed better learning than participants in the video condition. Further breakdown of the learning data showed that participants in the VR condition were better at ‘remembering’ than those in the video and traditional conditions, and participants in both VR and traditional conditions were better at ‘understanding’ than those in the video condition.

**Table 2.** Number of participants who responded with qualitative feedback in grouped types: positive, negative and mixed feedback.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Positive</th>
<th>Negative</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Video</td>
<td>2</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Textbook</td>
<td>1</td>
<td>15</td>
<td>6</td>
</tr>
</tbody>
</table>
That the VR condition showed better test results compared to the video condition suggests that the learning in the VR condition is not a result of the graphics or visuals of the equipment, as these were the same in both conditions. Instead, the learning appears to be attributable to either the 3D immersion or the interactivity of the VR environment. A further study may benefit from comparing VR to other active learning methods. This study compares interactive VR, an active learning method, to passive video watching and traditional textbook-based methods. The distinction between active learning and passive learning plays an important role in many existing educational theories. There is evidence that active learning is beneficial to students (e.g. Pereira-Santos, Prudêncio and Carvalho 2017), which could suggest that the benefits found for VR are simply the benefits of active learning. However, active learning is not always found to be better than passive learning (e.g. Haidet et al. 2004); therefore, the benefits shown in VR may also be a result of other factors.

The current results show a difference in learning stages as defined by Bloom’s taxonomy; further research into the other stages would be of interest. This study looked at the lower ends of the learning hierarchy, remembering and understanding. VR may compare differently to traditional methods for applying, analysing, evaluating and creating. In particular, the 3D aspects of VR, along with the interactivity it affords, may be beneficial for ‘creating’ in many subjects. Alternatively, participants’ unfamiliarity with the equipment, which they hadn’t used before, may mean that improvements of the VR condition were diminished, as individuals need time to adapt to new technology systems (e.g. Cook and Woods 1996). This could explain why participants in the VR condition were not significantly better at ‘understanding’ compared to participants in the traditional condition.

VR was also found to have a very positive impact on mood, with participants having an overall increase in positive emotions and an overall decrease in negative emotions. Conversely the other conditions showed a decrease in positive emotions. Enjoyment has been previously linked as an important part of student performance (e.g. Goetz et al. 2006; Valiente, Swanson, and Eisenberg 2012). This suggests that using VR headsets can have a positive impact on the learning experience.

The WBLT Evaluation Scale also shows that engagement can be increased through the use of VR. The importance of student engagement has been recognised previously (e.g. Kuh 2009; Strydom, Mentz and Kuh 2010; Wolf-Wendel, Ward and Kinzie 2009). Participants also rated the VR environment higher for learning, demonstrating that they felt that they had learnt better from the VR. Student self-rating of learning has been shown to be a valid measure of student performance (Benton, Duchon, and Pallett 2013), with participants here reporting higher learning in the VR condition, which was found for ‘remembering’.

The positive effects on emotion and engagement in VR are important benefits for both within and outside classroom learning (e.g. distance learning, self-teaching). These aspects of learning are sometimes overlooked, with the focus being on other outcomes, such as test scores. However, it has been demonstrated that individuals’ emotions, engagement and motivation are highly linked with each other and they are all important aspects of learning (Pintrich 2003).

This research has demonstrated how VR can replicate or complement traditional learning methods. It is important to consider how VR technology allows for learning beyond the classroom. The technology, though suitable for classroom use, is also particularly suitable for distance learning, self-teaching and other learning environments.
This can be achieved, as the equipment can allow for rich, detailed learning environments that can be programmed to any scenario. Such VR environments can allow for learning that could not be replicated in reality (e.g. dangerous environments or experiments) or would be too costly to be accessible (e.g. expensive equipment or materials).

Future studies, for example, may want to consider the possible advantages of the auditory options available with equipment such as VR, which were not utilised for this project as they may be a confounding variable. As discussed, this could be of interest in relation to learning styles, which are prevalent in a number of learning theories, though the concept of learning styles has received some criticism (e.g. Pashler et al. 2008). Regardless of learning styles, there may be some benefit to including audio to increase immersion and engagement (e.g. Paterson and Conway 2014; Wharton and Collins 2011).

Many VR headsets also share the benefits of mobile learning, most obviously the VR headsets that run through mobile phones. Though not as powerful and capable of detailed environments as PC-based VR headsets like the HTC Vive and Oculus Rift, these mobile headsets share many of the same benefits. Though the headset used in the study, the HTC Vive, is currently only mobile with the use of a portable backpack PC, there is a new mobile, portable version of the HTC Vive headset called the ‘Vive Focus’. The Vive Focus is currently available for developers and is expected to be released later this year, which means that applications such as the one used in this study will be fully mobile, allowing for more flexible learning.

Overall, VR does seem to be a potential alternative to traditional textbook-style learning, with similar performance levels and improved mood and engagement. These benefits may have a longer-term impact on learning, such as improvements resulting from the learning experience. However, the results may be partially because of the novelty of the VR equipment, so the improvements may not be sustained over longitudinal studies. Conversely, these improvements could increase over time, as individuals become more familiar with the equipment and more able to navigate it easily. Therefore, further longitudinal studies are needed to address these questions. VR does show great potential, not only as an option to supplement or replace traditional learning methods, but to develop novel learning experiences that have not been used before.

References


