MASTER

Learning in virtual reality
a quantitative study on the effects of immersion and interactivity on procedural training

Augustijn, G.J.

Award date:
2018

Link to publication

by Yannick Augustijn

identity number 0856560

in partial fulfilment of the requirements for the degree of

Master of Science
in Human Technology Interaction

Supervisors:
dr. ir. A. Haans
dr. A.A.J van den Beemt
Abstract

This thesis is based on a graduation internship done at Ordina in Nieuwegein, and studies the effects of immersion and interactivity on procedural training. This was investigated by dividing 44 male and 20 female participants ($M_{age} = 37.9$, $SD_{age} = 12.4$, Range = 21 to 62) over a 2 x 2 design, with interactivity and immersion as between subject variables. A quantitative analysis showed that participants that trained in the VR condition (high immersion, high interactivity), significantly outperformed the other conditions on the practical test, but did not differ significantly on the theoretical test. The higher performance on the practical test can be attributed to the small to medium positive effects of Interest, Involvement, Spatial Presence, and Realism caused by higher Immersion and Interactivity. The negative differences that were found on the theoretical test can be explained by the small detrimental effect of Spatial Presence, due to higher Interaction and Immersion. Future research is encouraged to replicate this study in a real-life setting, to see whether the knowledge gained while training with the VR application transfers to performance on the real-life task, and whether these found effects still hold.

Keywords: procedural training, virtual reality, immersion, interactivity, presence, involvement, realism, interest
It was in January 1896 that the Lumière brothers showed their movie, L’Arrivée d’un train en gare de la Ciotat (Arrival of a train at the station of La Ciotat), in a cinema. It was rumoured to be such an immersive experience for the public, that they flinched and tried to escape from the train that was obviously going to hit them (Loiperdinger & Elzer, 2004). Now, over one hundred years later, there is another type of media that lets users flinch as if something is coming at them.

This type of media is called Virtual Reality (VR) or a Virtual Environment (VE), and can be described by the definition of Merriam Webster (2016), which states that VR is:

“an artificial environment which is experienced through sensory stimuli (such as sights and sounds) provided by a computer and in which one's actions partially determine what happens in the environment”

It is important to note that this is just one of the definitions of VR and several others exist, but that this will be the definition used in the remainder of this thesis. An example that provides insight into this definition would be of a driving simulator where users can turn the car by using the steering wheel.

Ever since its first use, VR has found many different applications; which span many different fields. In entertainment its uses range from videogames (Skalski, Tamborini, Shelton, Buncher, & Lindmark, 2011; Zyda, 2005) to virtual mazes (Shirai, Kobayashi, Kawakita, & Hasegawa, 2004) and even to “adult entertainment” (Zhang, Weidner, & Broll, 2017). But besides having a value for entertainment, VR has also been used as a practical tool; for example in visualizing data (Arns, Cook, & Cruz-Neira, 1999) or for planning new pipelines for oil fields (Gruchalla, 2004). Other applications include the domain of psychology, for example to treat phobias like fear of heights (Hodges et al., 1995), fear of public speaking (Slater, Pertaub, Barker, & Clark, 2006), fear of spiders (Garcia-Palacios,
EFFECTS OF IMMERSION AND INTERACTIVITY

Hoffman, Carlin, Furness, & Botella, 2002), and several other fears (Krijn, Emmelkamp, Olafsson, & Biemond, 2004). Some other applications include the use of VR to “visit” different places, so that a tourist can visit threatened sites (Guttentag, 2010) or that a prospective home-buyer can visit many potential homes in a VR setting, and choose their top five houses to visit in person (Kun & Zong, 2009). These examples show that the applications of VR are plenty, and that this technology can be used in many different instances.

One other instance where VR can be used, is education. Over the years VR has been used for training in aviation (Gopher, Weil, & Bareket, 1994), medicine (Grunwald & Corsbie-Massay, 2006; Seymour et al., 2002; Stansfield, Shawver, Sobel, Prasad, & Tapia, 2000), mechanical engineering (Ren et al., 2015), the navy (Johnson, 1997; Ordina, 2017), the army (Hill et al., 2003), or even Tai-Chi (Chua et al., 2003; Patel, Bailenson, Hack-Jung, Diankov & Bajcsy, 2006). In many of these instances VR has been used to train the skills that are needed in a specific job, such as performing surgery or critical decision making. Furthermore, most of these instances also highlight the kinds of training where VR is most applicable: When the “real” teaching or training is expensive, dangerous, or difficult regarding logistics (Pantelidis, 2010).

This thesis will use the Virtual Reality application developed by Ordina for the royal Dutch Navy to investigate the effectiveness of training in VR, where the research will focus on procedural knowledge. Hereby, procedural knowledge is defined as knowledge applied in the performance of a task, and effectiveness as how much someone learns using a certain learning method. The aim is to increase insight into the effectiveness of VR for training, and to get more insight in specific factors that influence the effectiveness of the VR training. Therefore, the following research question will be investigated:

What is the effectiveness of training in a VR application, with regards to procedural knowledge, and which factors influence this?
The exact aims and hypothesis of the present study will be given after a more detailed literature review on the topics that are of relevance for this research question. These topics include important aspects of VR for educational use, its current uses, and what kind of effects can be expected.

**Virtual Reality**

Two important terms in VR research are *spatial* presence and immersion; where higher presence can cause people to behave as if the simulated content were real, which is useful for training (Slater, Linakis, Usoh, & Kooper, 1996), and where higher immersion has been linked to higher presence (Gorini, Capideville, De Leo, Mantovani, & Riva, 2011) and higher performance (Sowndararajan, Wang, & Bowman, 2008).

**Presence.** One of the definitions of presence, and also the one to be used in this thesis, is the feeling of being in a virtual environment when using a VR system (Slater & Wilbur, 1997). Another definition of presence that is used is the feeling of being in a virtual location with someone else, who is not there in real life. However, this last definition is often also called co-presence or social presence (Heeter, 1992).

According to the current literature there are several different components that give rise to this feeling of presence (Cummings & Bailenson, 2016), such as: the technical capabilities, or immersion, of the medium used (Gorini et al., 2011), the presence of a narrative (also Gorini et al.), differences in spatial ability (Coxon, Kelly, & Page, 2016), a match between what someone sees, does and feels (whether someone walks for real and moves in the VR environment) (Slater, Usoh, & Steed, 1995; Usoh et al., 1999), and emotion in the environment (Riva et al., 2007).

Presence is a psychological feeling, consisting of several different attributes (Biocca & Delaney, 1995). This makes it harder, but not impossible, to quantify in one single score.
Several types of measurements have been proposed to be able to get a “score” for presence. IJsselsteijn, de Ridder, Freeman and Avons (2000) gave a good overview of those different types of measurements; which include subjective and objective measurements.

Examples of subjective, self-reported, measurements include questionnaires, for example the Presence Questionnaire (PQ; Witmer & Singer, 1998) and the Igroup Presence Questionnaire (IPQ; Schubert, Friedmann & Regenbracht, 2001), continuous presence assessment, such as moving a slider in real-time to indicate your sense of presence (IJsselsteijn et al., 1997), and psychophysical methods, such as matching the tone of a sound to the strength of the feeling of presence (Stanney & Kennedy, 1998).

The objective measures that have been used to measure presence, include postural responses, whether someone shows self-movement in order to relate with the experience shown (IJsselsteijn, Freeman, De Ridder, Avons, & Pearson, 2000), physiological measures, including heart rate, respiration rate, skin temperature and skin conductance response (Wiederhold, Davis, & Wiederhold, 1998), dual task measures, like secondary reaction time measures to a tone or a light (Witmer & Singer, 1998), and social responses, for example the reaction to the behaviour of virtual agents (Bailenson, Yee, Merget, & Schroeder, 2006).

Another proposed way of measuring presence is by using a wide variety of the aforementioned measurements on a unidimensional scale using a Rasch model (Haans & IJsselsteijn, 2018). Meaning that any of the aforementioned measures is probably a valid indicator, if it reflects that someone has accepted the virtual environment as real to a certain degree.

These examples show that there are many ways to try to measure the feeling of presence, its effects on the behaviour and experiences of users, and to try to relate this to its
origin. One of the ways in which this feeling might originate, as aforementioned, is through immersion (Cummings & Bailenson, 2016; Gorini et al., 2011).

**Immersion.** An explanation for the concept of immersion is the potential of the technology to induce the feeling of presence (Slater & Wilbur, 1997); meaning that the level of immersion is defined, and can be assessed, by the technical capabilities of the system used for the VR experience. Changing the specifications of the system, can change the level of immersion achieved. This was also described by Bowman (2007) who stated that Field of View (FOV, the angle that can be viewed at once), Field of Regard (FOR, the total size of the visual field), display size, display resolution, stereoscopy (the display of different images to each eye to provide an additional depth cue), head-based rendering (display of images based on the position and orientation of the users head), realism of lightning, frame rate, and refresh rate can all alter the level of immersion of a system. These same observations were also made by Cummings and Bailenson (2016), who furthermore found that tracking level, sound and user perspective were also of importance for immersion. A similar conclusion was also drawn by Steuer (1992), who said that the level of immersion is based on “vividness” and “interactivity”.

**Vividness.** Vividness was described as the way that the mediated environment presents information to the senses of the user (Steuer, 1992). Within this definition two other terms were defined, breadth and depth. Breadth refers to the number of sensory dimensions simultaneously presented; and depth being the resolution within these different channels. An example of this would be that a movie provides a more vivid experience than a book, because it caters to more different sensory dimensions, and because the moving images provide a higher resolution than the printed words.

**Interactivity.** The second term that Steuer (1992) described was interactivity, which is the level at which users can interact with the mediated environment. This interactivity
depends on the speed (the rate at which input can be incorporated in the mediated environment), the range (the number of possible actions), and mapping (the mapping of the controls in a natural and predictable way). Some examples that would make these concepts more clear are that a system with a delay in the visual feedback would score worse on speed, and thus interactivity, than a system without a visual delay (Welch, Blackmon, Liu, Mellers, & Stark, 1996); that a videogame where you can move in 360 degrees provides a better range of actions, and thus higher interactivity, than the same game where you can only move in eight predefined directions; and that a virtual car controlled by a steering wheel would score better on mapping and provide a more immersive experience than controlling the same car with keyboard controls.

These concepts of interactivity and vividness are of importance for achieving a higher level of immersion (Steuer, 1992). This can subsequently be linked to a higher level of presence (Slater & Wilbur, 1997), which can result in several benefits in education.

**Virtual Reality in Education**

**Benefits.** These benefits include from the possibility for learning on demand (LoD) (Trondsen & Vickery, 1997) and active learning (Cuseo, n.d.).

The first benefit of VR in education is the concept of Learning on Demand (LoD). The main idea behind this term is that someone can follow a training whenever or wherever they want to, due to videos, VR lectures or lessons. While using this the user can learn without being dependent on experts or equipment (Ren et al., 2015; Trondsen & Vickery, 1997). An example of this would be learning how to perform a procedure on a navy ship that is often away for a mission. But when using a VR training the user could practice the procedure as often as they want, without having to be on the actual ship.
Another benefit of VR is the support for active learning. Active learning means that students do not only sit and listen to a lecture, but are invited to actively participate in the course. For example by holding debates, teaching each other, and doing experiments (Bonwell & Eison, 1991; Lujan, 2006). Freeman et al. (2014) found in their meta-analysis that student learning increased under active learning, and that it would increase the average grade of a class by approximately six percent.

This rise in performance is probably due to the fact that under active learning students are more likely to be actively involved with the learning materials, compared to when they are just listening (Cuseo, n.d.). This was also indicated by Fletcher (1990), who stated that students tend to retain 20% of what they hear, 40% of what they see, but 75% of what they see, hear, and interact with. Thus, a student reading about a procedure, watching someone do the procedure and performing the procedure themselves would score better on performance compared to a student who would only read about the procedure for the same amount of time.

With VR, you can support active learning, by providing higher interactivity, and in such a way let the students interact with the information or the procedure that they need to learn. Being able to practice the procedure that they need to learn will help the students develop their performance on this procedure. This was also shown in the paper of Ren et al. (2015), who used 3D virtual laboratories for mechanical engineering students to enable them to practice with certain apparatus, instead of using a paper-based manual. During this experiment, they demonstrated that the students that practiced with the 3D virtual labs significantly outperformed the students that used the paper-based manual, when working with the actual equipment.

The final advantage of using VR that will be discussed in this thesis is the fact that VR is a rather new and interesting technology in the educational domain. This was also indicated by Ren et al. (2015), who showed that students in virtual conditions were more interested in
learning than students in paper-based manual conditions. Furthermore, this was also indicated by Pantelidis (2010) who said that VR should be used if it can make the subject to learn more interesting, and the students more motivated. These more interested and motivated students will then perform better (Castillo-Merino & Serradell-López, 2014). Although this might seem as a temporary benefit, due to the newness of the medium, it is still one to consider when investigating the idea of using VR in education.

**Drawbacks.** Although there are many benefits to VR in training and education, there are also some drawbacks of VR that could make the use of this technology less of a suitable choice.

One of these is the fact that some participants might feel nauseous when using a VR system. This feeling is being generated by the fact that the motion that you sense is different from the motion that you visualize. This causes the body to think that it has been poisoned and that it needs to lose the poison, by vomiting (Golding, 2006). This feeling can occur in VR because the match between what someone sees and what someone does in reality is not always correct. This means that users of the system might become nauseated, and suffer from some of the feelings associated with it, like being sick to their stomach or feeling dizzy, light-headed, sweaty, hot, or annoyed (Gianaros, Muth, Mordkoff, Levine, & Stern, 2001). These feelings might distract from the attention given to the task, which will probably result in a worse performance (Comer, Gould, & Furnham, 2013).

The next potential drawback of VR is that the initial cost of creating a VR simulation is rather large, both in time and in money. This was also stated in the paper by Pantelidis (2010), who discouraged using VR when creation of the environment is “too expensive to justify using, considering the expected learning outcome.” (p. 65). This shows that clear thought should be set in the creation of VR, and that the pros and cons should be weighed before implementing VR in the curriculum.
The final problem is not with VR in use of education, but with regards to the use of ICT in education in general. Schneckenberg (2009) has indicated that the overall diffusion of ICT in educational settings is very low. This is partly because of the low levels of competence with ICT of the faculty staff (Allen & Seaman, 2007), the attitudes of the staff (Van den Beemt & Diepstraten, 2016), and the expectations of the students (King & Boyatt, 2015). These papers show that, even though VR might be an appropriate solution there still is a long way to go before it gets implemented in education. This slow diffusion of ICT in education might counter the benefit of VR being a rather new and exciting medium and is therefore important to keep in mind.

The combination of benefits and drawbacks of VR shows that some situations can gain from using VR in education. These added benefits are also why the use of VR in education has been around for over fifty years (McLellan, 1996). Over those years, VR has had several different uses, which highlight where it can be used best.

**Current use.** Most of the instances where VR was used, included training of skills or knowledge that is expensive to train regarding materials, dangerous to train when mistakes are made, or hard to train regarding logistics; but it has also been used in the case of distant learning.

**Expensive to train regarding materials.** The first useful application is in instances where training with the real material is expensive. Good examples of this can be found in medical literature, where training for operations requires bodies to practice on (Grunwald & Corsbie-Massay, 2006). With VR, these bodies can be simulated and therefore be practiced on infinite times, without any added cost. This will subsequently lead to improved performance on these operations (Seymour et al., 2002).
Another example of a good match for VR is in the field of aviation. Simulated flight training has been around since the 1960s, where the Air Force has been first developing flight simulators with head-mounted displays (McLellan, 1996). Flight simulators are an appropriate use for VR, since it is easier, cheaper, and safer to train in a simulator, compared to making a real flight. Furthermore, it has been found that training from a flight simulator transfers to real life performance (Gopher et al., 1994).

These two examples also show instances where training in VR does not suffer from some of the potential drawbacks. This is because the “real” training is also very expensive, meaning VR training does not suffer from the cost drawback, and therefore implementation and use of the VR in training is rather affordable.

**Dangerous to train when mistakes are made.** A second aspect of training that can be done using VR is training which is dangerous when mistakes are made. An example of this is in the aforementioned case of medical training, where small mistakes can result in injuries or death of patients. Another example of training under dangerous circumstances was described by the research of Hill et al. (2003), who described a VR system that teaches army officers, and lets them train with principles of critical decision making. This is also a case which is hard to train regarding logistics, but also one where mistakes can have severe consequences.

Furthermore, this kind of training consists of army officers making the correct decisions under critical circumstances, which is a task that did not require much movement besides walking and watching. Since these aspects of the training can be incorporated in the training, the user will not have a mismatch between what they sense and what they see, and will therefore not suffer from potential nausea.
Hard to train regarding logistics. VR is also an appropriate solution when training with the “real thing” is hard or difficult to do regarding logistics. One case in which this was done was in the study of Ren et al. (2015), who used virtual laboratories to let their students train with certain expensive machines, which were needed to do certain types of measurements, which would otherwise be hard to realize due to a high number of students and a low number of machines.

A second example of training where logistics can be difficult is in the case of presenting in front of an audience, where the most frightening part is to speak in front of a group of people (Slater et al., 2006). However, finding a group of people that want to listen to a presentation for practice can be rather difficult. In this case, using a VR system to practice your presentation, as was done in the experiment of Slater et al., can be a good way to practice this.

A final case in which VR can be used when logistics makes training hard to do is in the case of the Dutch Navy. Students from this organisation often need to train procedures on the ships on which they will work, but most of these ships are often not available since they are on a mission. A proposed solution was a VR application which allowed the students to train the procedures on a virtual ship, meaning that there is less stress on the actual ship to return for training (Ordina, 2017).

These examples show instances where the benefits of VR training outweigh the drawbacks. For example, in the study of Ren et al. the costs of buying extra machines to provide all the students would outweigh the costs of a VR application. Furthermore, students would see what they sensed, as it only incorporated using the machine; so, students would not feel nauseated.
**Distant learning.** VR has also been shown to facilitate learning from a distance, for example in the study of Soblechero, Gayz, and Hernández Ramírez (2014), who compared online learning with onsite learning, and found no significant differences in results.

A second example where a VR solution was used was in the study of Lau (2015), who used 3D immersive VR in a comparison to traditional in-house training, and did not find any significant differences. Based on these results they argued that future employees do not need to travel to an in-house training facility, but can instead train from any location. These two examples show that there is evidence to believe that lectures or training can be followed from a distance, with the use of immersive VR.

However, even though beneficial effects have been found in several cases, there can still be more insights into why these effects exist in the first place, and what specific explanatory factors give rise to the beneficial effects of VR.

**Research aims.** This thesis will focus on why learning using VR applications help students perform better, and why the beneficial effects found in the previous mentioned cases exist. There are several factors that might explain the benefits of a VR application over other ways of learning, such as videos or manuals. These factors are expected to be the higher levels of immersion (Sowndararajan et al., 2008) and the beneficial effects of interactivity (Fletcher, 1990; Jang, Vitale, Jyung, & Black, 2017).

Immersion is expected to have a positive impact on performance of a procedure because of more available spatial cues (Sowndararajan et al., 2008) and because higher immersion has been linked to higher presence (Gorini et al., 2011). This can lead to more realistic behaviour of the users, which is useful when training for a procedure (Slater et al., 1996).
Furthermore, Interactivity is expected to have a positive effect because it has been linked to higher immersion (Steuer, 1992), and because students are proven to retain information better when they get to interact with the information, compared to passive viewing (Fletcher, 1990; Jang et al., 2017).

Other factors that are expected to be of effect, besides interest and (spatial) presence are involvement, which is the degree of how much the materials presented can captivate and maintain the attention of the user, and realism, how real the virtual environment looks (Witmer & Singer, 1998). Involvement was found to have a positive effect on retention after learning a procedure, since more involvement means that users are more engaged with the materials presented (Chittaro & Buttussi, 2015). Realism is expected to be of effect as it will contribute to the sense of presence experienced by the user (Slater et al., 1996).

The aim of this thesis is to find an answer to the research question “What is the effectiveness of training with a VR application, with regards to procedural knowledge and which factors influence this?”, and to answer it, it will be split into the two following sub questions:

What is the effectiveness of training with a VR application, with regards to procedural knowledge?

How do immersion and interactivity influence the effectiveness of the VR application?

This will be examined by teaching participants a procedure on a virtual navy ship, using four different teaching methods, which vary on Immersion and Interactivity. These conditions will be: a physical manual (low Immersion, low Interactivity), an interactive manual (low Immersion, high Interactivity), passive viewing of the procedure in VR (high Immersion, low Interactivity), and practicing the procedure in VR (high Immersion, high
Interactivity). Each participant will train using one of these methods. Their performance will then be assessed by a practical test and a theoretical test.

It is expected that the earlier stated effects of immersion and interactivity will have an effect on performance through interest (because of the novelty of immersive and interactive media) (Castillo-Merino & Serradell-López, 2014; Pantelidis, 2010), spatial presence (Ai-Lim Lee, Wong, & Fung, 2010; Witmer & Singer, 1998), involvement (Chittaro & Buttussi, 2015) and realism (Slater et al., 1996). Because of these expectations the following hypotheses have been chosen:

- **H1**: Participants in the VR condition score significantly higher on performance than the groups in the other conditions.
- **H2**: Immersion has a significant positive effect on performance.
- **H3**: Interactivity has a significant positive effect on performance.
- **H4**: Interactivity and Immersion have a significant positive effect on performance through interest, spatial presence, involvement, and realism as mediators.

**Method**

**Participants and design**

66 participants (46 males and 20 females; \( M_{age} = 38.2, SD_{age} = 12.4, \) Range = 21 to 62) were recruited using convenience sampling. The experiment lasted about thirty minutes for which participants received no compensation. Participants were randomly assigned to a 2 (Interactivity: yes vs no) x 2 (Immersion: high vs. low) design with both Interactivity and Immersion as between subject variables. During data collection it was noted that two participants were behaving differently (one rushed the study and the other later indicated to not have understood the study), these two also proved to be significant outliers during data
analysis and were thus removed from the observations. Hereafter, 64 participants (44 males and 20 females; $M_{age} = 37.9$, $SD_{age} = 12.4$, Range = 21 to 62) remained.

**Apparatus and Materials**

Within this study four distinct kinds of media were used to teach participants how to let down a Rigid Inflatable Boat (RIB) from a navy ship. To complete this task, they needed to visit three separate locations on the ship, turn on three separate systems in the correct order and finally press a button at the correct time. This was taught using one of the following materials, which were all presented in the native language of the participants (Dutch):

- **Manual**: A paper-based manual explaining the various steps that are needed to complete the task. This manual included text and pictures to teach the participants the task.

- **Interactive Manual**: This interactive manual was based on the previous manual. It was made and presented in PowerPoint and provided participants the opportunity to e.g. press the correct buttons to turn on the system, or move to the correct places by pressing the arrows on the screen.

- **Passive VR**: This condition included a “first-person view” of the experimenter doing the procedure in the VR application, while using a text-based manual on a virtual tablet. This video was recorded using “VR Capture” and was presented on a HTC Vive by using “MermaidVR Video Player”.

- **Active VR**: In this condition participants could practice the task by using the VR application. They had access to an in-game text-based manual, presented on a virtual tablet. This application was made by Ordina using Unity, and was presented using a HTC Vive with motion controls.
The HTC Vive that was used consisted of a 2160x1200 AMOLED Head Mounted Display (HMD) (1080x1200 per eye), with a refresh rate of 90 Hz. It also supported two handheld, wireless, motion controllers, who both had 24 sensors, a multi-function trackpad, dual stage triggers and HD haptic feedback. The player was tracked using two HTC Vive Base stations (or “Lighthouses”) which provided a 360-degree virtual space up to a 4,5x4,5m radius.

The computer that was used to run the VR application and support the HTC Vive was a MSI GT73VR 7RF-409NL Titan Pro laptop, containing an Intel core i7 processor, 16 GB RAM and a NVIDIA GeForce GTX 1080 video card. This laptop ran Windows 10 as its operating system.

Procedure

Upon entering the lab, participants were informed that all collected data would be analysed anonymously, and their rights to withdraw at any time were explained. By signing a form, they agreed to participate in the experiment.

Next, participants received the materials to learn the procedure based on their experimental condition. They had ten minutes to learn the procedure, which started as soon as they finished reading the instructions in the “low-Immersion” conditions, or when they read the instructions and were used to the controls in the “high-Immersion” conditions. Based on their condition they received the materials as described in Table 1:
Table 1
Overview of the materials received per experimental condition

<table>
<thead>
<tr>
<th>Low Immersion</th>
<th>No Interactivity</th>
<th>Interactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>Manual a</td>
<td>Interactive Manual b</td>
</tr>
<tr>
<td>High Immersion c</td>
<td>First person View Video</td>
<td>VR Application</td>
</tr>
</tbody>
</table>

a. This was presented on paper
b. This was presented on a laptop
c. The “high Immersion” conditions were presented on a HTC Vive

After studying the materials for ten minutes the participants had to complete a theoretical test, as will be described in the “Measures” section. Subsequently, the participants were presented the questionnaire.

Next, participants were asked to perform the practical test. They were given the HTC Vive and received an explanation of the different controls. After this they got a few minutes to get used to the controls that were needed to complete the practical part of the experiment. When they indicated to be familiar with the controls, the VR application was started and participants could perform the procedure in VR.

After the participants completed the practical test, they were debriefed and thanked for their contribution.

Measures

Theoretical test. The written test that was used was created with the help of the programmers of the VR application, as well as with teachers and professionals who have experience with teaching tasks. The written test consisted of two parts: ten questions about the procedure (e.g. Where is the button located to turn on the Electric Cabinet?”) and of a task analysis part where participants had to put the specific parts of the task in the correct order (e.g. “Go down the stairs” and “Turn on the Electric Cabinet”). The number of correct answers per part and time taken per part to answer the questions was recorded to create a measure of performance on the theoretical test. The logarithms of the time needed to complete each part were transformed to a scale ranging from 0 to 100 resulting in a score for time
needed for the questions \( (M = 47.84, SD = 19.60) \), and for the tasks \( (M = 49.30, SD = 22.58) \). This was done to be able to get more intuitive and understandable values on both these measures.

It was found that these four variables, that were used to measure performance on the theoretical test, correlated; as shown in Table 2. To control for the number of tests ran, a factor analysis without rotation was done on these different variables to see whether they could be combined into a smaller set of measures, for example one regarding time and one regarding questions correct. It was found that there was one resulting factor which had an eigenvalue of 1.25. On this factor all loadings were above 0.45, Bartlett’s test of sphericity was significant \( (\chi^2 (64) = 39.07, p < .00) \) and the Kaiser-Meyer-Olkin measure of sampling adequacy reported 0.59, very close to the commonly recommended 0.60. Based on these results, a reduction of the four measures into one measure was deemed correct, and the four measures were averaged into one score for theoretical test \( (\alpha = 0.65) \), and scaled to range from 0 to 100. This score \( (M = 46.19, SD = 21.42) \) will be used for the remainder of the analyses as the variable to measure performance on the theoretical test.

Table 2

<table>
<thead>
<tr>
<th>Variable:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Time taken (questions)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Time taken (task)</td>
<td>0.41**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Percentage correct (questions)</td>
<td>0.25*</td>
<td>0.23</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4. Percentage correct (task)</td>
<td>0.17</td>
<td>0.47**</td>
<td>0.40**</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. \( n = 65 \)

* \( p < 0.05 \) ** \( p < 0.01 \)

**Questionnaire.** The study also contained a questionnaire with 13 questions from the IPQ on a 7-point Likert scale (Schubert, 2003), six questions of the Intrinsic Motivation Inventory (IMI) regarding interest in the used teaching method, also on a 7-point Likert scale (Intrinsic Motivation Inventory, 1994; Visser-Wijnveen, Stes, & van Petegem, 2011), and
demographic questions regarding the participants age, sex, education, videogame experience and VR experience.

The answers to the six questions from the IMI questionnaire were used to generate a value for interest (6 questions, $\alpha = 0.92$). The answers on the questions from the IPQ to receive values for Spatial Presence (5 questions, $\alpha = 0.83$), Involvement (4 questions, $\alpha = 0.64$), and Realism (4 questions. $\alpha = 0.73$). The questions regarding demographics were used to be able to control for any significant differences over these factors.

**Practical test.** The practical test consisted of the Virtual Reality application, made by Ordina. In this application the participants could perform the learned procedure. The environment consisted of three different locations, that needed to be visited in a specific order and in two of the rooms the participant had to interact with the environment. The three locations in the application were:

- **RIB-Room.** This is where the participants started. From here they could reach the “Engine Room” and “Control Room”. They could also see the RIB from here.

- **Engine Room.** This room could be reached by going down the ladder from the RIB-Room. Here participants had to turn on the “Electric Cabinet” and after this go back to the “RIB-Room”

- **Control Room.** After turning on the “Electric Cabinet” the participants had to go to the “Control Room”. Here they first had to open the door of the ship by using the “Hydraulic System”. When the door was opened they had to let down the RIB by using the “Operator Console”.

Once they let down the RIB, they had completed the procedure. The logarithm of the time needed to complete the procedure was transformed to a scale ranging from 0 to 100 ($M = 55.50, SD = 24.77$), and taken as the variable that measures performance on the
practical test. This variable was used as it was easiest to record, and because mistakes made only added time for completing the procedure, but did not result in other negative consequences on the procedure.

Results

During an initial investigation of the data it was found that of the various descriptive variables age, $r = -0.51$, $n = 64$, $p < 0.01$, and VR experience, $r = 0.33$, $n = 64$, $p < 0.01$ significantly correlated with performance on the practical test. However, there were no significant differences between age and VR experience between the different conditions. Furthermore, no significant correlations were found between the descriptive variables and performance on the theoretical test. All used tests were tested for their relevant assumptions and were tested on a two-sided basis on a five percent alpha level. In case of violations of the assumptions, these will be reported and the correct measures will be taken.

For the first hypothesis H1: *Participants in the VR condition score significantly higher on performance than the groups in the other conditions*, two contrast analyses were run (Haans, 2018). One on performance on the theoretical test and one on performance on the practical test. A graphical overview of the differences in performance between the different conditions can be found in Figure 1.

The contrast analysis for performance on the practical test tested whether the participants in the VR condition scored significantly higher on performance on the practical test, resulting in contrast weights of 1 for the VR condition, and -1/3 for the other three conditions. This test showed a difference of 33.97 points between the scores of the VR group, that trained using the VR application, ($M = 80.61$, $SD = 12.20$) and the scores of the other conditions ($M = 46.42$, $SD = 21.7$). The difference between the VR and non-VR groups proved to be a significant positive difference, of which around 85% of the found difference
between the groups can be explained by the expectation tested, with $F(1,60) = 39.87, p < 0.01, \eta^2_{alerting} = 0.85, C = 33.97$. This result supports hypothesis H1: *Participants in the VR condition score significantly higher on performance than the groups in the other conditions.*

The contrast analysis for performance on the theoretical test indicated that there was a negative difference of 5.5 points between the VR condition ($M = 42.13, SD = 18.92$) and the mean of the other conditions ($M = 47.67, SD = 22.26$). This means that VR did not score significantly higher on the theoretical test than the other conditions, with $F(1,60) = 0.92, p = 0.34$.

These two analyses showed that participants in the VR condition performed significantly better on the practical test, while they failed to outperform on the theoretical test. Therefore, it finds support for hypothesis H1: *Participants in the VR condition score significantly higher on performance than the groups in the other conditions,* in case of the
practical test, but it fails to find support for this hypothesis in case of performance on the theoretical test.

The second hypothesis $H2$: *Immersion has a significant positive effect on performance* was answered using a two-sided $t$-test. It was found that participants in the “high Immersion” conditions ($M = 34.72, SD = 11.06$) were significantly younger than the participants in the other conditions ($M = 41.13, SD = 12.93$), $t = 2.13, p = 0.04$. Age showed to influence performance on the practical test directly, but no interaction effects were found. Because of this, performance on the practical test will also be investigated using a regression analysis to be able to control for this effect of age on performance on the practical test. A graphical overview of the difference in performance on the two tests can be found in Figure 2, before controlling for age, and Figure 3, after controlling for age.

![Graph showing performance on practical and theoretical tests before and after controlling for age](image)

*Figure 2. Performance on the practical and the theoretical test for high vs. low Immersion before controlling for age, error bars represent 95% confidence intervals*
After controlling for age, $b = -0.79$, $t = -4.03$, $p < 0.01$, $\eta^2 = 0.21$, it was found that Immersion still had a significant positive effect on performance on the practical test, $b = 22.7$, $t = 4.75$, $p < 0.01$, $\eta^2 = 0.27$.

The results of another two-sided $t$-test showed that there was no significant difference, $t = 0.60$, $p = 0.55$, between the scores of the low Immersion groups ($M = 47.82$, $SD = 23.61$) and the high Immersion groups ($M = 44.58$, $SD = 19.23$) regarding performance on the theoretical test. This was not controlled for age, as there was no significant effect of age on performance on the theoretical test.

These results support the hypothesis in case of performance on the practical test, where participants in the “high Immersion” group outperform those in the “low Immersion” group, but fail to support the hypothesis in case of the theoretical test, where no significant difference can be found.
Another two-sided $t$-test was used for the third hypothesis $H3$: Interactivity has a significant positive effect on performance. A graphical overview of the differences between the two groups (high vs. low Interactivity) can be found in Figure 4.

The results of the $t$-test on performance on the practical test by Interactivity (high vs. low) showed that there was a significant positive difference of 14.5 points between the scores of the groups in the “high Interactivity“ ($M = 62.53$, $SD = 26.34$) and “low Interactivity” condition ($M = 48.02$, $SD = 20.89$), $t = 2.43$, $p = 0.02$, $d = 0.61$.

![Figure 4, Performance on the practical and the theoretical test for high vs. low Interactivity](image)

The other $t$-test also showed a significant negative difference of -13.2 points, $t = -2.57$, $p = 0.01$, $d = -0.64$, between the two groups regarding performance on the theoretical test. In this case the participants in the “high Interactivity” group ($M = 39.82$, $SD = 20.87$) scored worse than the participants in the “low Interactivity” group ($M = 52.99$, $SD = 20.16$).
These results show that the data supports the third hypothesis, \( H3: \) \textit{Interactivity has a significant positive effect on performance}, in case of performance on the practical test, but rejects this hypothesis with regards to performance on the theoretical test.

The fourth hypothesis, \( H4: \) \textit{Interactivity and Immersion have a significant positive effect on performance through interest, spatial presence, involvement, and realism as mediators}, was evaluated using structural equation modelling. The two variables that measured performance, one on the theoretical test and one on the practical test, were each modelled three times, using different independent variables: one of the models used Immersion, another used Interactivity, and the final model used both Interactivity and Immersion as independent variables. The correlation matrix can be found in Table 3, graphical overviews of the models can be seen in Figure 5 to Figure 10, and an overview of the coefficients can be found in Table 4 to Table 11.

Table 3  
\begin{tabular}{|l|c|c|c|c|c|c|c|c|c|}
\hline
Variable: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
1. Immersion & - & & & & & & & & \\
2. Interactive & 0.03 & - & & & & & & & \\
3. Interest & 0.76** & 0.26* & - & & & & & & \\
4. Spatial & 0.73** & 0.24 & 0.77** & - & & & & & \\
5. Involvement & 0.45** & 0.12 & 0.50** & 0.55** & - & & & & \\
6. Realism & 0.55** & 0.31* & 0.68** & 0.68** & 0.46** & - & & & \\
7. Theory a & -0.08 & -0.31* & -0.11 & -0.24 & -0.03 & -0.13 & - & & \\
8. Practical b & 0.56** & 0.29* & 0.50** & 0.40** & 0.40** & 0.35** & 0.22 & - & \\
9. Age c & -0.26* & -0.12 & -0.33** & -0.26** & -0.14 & -0.18 & -0.08 & -0.51** & - \\
\hline
\end{tabular}

\textit{Note. n = 65}
\begin{itemize}
  \item a. Performance on the theoretical test
  \item b. Performance on the practical test
  \item c. Age was found to be of effect, and was thus included
\end{itemize}
* \( p<0.05 \) ** \( p<0.01 \)

Table 3 shows the positive significant correlations between Immersion and performance on the practical test, \( r = 0.56, n = 64, p < 0.01 \), and between Interaction and performance on the practical test, \( r = 0.29, n = 64, p = 0.02 \). Furthermore, it provides a
significant negative correlation between Interaction and performance on the theoretical test (ptt), $r = -0.31, n = 64, p = 0.01$.

It also presents the significant correlations between Immersion and Interest, $r = 0.76, n = 64, p < 0.01$, Immersion and Spatial Presence, $r = 0.73, n = 64, p < 0.01$, Immersion and Involvement, $r = 0.45, n = 64, p < 0.01$, and Immersion and Realism, $r = 0.56, n = 64, p < 0.01$. Besides this, the table shows that Interactivity has a significant effect on Interest, $r = 0.26, n = 64, p = 0.04$, and on Realism, $r = 0.31, n = 64, p = 0.01$. This same effect seems to exist between Interactivity and Spatial Presence, $r = 0.24, n = 64, p = 0.05$, but this cannot be fully supported by the data. No significant effect of Interactivity can be found on Involvement, $r = 0.12, n = 64, p = 0.33$.

It can also be concluded that there are positive significant correlations between Interest and performance on the practical test, $r = 0.50, n = 64, p < 0.01$, Spatial Presence and performance on the practical test, $r = 0.40, n = 64, p < 0.01$, Involvement and performance on the practical test, $r = 0.40, n = 64, p < 0.01$, and Realism and performance on the practical test, $r = 0.35, n = 64, p = 0.01$. However, for the theoretical test, no significant correlations can be found between the variables and performance on the theoretical test, $r < 0.25, p > 0.05$, see Table 3 for full results.

Finally, it shows that there are moderate to strong positive correlations between the variables of Interest, Spatial Presence, Involvement, and Realism. When checking the VIF values of these variables, they are not deemed as extremely problematic (VIF values all below 3.0). However, during creation of the models problems were encountered that are often attributed to multicollinearity errors, such as a change in the signs as well as in the magnitudes of the coefficients, and difficulties when trying to assess the relative importance of the variables. These problems were tried to be solved by centring all the different variables,
trying to create factors of them, and even dropping some of the variables and investigating what happens, but no improvements were found.

The first model investigated how Immersion affects performance on the practical test, especially through the variables of Interest, Spatial Presence, Involvement, and Realism. This model also controlled for age of the participant, as this was found to correlate significantly with the performance on the practical test, as could be seen in Table 3. Before creating the model, a regression analysis expected the effect of Immersion to be around 23 points, after controlling for age, \( b = 22.70, t = 4.75, p < 0.01 \).

This model, \( \chi^2(10) = 55.07, R^2 = 0.83, p < 0.01 \), showed that Immersion has a significant positive effect on all the variables, as can be seen in Table 4. However, of all these variables only Involvement had a significant positive effect on the performance on the practical test, \( C = 4.31, Z = 2.01, p = 0.04 \).

Table 4 shows that none of the explanatory variables can significantly explain the indirect effects of Immersion, but it seems that it can be partly attributed to Involvement, \( C = 4.78, Z = 1.80, p = 0.07 \). A graphical overview of this model can be found in Figure 5.

### Table 4

Overview of the different coefficients to and from the explanatory variables as used in the model of Immersion on performance on the practical test (ppt), after controlling for age.

<table>
<thead>
<tr>
<th>Variable:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immersion to variable:</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Interest</td>
</tr>
<tr>
<td>Spatial Presence</td>
</tr>
<tr>
<td>Involvement</td>
</tr>
<tr>
<td>Realism</td>
</tr>
<tr>
<td>Immersion</td>
</tr>
</tbody>
</table>

Note: Overall \( R^2 = 0.83, p < 0.01** \). C stands for Coefficient

a. This is the direct effect of Immersion

b. This is the total effect of Immersion, which is the direct effect of Immersion + the sum of the other indirect effects

\* p<0.05 \**p<0.01
In the second model, with $\chi^2(10) = 127.36$, $R^2 = 0.41$, $p < 0.01$, as presented in Figure 6, it was investigated how Interactivity affects performance on the practical test. This model also controlled for age of the participants, and investigated the effects of Interactivity through Interest, Spatial presence, Involvement, and Realism. Interactivity was expected, by an initial regression analysis, to have a significant positive effect on performance on the practical test, $b = 11.56$, $t = 2.21$, $p = 0.03$.

The model showed that Interactivity significantly increases Interest, Spatial Presence, and Involvement, see Table 5. Furthermore, it was found that Involvement leads to significantly increased scores on the practical test, $C = 4.67$, $Z = 2.07$, $p = 0.04$. It also seems to be that interest influences performance, but the data cannot provide significant results. However, none of the explanatory variables could explain the differences found in performance on the practical test due to Interactivity, as all the indirect effects did not lead to significant differences.
Table 5
Overview of the different coefficients to and from the explanatory variables as used in the model of Interactivity on performance on the practical test (ppt), after controlling for age.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interactivity to variable</th>
<th>Variable to ppt</th>
<th>Indirect effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>Z</td>
<td>p</td>
</tr>
<tr>
<td>Interest</td>
<td>0.72</td>
<td>2.16</td>
<td>0.03*</td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>0.72</td>
<td>2.01</td>
<td>0.04*</td>
</tr>
<tr>
<td>Involvement</td>
<td>0.30</td>
<td>0.99</td>
<td>0.32</td>
</tr>
<tr>
<td>Realism</td>
<td>0.74</td>
<td>2.63</td>
<td>0.01**</td>
</tr>
<tr>
<td>Interactivity</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: Overall $R^2 = 0.41$, $p < 0.01**$, C stands for Coefficient
a. This is the direct effect of Interactivity
b. This is the total effect of Interactivity, which is the direct effect of Immersion + the sum of the other indirect effects
*p<0.05 **p<0.01

Figure 6, Model of Interactivity on performance on the practical test

The third model, $\chi^2(10) = 42.02$, $R^2 = 0.82$, $p < 0.01$, combined both Model 1 and Model 2, and used both Immersion and Interactivity as independent variables. They were both expected to have a positive effect on performance on the practical test, as was found by an initial regression analysis, where Immersion, $b = 22.71$, $t = 4.97$, $p < 0.01$, was expected to
have a greater effect than Interactivity, $b = 11.59, t = 2.60, p = 0.01$. The model can be seen in Figure 7, and the different coefficients can be found in Table 6.

Immersion has a significant positive effect on all the variables, after controlling for age. The same cannot be said for Interactivity, which does not have a significant effect on Involvement. From the mediators both Spatial Presence and Involvement have a significant effect on Performance on the Practical test. However, Spatial Presence has a significant negative effect, $C = -4.90, Z = -1.96, p < 0.05$, and Involvement a positive effect, $C = 4.58, Z = 2.30, p = 0.02$.

The indirect effect of Immersion via Involvement also provides a significant difference on Performance, $C = 5.05, Z = 2.00, p = 0.05$, where the indirect effect of Immersion via Spatial presence fails to provide a significant result, $C = -10.37, Z = -1.91, p = 0.06$.

In Table 7 the total effects of the mediators are presented. This table combines both the “indirect effects” coefficients from Table 6. This provides an overview of how the different mediators influence the performance on the practical test. It fails to support the hypothesised effect of the mediators at the five percent level, but it seems that Spatial Presence and Involvement have the effects as earlier specified, where Spatial Presence shows a negative effect on performance on the practical test, and Involvement a positive effect.
Table 6
Overview of the different coefficients to and from the explanatory variables as used in the model of Interactivity and Immersion on performance on the practical test (ppt), after controlling for age

<table>
<thead>
<tr>
<th>Immersion</th>
<th>Immersion to variable:</th>
<th>Variable to ppt:</th>
<th>Indirect effect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable:</td>
<td>C Z p</td>
<td>C Z p</td>
<td>C Z p</td>
</tr>
<tr>
<td>Interest</td>
<td>2.08 9.88 0.00**</td>
<td>-0.46 -0.17 0.87</td>
<td>-0.97 -0.17 0.87</td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>2.12 8.88 0.00**</td>
<td>-4.90 -1.96 0.05*</td>
<td>-10.37 -1.91 0.06</td>
</tr>
<tr>
<td>Involvement</td>
<td>1.10 4.04 0.00**</td>
<td>-4.58 2.30 0.02*</td>
<td>5.04 2.00 0.05*</td>
</tr>
<tr>
<td>Realism</td>
<td>1.29 5.57 0.00**</td>
<td>-0.53 -0.21 0.83</td>
<td>-0.68 -0.21 0.83</td>
</tr>
<tr>
<td>Immersion</td>
<td>- - -</td>
<td>29.49* 4.25 0.00**</td>
<td>22.51b 4.98 0.00**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interactivity</th>
<th>Interactivity to variable:</th>
<th>Variable to ppt:</th>
<th>Indirect effect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable:</td>
<td>C Z p</td>
<td>C Z p</td>
<td>C Z p</td>
</tr>
<tr>
<td>Interest</td>
<td>0.66 3.12 0.00**</td>
<td>-0.46 -0.17 0.87</td>
<td>-0.31 -0.17 0.87</td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>0.65 2.73 0.01**</td>
<td>-4.90 -1.96 0.05*</td>
<td>-3.18 -1.59 0.11</td>
</tr>
<tr>
<td>Involvement</td>
<td>0.27 0.98 0.33</td>
<td>4.58 2.30 0.02*</td>
<td>1.22 0.90 0.37</td>
</tr>
<tr>
<td>Realism</td>
<td>0.70 3.02 0.00**</td>
<td>-0.53 -0.21 0.83</td>
<td>-0.37 -0.21 0.83</td>
</tr>
<tr>
<td>Interactivity</td>
<td>- - -</td>
<td>14.13c 3.15 0.00**</td>
<td>11.50d 2.61 0.01**</td>
</tr>
</tbody>
</table>

Overall $R^2 = 0.82$, $p < 0.01$, C stands for Coefficient

a. This is the direct effect of Immersion
b. This is the total effect of Immersion, which is the direct effect of Immersion + the sum of the other indirect effects
c. This is the direct effect of Interactivity
d. This is the total effect of Interactivity, which is the direct effect of Immersion + the sum of the other indirect effects

*p<0.05  **p<0.01

Table 7
Overview of the different total effects of the explanatory variables as used in the model of Interactivity and Immersion on performance on the practical test (ppt), after controlling for age

<table>
<thead>
<tr>
<th>Total effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable:</td>
</tr>
<tr>
<td>Interest</td>
</tr>
<tr>
<td>Spatial Presence</td>
</tr>
<tr>
<td>Involvement</td>
</tr>
<tr>
<td>Realism</td>
</tr>
</tbody>
</table>

Note: C stands for Coefficient

*p<0.05  **p<0.01
From the different analyses performed for the performance on the practical test both Immersion and Interactivity showed to have a significant positive effect on performance. These effects are, according to the models, mostly because of improved Involvement, and perhaps due to higher Interest.

The fourth model investigated the effects of Immersion on Performance on the theoretical test. Immersion was expected to not have a significant effect on Performance on the theoretical test, \( b = -3.23, t = -0.50, p = 0.55 \), this was based on an initial regression analysis and in accordance with the result of the second hypothesis, where there were no significant differences between the two immersion groups regarding performance on the theoretical test. However, the model was still created to check for any relevant insights. The model, \( \chi^2 (6) = 51.45, R^2 = 0.76, p < 0.01 \), as can be seen in Figure 8, does not control for age, as this was not found to be of effect on performance on the theoretical test. Table 8 presents the different coefficients of this model. It shows that Immersion has a significant negative
The fifth model, $\chi^2(6) = 120.03$, $R^2 = 0.26$, $p < 0.01$, investigated the effects of Interactivity on Performance on the theoretical test. Here an initial regression analysis expected Interactivity to have a significant negative effect on Performance on the theoretical
test, $b = -13.17, t = -2.57, p = 0.01$, as was also shown by the third hypothesis. The different coefficients of the model can be found in Table 9 and a graphical overview in Figure 9.

The table shows that Spatial Presence has a significant negative effect on Performance, $C = -6.05, Z = -2.14, p = 0.03$, but that the significance of the indirect effect of Interactivity via Spatial Presence cannot be supported by the current data. This model does not provide any significant indirect effects of the different mediators, and can therefore not explain the effects of Interactivity on performance on the theoretical test, although it seems that it can be partially attributed to Spatial Presence.

| Table 9 |
|---|---|---|---|---|---|---|---|---|
| Overview of the different coefficients to and from the explanatory variables as used in the model of Interactivity on performance on the theoretical test (ptt) |

<table>
<thead>
<tr>
<th>Variable:</th>
<th>C</th>
<th>Z</th>
<th>p</th>
<th>C</th>
<th>Z</th>
<th>p</th>
<th>C</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest</td>
<td>0.72</td>
<td>2.16</td>
<td>0.03*</td>
<td>2.67</td>
<td>0.91</td>
<td>0.36</td>
<td>1.94</td>
<td>0.84</td>
<td>0.40</td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>0.72</td>
<td>2.01</td>
<td>0.04*</td>
<td>-6.05</td>
<td>-2.14</td>
<td>0.03*</td>
<td>-4.33</td>
<td>-1.46</td>
<td>0.14</td>
</tr>
<tr>
<td>Involvement</td>
<td>0.30</td>
<td>0.99</td>
<td>0.32</td>
<td>1.96</td>
<td>0.81</td>
<td>0.42</td>
<td>0.59</td>
<td>0.63</td>
<td>0.53</td>
</tr>
<tr>
<td>Realism</td>
<td>0.74</td>
<td>2.63</td>
<td>0.01**</td>
<td>1.31</td>
<td>0.43</td>
<td>0.67</td>
<td>0.97</td>
<td>0.43</td>
<td>0.67</td>
</tr>
<tr>
<td>Interactivity</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-12.34</td>
<td>-2.40</td>
<td>0.02*</td>
<td>-13.18</td>
<td>-2.42</td>
<td>0.02*</td>
</tr>
</tbody>
</table>

Note: Overall $R^2 = 0.26, p < 0.01**, C$ stands for Coefficient
a. This is the direct effect of Interactivity
b. This is the total effect of Interactivity, which is the direct effect of Immersion + the sum of the other indirect effects
* $p<0.05$ **$p<0.01$
The final and sixth model investigated the effects of Immersion and Interactivity on performance on the theoretical test. Immersion was not expected to have a significant effect, $b = -2.82, t = -0.44, p = 0.59$, but Interactivity was expected to have a significant effect on performance on the theoretical test, $b = -13.09, t = -2.53, p = 0.01$. The results of this model, $R^2 = 0.82, p < 0.01$, can be found in Table 10 and Table 11. Table 10 shows that the indirect effect of Immersion through Spatial Presence gives a significant negative coefficient, $C = -13.26, Z = -2.02, p = 0.04$. All the other mediators provided no significant results. The total effect of Spatial Presence, as can be seen in Table 11, provided a significant negative result on performance on the theoretical test, $C = -17.33, Z = -2.01, p = 0.04$. 

Figure 9. Model of Immersion on performance on the theoretical test
### Table 10
Overview of the different coefficients to and from the explanatory variables as used in the model of Immersion and Interactivity on performance on the theoretical test (ptt)

<table>
<thead>
<tr>
<th>Variable:</th>
<th>Immersion to variable:</th>
<th>Variable to ptt:</th>
<th>Indirect effect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Z</td>
<td>p</td>
<td>C</td>
</tr>
<tr>
<td>Interest</td>
<td>2.08</td>
<td>9.88</td>
<td>0.00**</td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>2.12</td>
<td>8.88</td>
<td>0.00**</td>
</tr>
<tr>
<td>Involvement</td>
<td>1.10</td>
<td>4.04</td>
<td>0.00**</td>
</tr>
<tr>
<td>Realism</td>
<td>1.29</td>
<td>5.57</td>
<td>0.00**</td>
</tr>
<tr>
<td>Immersion</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable:</th>
<th>Interactivity to variable:</th>
<th>Variable to ptt:</th>
<th>Indirect effect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Z</td>
<td>p</td>
<td>C</td>
</tr>
<tr>
<td>Interest</td>
<td>0.66</td>
<td>3.12</td>
<td>0.00**</td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>0.65</td>
<td>2.73</td>
<td>0.01**</td>
</tr>
<tr>
<td>Involvement</td>
<td>0.27</td>
<td>0.98</td>
<td>0.33</td>
</tr>
<tr>
<td>Realism</td>
<td>0.70</td>
<td>3.02</td>
<td>0.00**</td>
</tr>
<tr>
<td>Interactivity</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Overall $R^2 = 0.82$, $p < 0.01**$, C stands for Coefficient

- **a.** This is the direct effect of Immersion
- **b.** This is the total effect of Immersion, which is the direct effect of Immersion + the sum of the other indirect effects
- **c.** This is the direct effect of Interactivity
- **d.** This is the total effect of Interactivity, which is the direct effect of Immersion + the sum of the other indirect effects

* $p<0.05$ **$p<0.01$

### Table 11
Overview of the different total effects of the explanatory variables as used in the model of Interactivity and Immersion on performance on the theoretical test (ptt).

<table>
<thead>
<tr>
<th>Variable:</th>
<th>Total effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Z</td>
</tr>
<tr>
<td>Interest</td>
<td>6.49</td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>-17.33</td>
</tr>
<tr>
<td>Involvement</td>
<td>2.68</td>
</tr>
<tr>
<td>Realism</td>
<td>2.60</td>
</tr>
</tbody>
</table>

Note: C stands for Coefficient

* $p<0.05$ **$p<0.01$
The different analyses performed for the performance on the theoretical test showed that the degrading performance on the theoretical test can probably be attributed to higher Spatial Presence.

Therefore, this data fails to find support for the fourth hypothesis, *H4: Interactivity and Immersion have a significant positive effect on performance through interest, spatial presence, involvement, and realism as mediators*, but it does provide insights into how Immersion and Interactivity influence the performance on both the theoretical and practical test.

**Discussion**

This study was conducted to gain more insight into the effectiveness of VR training for procedural knowledge. A second aim of the study was to get more insight into how the level of immersion and interactivity of teaching media influence performance on subsequent
theoretical and practical tests. These aims were also covered by the two initial research questions, which were: “What is the effectiveness of training with a VR application, with regards to procedural knowledge?” and “How do Immersion and Interactivity influence the effectiveness of the VR application?”.

To answer these questions an experiment was conducted. Here, participants were divided over a 2 x 2 subject design, got to learn a procedure on a navy ship, and were assessed by a theoretical test as well as a practical test. The performance on the theoretical test was assessed by the number of questions correct and the number of steps correct on the task, as well as the time that was necessary to complete both parts of the test; these four numbers were all averaged into a single score for theoretical performance. Performance on the practical test was assessed by the time needed to finish the learned procedure in a simulated environment. Both scores ranged from 0 to 100, where a higher score is a better score. Performance on the theoretical test and performance on the practical test will be discussed separately, as it is hard to validly compare performance between different means of testing (Clark, 1989).

**Practical Test**

It was found that participants who trained with the VR application outperformed the participants in the other conditions on the practical test, due to them being faster on completing the simulation.

**Immersion.** It has been found that Immersion had a large positive effect on performance in the theoretical test. This can be because that Immersion showed to have a large effect on Interest and Spatial Presence, a medium effect on Realism, and a small effect on Involvement. Additionally, Interest also showed to have a medium effect, and Spatial Presence, Involvement, and Realism all showed small effects on Performance on the practical test, according to the correlation matrix. These findings were in accordance with the expected beneficial effects of Interest (Castillo-Merino & Serradell-López, 2014; Pantelidis, 2010),
Spatial Presence (Witmer & Singer, 1998), Involvement (Chittaro & Buttussi, 2015), and Realism (Slater et al., 1996).

Contrary, the mediation models assessing performance on the practical test showed a different view. These difference in found results can be because of multicollinearity, which is suggested because multicollinearity can cause a change in the signs as well as in the magnitudes of the coefficients. Another problem that can occur because of multicollinearity is difficulty when assessing the relative importance of the variables.

In the different mediation models only Involvement was shown to have a small effect on performance on the practical test. In Model 1, which investigated the effects of immersion on performance on the practical test, it seemed that the indirect effect of Immersion via Involvement had a positive effect on performance, but it could not be supported at the five percent level. A small but significant indirect effect was shown in Model 3, which used both immersion and interactivity as independent variables.

In these models, Involvement relates to how well the activities and events in the Virtual Reality application can captivate and maintain the attention of the user (Witmer & Singer, 1998, p. 227). This states that a higher level of involvement means that the user is more engaged with the material presented, which is also often the case in VR applications (Regian, Shebilske, & Monk, 1992). It has been shown that forms of media that are more engaging provide better learning experiences when learning procedures (Chittaro & Buttussi, 2015), where participants scored higher on retention when redoing the task after a two-week period. Based on this earlier research and the found effects, it can be argued that the improved performance on the practical test, as observed for the participants trained in the VR application, can be explained by higher levels of Involvement due to a more immersive experience.
However, not all of the observed beneficial effects of Immersion on performance on the practical test can be explained by an increase in Involvement. A large part of the total effect is unaccounted for, and other explanations, which were not covered in the scope of this study, might shed insight into what causes the beneficial effect of immersion on performance on the practical test.

One of these might be the explanation offered by Sowndarajan, Wang and Bowman (2008). They found that higher levels of Immersion lead to higher performance when learning procedures, and that this effect is mostly due to a higher field of vision (FOV). This higher field of vision would, as argued, lead to more spatial cues, and therefore a higher performance on the task. This could have been of influence, as the more immersive forms of media provide a higher FOV compared to the pictures used in the tow different manuals.

**Interactivity.** Besides Immersion, Interactivity also showed to have a medium positive effect on performance on the practical test. This beneficial effect was as expected, since several papers showed that Interactivity could be linked to an improved performance. (Fletcher, 1990; Jang et al., 2017; Steuer, 1992).

This study showed a small positive effect of Interactivity on Interest and Realism. Furthermore, improved performance on the practical test has been linked to higher Interest, Spatial Presence, Involvement, and Realism.

Of these different mediators, only Interest showed to have a small effect on performance. This effect was in accordance with the work of Pantelidis and Ren et al. (2010; 2015), who indicated that VR applications can be used to make learning more interesting. As interest is linked to a more engaging experience, and more engagement is linked to higher performance on procedural tasks (Chittaro & Buttussi, 2015), this could be seen as an explanation of why the participants in the interactive conditions outperform the other
participants with regards to the practical test. However, the indirect effect of Interactivity through Interest cannot be supported by our model, and can therefore not be taken as a definitive explanation of the improved performance in the practical test.

Since this analysis cannot explain the effects of Interactivity, other explanations besides the one previously given will be explored of why participants with higher interactivity outperform the other participants on the practical test. One of these is given by Fletcher (1990), who stated that students retain more of the material when they get to interact with it, as opposed to merely reviewing it, which would result in higher performance when having to do the procedure for the test.

Another reason why Interactivity was found to have a beneficial effect on performance in the practical task was that being able to interact with an environment, and to be able to walk around in it, was linked to a better spatial layout of a virtual environment compared to viewing it passively (Brooks, Attree, Rose, Clifford, & Leadbetter, 1999; Tüzün & Özdinç, 2016). This means that participants in the more interactive conditions would remember the lay-out of the ship better and therefore be able to better recall where each room and control console is situated, which would result in a faster time on the procedure.

This study showed that participants in the VR condition were faster when redoing the simulation, compared to the other participants. This can be due to the beneficial effect of Immersion and Interactivity and their effect on Interest, Spatial Presence, Involvement, and Realism. However, this explanation was not supported by all of the analyses, and other explanations were investigated to account for the differences in performance on the practical test.
Theoretical test

The differences in the practical test were not reflected on the theoretical test. Here the participants that trained using the Virtual Reality application were not found to be better at the theoretical test, compared to the participants that used other learning methods. Furthermore, no significant effect of Immersion was found that would lead to differences in performance on the theoretical test.

Contrary, it has been shown that the participants in the more Interactive conditions, the VR application and the interactive manual, scored significantly worse than the participants in the other two conditions. This could be because of detrimental effects of Spatial Presence, which were found by the different models.

This result was not as expected based on the earlier discussed literature. For example, Ai-Lim Lee, Wong and Fung found that higher spatial presence was linked to higher performance on a test after using a desktop VR to learn the anatomy of a frog (2010).

However, an explanation of why Spatial Presence has a detrimental effect was given by Makransky, Terkildsen and Mayer (2017). They conducted a study where they compared knowledge gain with low immersive (PC) and high immersive (HMD) VR, and measured the brain activity while training. They found that the high immersive VR conditions reported significantly more presence, had significantly less knowledge gain, and had more cognitive load than the low immersive conditions. Based on this they argued that more presence might distract the learner, resulting in worse performance.

This same explanation was given by Andrews, Rose, Leadbetter, Attree, and Painter (1995), who argued that a more interactive experience would provide more distractions for participants. This would influence their capability to remember details about the environment, which were asked during the theoretical test. This explanation was also suggested by Brooks
et al. (1999), but this could not be deemed significant in their research. These findings show that spatial presence has a negative effect on performance on the theoretical test, due to distraction of the user.

Based on this study it can be argued that training in VR should be encouraged when someone needs to learn how to do a certain procedure, where theoretical knowledge is of less importance. Here, the user will benefit from the higher Immersion and Interactivity, and thus be able to perform the procedure faster than when trained using one of the other conditions.

However, when someone also needs to have theoretical knowledge of the procedure, it can be argued that it is more beneficial to use a source of media that is less immersive and less interactive than VR. This is encouraged because of the small, but significant, detrimental effect of Spatial Presence on performance on the theoretical test, and the large effect that Immersion has on Spatial Presence.

This research has managed to produce more insight into the effectiveness of using Virtual Reality for procedural training. It also worked to explain these sometimes-beneficial effects by investigating how Immersion and Interactivity influence performance, and thereby managed to provide suggestions for when to use, or not use, VR for education. However, there were also several limitations that might have altered the validity of this research.

**Limitations**

The first limitation of this study was that the final practical test consisted of a virtual task, instead of performing the task in real life. The virtual task was used because of the unavailability of the real ship during data collection, and using the virtual reality application was the most similar environment available. Because of this, it could have been likely that the users that were trained using the VR application performed better than the other users, because they were familiar with the controls of the application. This could have caused them
to be more competent in moving around in the virtual environment and interacting with the
different objects; which can result in a higher score for performance, since they will be faster.
These faster times, due to being used to the controls, can be interpreted as faster times due to
having retained the various steps in the procedure better, which might lead to wrong
conclusions.

It has been tried to control for this by letting all participants practice with the controls
of the simulation until they indicated that they understood and knew them. However, it might
be possible that these few minutes were not enough to give all the participants the same level
of competency regarding the controls used in the simulation.

The second limitation was that the sample used in this study consisted of employees of
an IT company, who are not a correct representation of the future users of the system. This
sample was used due to their availability. The sample used could have had an impact on the
results since they were almost all highly educated, and could have had a predisposition for
learning new procedures. Furthermore, the fact that all participants work for an IT company
could have caused them to be more interested in working with computer based technologies;
causing them to rate the conditions based on computers higher than the others. This means
that the results found can be indicative for this specific population, and it might be that they
are not representable for the population that must learn the procedures.

The final limitation of this study was in the quality of the data. It was found that many
of the investigated variables correlated with each other. Even though VIF values of these
variables were deemed acceptable, some of the results conflicted with each other, and can
therefore be deemed a possible victim of problems associated with multicollinearity. An
example of these are the conflicting signs of Spatial Presence between the correlation matrix
and the models in case of Immersion with regards to performance on the practical test.
Several solutions were investigated, and were not able to provide significant improvements in
the results. It was therefore hard to investigate which factor is of the most importance for the success, or failure, of VR-based technologies for procedural training.

**Future work**

Based on the limitations there are several suggestions for future research. The first suggestion is to replicate this study, but to conduct the practical test in real-life on a real ship; and to see whether the conclusion of this study still holds. This is suggested because this will provide better insights into whether the improved performance transfers from the simulation to the real-life situation, and thus whether training in VR provides satisfactory results.

A second suggestion for future research is that the real future users of the system will be used as participants, instead of the readily available participants used in this study. This can be combined with the first suggestion to get a more representative result of the performance of the users.

These two suggestions can be combined into one replication of this study, but then conducted in the real setting of future use. To do this the practical test should be conducted on a real ship, and the future users of the system should be used as participants, instead of the more easily available participants used in this study. This suggestion will be able to give more insights into the real-life application of VR for procedural training.

The final suggestion for future research is to aim and get more insight into the relative importance of the several factors. This is suggested because of the multicollinearity that was present in this study. By tweaking the experiment set-up, it should be possible to differ on several of the variables, and thus get more insight into their relevant importance for procedural training.
Conclusion

This study aimed to quantify the benefits of Virtual Reality training for procedural training. It found that participants who trained using the virtual reality application performed better than the other conditions on the practical test, but did not score significantly different on the theoretical test. The higher performance on the practical test can be attributed due to the beneficial effects of Interest, Involvement, Spatial Presence, and Realism caused by higher Immersion and Interactivity. Of these variables only, Involvement was shown by the models to have a small significantly positive effect on performance on the practical test. However, it cannot be ruled out that these found beneficial effects on performance on the practical test were because the test consisted of the same VR application that was used for training in the VR condition.

The negative differences that were found on the theoretical test can be partly explained by the small detrimental effects of Spatial Presence, due to higher Interaction and Immersion. In this case more Spatial Presence will result in more distractions, and thus less ability to learn new knowledge. Future research is encouraged to replicate this study in a real-life setting, to see whether the training with the VR application transfers to real-life performance on the task.
References


EFFECTS OF IMMERSION AND INTERACTIVITY


https://doi.org/10.1518/107118192786749243


2018, from https://www.merriam-webster.com/dictionary/virtual reality


Wiederhold, B. K., Davis, R., & Wiederhold, M. D. (1998). The effects of immersiveness on physiology. In *Virtual environments in clinical psychology and neuroscience: Methods and techniques in advanced patient–therapist interaction.* (pp. 52–60). Amsterdam,
