Don't Judge a Book by its CoVR: Learning and Training in virtual reality; the effects of two levels of immersion

Kendall Carmody, Meredith Carroll
Florida Institute of Technology
kcarmody2016@my.fit.edu, mcarroll@fit.edu

ABSTRACT

Immersive devices, such as virtual reality (VR) devices are swiftly being integrated into military and civil operations as advanced tools for training, with studies observing the beneficial aspects of employing these devices for training cognitive skills (Xie et al., 2021). However, further research is needed to understand specific characteristics of the learning effectiveness of the technology and identify which levels of immersion benefit specific learning constructs in training. This study examined the effect of immersion on four constructs related to learning, including memory retention, learner engagement, learning performance in the form of knowledge acquisition, and perceived learning in a virtual maintenance training task. The study aimed to determine how these constructs are impacted by two different levels of immersion during a virtual maintenance training task by means of either (a) a desktop computer and keyboard; Low-immersion Virtual Reality (LiVR) or (b) a virtual reality headset and controllers; High-immersion Virtual Reality (HiVR). To achieve this, a between-subjects experimental design was employed with 25 participants completing a maintenance training task on a simulator developed in collaboration with the Air Force Research Lab's (AFRLs) Gaming Research Integration for Learning Laboratory (GRILL). A memory retention quiz, the User Engagement Scale (UES) short-form, a pre-post knowledge assessment, and the Cognitive, Affective and Psychomotor (CAP) Perceived Learning Scale were administered to capture the dependent variable learning constructs, with level of immersion acting as the independent variable. Results revealed that both immersive testbeds enhanced participants' knowledge acquisition and that engagement was significantly higher in the HiVR condition. Trends were also observed with respect to the impact that immersion had on participant's memory retention. The findings can help to inform the procedural training community on the benefits of immersive devices with respect to virtual environment training for hands-on spatial tasks

ABOUT THE AUTHORS

Kendall Carmody is a current Ph.D. student, who graduated with a MS in Aviation Human Factors from the College of Aeronautics at Florida Institute of Technology and works at Florida Tech's ATLAS lab as a lead researcher. She has a B.S in Aeronautical Science, and a Minor in Aviation Environmental Science. Kendall has worked as an Airport Operations Specialist intern at Rock Hill-York County airport and is a current research fellow for the Air Force Research Lab (AFRL). Kendall's areas of focus include human factors, virtual environments (VE), virtual/augmented reality, human-agent teaming (HAT) and urban air mobility (UAM).

Meredith Carroll is a Professor of Aviation Human Factors and Director of the Advancing Technology-interaction and Learning in Aviation Systems (ATLAS) Lab at Florida Institute of Technology's College of Aeronautics. She has nearly 20 years of experience studying human/team performance and training in complex systems, with focuses in cognition and learning, decision making, human-machine teaming, training system design/evaluation, and performance assessment. She received her bachelor's degree in Aerospace Engineering from the University of Virginia, her Master's degree in Aviation Science from Florida Institute of Technology and her Ph.D. in Applied Experimental Psychology and Human Factors from the University of Central Florida.

Don't judge a book by its coVR: Learning and Training in Virtual Reality, the Effects of Two Levels of Immersion.

Kendall Carmody, Meredith Carroll
Florida Institute of Technology
kcarmody2016@my.fit.edu, mcarroll@fit.edu

INTRODUCTION

Aviation and industrial maintenance tasks contain complex procedures that require a significant amount of training (Gavish et al., 2015). Therefore, identifying ways to enhance the efficiency of training in these fields would be operationally and economically beneficial. Due to the inexpensive, immersive, and interactive capabilities of VR technology, industrial and aviation maintenance training are domains that have begun utilizing VR to enhance training. However, current training mechanisms in these domains have not taken full advantage of these devices to enhance memory recall and spatial awareness in a more affordable manner (Renganayagalu et al., 2021). This is due in part because the technology is constantly evolving, and more research is needed to understand VR training effectiveness in various contexts.

Different levels of immersion afforded by VR devices can deliver heightened integration and realism between the human and the simulation, allowing for accessible and affordable training options across domains. Before implementing VR as a new training mechanism into these industries, it is important to understand the learning effects and which levels of immersion render the best results in a particular training context. Therefore, the purpose of the study was to examine the effect of immersion on four constructs related to learning: memory retention, learner engagement, learning performance in the form of knowledge acquisition, and perceived learning in a virtual maintenance training task.

Defining Immersion

Immersion, in its most simple form, is an advanced sense of "being there" or being present in a virtual world (Pausch, Proffitt, Williams 1997). Quantifying different levels of immersion has been a difficult task ever since VR's early conceptions, due in part to the continuously evolving and complex nature of VR systems (Rakowski & Gruber, 2019). Although various distinctions have been made when defining the level of immersion in a virtual environment, the definitions are complex, leading to confusion and misinterpretation when trying to simply define the terms (Girvan, 2018; Pan & Hamilton, 2018). However, there has recently been two widely accepted definitions of levels of immersion in virtual environments: Low Immersion Virtual Reality (LiVR) and High Immersive Virtual Reality (HiVR) (Rakowski & Gruber, 2019; Lee & Wong, 2014; Markansky et al., 2019). LiVR as defined by Rakowski and Gruber (2020) is "a computer-generated, three-dimensional virtual space experienced through standard audio-visual equipment, such as a desktop computer with a two-dimensional monitor" (p. 1). Rakowski and Gruber define HiVR as a "computer generated 360° virtual space that can be perceived as being spatially realistic, due to the high immersion afforded by a head-mounted device" (p. 1). These definitions simplify the complexities involved in the categorization of immersive virtual environments and allow for easy differentiation between the two levels.

Immersion and Memory Retention

Another benefit of increased immersion is increased memory retention. The ability for VR headsets to induce high levels of immersion has been shown to aid users in their memory retention, due to the affordance of additional spatial awareness (Krokos, Plaisant, & Varshney 2018). Taking advantage of the ability to assemble one's thought processes spatially and visually can lead to advanced capabilities in memory recall (Jaynes, 1976; Novak & Canas, 2008). Prior studies have been conducted to understand the discrepancies between memory recall in differing levels of immersion. This has been accomplished through creating virtual memory palaces in different levels of immersion, in which the user can visualize and move around a simulated environment to retain information (Krokos, Plaisant & Varshney, 2018). A memory palace is a visually organized structure in one's mind that is navigated through and used to remember information, whether it be a song, a speech, or an image (Yates, 1992). In Krokos, Plaisant &

Varshney's study participants were sectioned into HiVR or LiVR groups and were given five minutes in each scenario to memorize the images, until the image location was then replaced with a number. The participants were then required to verbally recall the name of the face that was previously in each numbered location. Results indicated that there was a statistically significant effect for the display condition: participants were able to recall more accurately in the HiVR head mounted display condition compared to the LiVR desktop condition (Krokos et al., 2018). A similar study, which highlights that a more immersive virtual environment can induce a higher sense of presence and in turn memory retention, was conducted by Randy Pausch, Dennis Proffitt, and George Williams to assess how immersion and presence impact a participant's memory retention in differing virtual environments (1997). In this case, participants were tasked with searching for uncamouflaged targets at random locations in a virtual room. There were "VR Users" and "Desktop Users". During each search task, the room contained letters arranged on the walls, ceiling and floor of the virtual room. The users were shown the target letter once and then the room would reset with the target letter in a different location among camouflaged letters. When they found the target letter, users said "there it is". Each user performed five searches that were then averaged to form a single data point for the specific user. Following the five searches, users then performed a sequence of searches in which there was a 50% chance the target letter was not there. In this case, the users were instructed to either locate the target or declare that no target existed. Time was measured for each search mission for both groups. Results indicated that when targets were present, both the VR Users and Desktop Users performed the same. However, when the target was absent, VR users were substantially quicker at noticing this than the Desktop Users. Pausch, Proffitt and Williams speculated that a reason for this is that VR Users were able to build a tighter mental model and frame-of-reference or "memory of spatial cues", due to the sense of presence or "being there" (p. 1). The ability to link movement to specific information greatly increase the chances of recalling the information, which is why immersive environments are extremely beneficial for learning novice operations, procedures and techniques (Murcia-Lopez & Steed, 2016).

Immersion and Engagement

Another construct that has shown to benefit from increased immersion is learner engagement. Understanding how to create more engaging experiences with users is a prevalent target design aspect for newer technologies (Hassenzahl & Tractinsky, 2006). Definitions of the construct of engagement range from something that attracts and holds our attention (Laurel, 1993), to a first-person experience that emphasizes playfulness and evokes our senses (Chapman, 1997). O'Brien and Toms (2008) define engagement within the realms of technology as "a quality of user experiences with technology that is characterized by challenge, aesthetic and sensory appeal, feedback novelty, interactivity, perceived control and time, awareness, motivation, interest and affect." (p. 23) Recent studies have examined how HiVR influences engagement. One such study, conducted in 2018 by Devon Allcoat and Adrian von Muhlenen assigned participants to one of three learning conditions: Traditional (Textbook), Video style, and VR (using the HTC Vive Headset). Each learning condition employed the same learning materials (images and information on a growing plant cell). The textbook condition contained screenshots of the 3D plant cell model, accompanied by informative text on the subject. The video condition included a 2D recording with the same informative text on the screen next to the video. The VR condition watched a simulation of the plant cell and were able to fully interact with the objects in the virtual room (i.e., the plant cell appeared as a floating object in the center of the virtual room). Results identified that engagement ratings were significantly higher in the VR learning experience than in the two other conditions, furthering the evidence that a HiVR condition can render higher engagement among participants compared to LiVR (Desktop Video) and Traditional (Textbook) conditions.

Allcoat et al. (2021) conducted a similar study that examined the learning effectiveness of virtual and mixed reality (MR) environments on student learning and engagement. MR can be broadly classified as a type of virtual environment in which virtual objects projected by the device interact with real world objects (Speicher, Hall & Nebeling, 2019; Allcoat, et al. 2021). In this study, participants were recruited to participate in a virtual environment-based class covering solar panels. Each student was randomly assigned to one of three conditions: Traditional, VR or MR. The learning materials utilized in each environment were based on real classroom information. The topic focused on solar-power panel efficiency and parameters that affect it. The results of the study indicate that participants' knowledge improved from pre- to post-test in all conditions, with a trend identified for more learning in the VR condition (although it did not reach statistical significance). However, there was a significant difference in engagement levels experienced in the VR and MR groups compared to the traditional groups, with the VR and MR groups reporting being significantly more engaged (Allcoat, et al., 2021). These results provide further support those higher levels of immersion generate increased engagement.

Immersion and Knowledge Acquisition

Immersion has also been shown to enhance learning performance in the form of Knowledge Acquisition. Students, employees, and trainees tend to learn more when the instruction method is aligned with their particular learning styles and there is the ability to interact with, manipulate and visualize an environment, which is beneficial to learning (Trindade, Fiolhais & Almeida, 2002). A large number of educational institutions have already begun implementing 3-dimensional virtual environments into online courses to aid in learning (Domingo & Bradley, 2017). A study that investigated the effects of HiVR on learning performance was conducted by Guiterrez et al. (2008). This experimental study aimed to identify significant differences in learning performance within differing virtual environments (LiVR desktop environment vs. HiVR HMD). Learning performance was measured through knowledge acquisition of medical students. Participants were randomly separated into two groups: a fully-immersed group (i.e., participants donned an HMD with eye tracking) and a partially-immersed control group (i.e., participants assessed and interacted with the VR environment using a computer and monitor). Each group was equipped with a joystick to manipulate the different medical utensils. The participants were tasked with learning how to diagnose the head injury using the joystick in either the VR condition, or the computer condition. A pre and post knowledge assessment test was administered to the participants, which consisted of "rating the relatedness of 72 pairs of concepts critical to the case" (Guiterrez et al., 2008, p. 158). An expert also took the knowledge assessment to use as a comparison for the assessment. An ANOVA was conducted on the scores with trial (pre vs. post test) as a within subjects variable and condition (fully vs. partially immersed) as a between subjects variable. The results indicated that both groups benefited from the VR simulation, and that the fully-immersed group showed significantly higher learner gains (increased learning performance through knowledge acquisition) than the partially immersed group when compared to the expert assessment.

Qian et al. (2020) conducted a study that looked at the effects of different levels of virtual environments on immersion, presence and learning performance within laboratory education. They defined a term known as VR fusion, described as utilizing some aspects of a virtual environment or VR and fusing it with the real world as an aspect of instruction. Forty-two participants were randomly assigned to one of three learning conditions: a PC environment that was be manually operated and navigated using a computer mouse, a VR environment that was manually operated and navigated using VR controllers, or a VR environment manually operated and navigated utilizing the participant's real hands. To evaluate the learning performance, participants were asked to describe the actions completed in the simulation or instruction method in detail, they were also given a subjective learning performance questionnaire that measured self-efficacy, intrinsic value and interest, and motivation. Their results indicated that immersive environments (VR controller group and VR real-hands group) improved learning performance and had strong positive effects on presence compared to the traditional instruction method. The results in the VR fusion condition indicated improvements in self efficacy, which increased student's confidence in learning and enhanced their performance. This is further supported by an incremental improvement from the PC condition to the VR and VR Fusion conditions during participant recall of the task they completed in the virtual environment. Participants were able to better recall and describe the task they completed in a VR environment compared to the traditional PC condition.

Immersion and Perceived Learning

Perceived learning is also an important aspect of virtual environments. Perceived learning refers to "The learning quality experienced by the participants" (Wong, et al., 2009). Studies have previously observed a student's "perceived learning" in a desktop virtual environment compared to traditional learning, (Wong et al., 2009; Chen et al., 2005; Yang & Heh, 2007); however, there is little empirical research on perceived learning in solely LiVR vs. HiVR. Understanding how level of immersion influences a user's perceived learning is important in understanding the impacts of advancing virtual learning and training.

A study conducted by Maffei (2020) aimed to examine the impact of AR on the geometry knowledge of high school students. Eighty-seven participants were utilized in the study and each participant was placed in either a traditional or AR group. The traditional group was taught geometry instruction via traditional methods utilizing two dimensional figures, while the AR group utilized AR technology to depict three dimensional figures. Maffei aimed to observe whether there was a difference in perceived learning scores utilizing the CAP Perceived Learning Scale. The CAP Perceived Learning scale ranges from zero to fifty-four, with zero being the lowest possible score (they did not perceive any learning gain). The difference between the two groups was not statistically significant;

however, the AR group had a higher mean score than the traditional group (M = 50.38 compared to M= 48.43). Future studies need to examine the effects of a higher level of immersion, such as fully-immersive virtual environment instead of an augmented one, on student perceived learning, to further understand how immersive technologies influence student learning perceptions.

Zhang, et al. (2017) conducted survey research using 180 participants to identify whether or not people believe utilizing VR would enhance their learning based on adapted items from Dalgarno, Hoehle, and Maor's prior studies (Dalgarno, Hedberg & Harper, 2002; Hoehle, Huff & Goode, 2012; Maor & Fraser, 2005). The items were modified to ensure relevancy in regards to VR-based learning. Items were measured with a 7-point Likert scale, from 1 (Strongly Disagree) to 7 (Strongly Agree). The test results of the responses concluded that participants believe that VR can influence reflective thinking and further improve perceived learning effectiveness. However, further research needs to be conducted to analyze this hypothesis (Zhang, et al., 2017).

METHODS

Participants

The target population for this study included maintenance trainees and workers in aviation and industrial domains. The accessible population for the study consisted of any students attending Florida Institute of Technology or within the Brevard County area, who were willing to participate in the study. The sample was selected utilizing convenience sampling through email, flyers, word of mouth, and snowballing. Twenty-five participants completed the study. The average age was 24 years old (S.D. = 1.33). The sample consisted of 18 males and 7 females. A total of 60% of participants were Caucasian, followed by 16% Asian, 12% African American, and 4% Hispanic, and 8% other.

Experimental Design & Set up

The research design was a randomized pretest-posttest true experimental design. The independent variable of this study was the level of immersion, with the dependent variables being the four constructs (Memory retention, learner performance, Engagement, Perceived Learning). Participants were randomly assigned to one of two groups: LiVR (Desktop Simulation) or HiVR (HMD-based Simulation). The LiVR condition completed the simulation scenarios on a desktop computer with a keyboard and mouse, while the HiVR condition utilized a VR headset device with VR hand controllers to complete the simulation.

Experimental Task

The experimental task was a virtual maintenance task that took place on a virtual airport tarmac. The simulation was a walk-around aviation maintenance training simulation designed to teach and evaluate the maintenance procedure of removing, inspecting, and replacing a tire on an aircraft (See Figure 1). Participants were tasked with utilizing specific tools, such as a tire jack, ratchet and multiple split-joint pliers to remove an aircraft tire. In the environment there was also a work bench containing all of the tools needed to complete the task. The workbench also contained an instructional clipboard which provided the participants with instructions presented in order, with the current step being highlighted. Participants were tasked with wiping all parts of the tire down, including the bolts, hub cap, the tire tube and other necessary objects related to the tire change. They then replaced the tire on the rim and fastened it back on to the aircraft using the necessary tools. The participants performed the task in either the HiVR environment, using a VR headset and VR hand controllers, or the LiVR environment, using a desktop monitor, a computer mouse, and a keyboard.

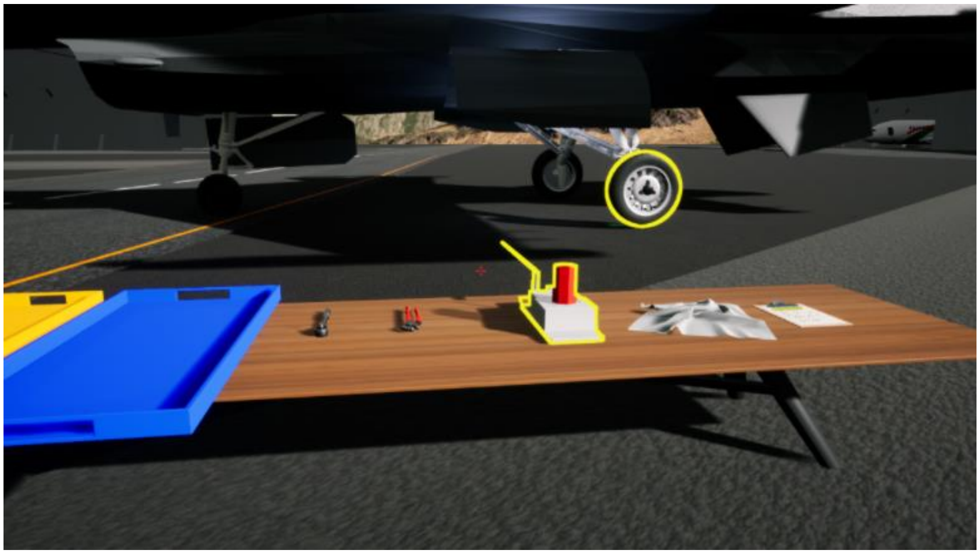

Figure 1. Participant Work Bench View in Either Condition

Experimental Setup

The simulation was a tire replacement training testbed, a simple testbed developed in Unreal Engine in collaboration with the Airforce Research Lab (AFRL)'s Gaming Research Integration for Learning Laboratory (GRILL). In the HiVR condition, this simulation was experienced with an Oculus Quest 2 VR headset and the Oculus Quest 2 VR haptic controllers, which were utilized to navigate and interact within the environment. In the LiVR condition, the simulation was experienced on a custom-built desktop computer equipped with an AMD Ryzen 7 3800X CPU, 32GB of RAM, an EVGA Geforce RTX 2080Ti graphics card, and a Windows 10 Home operating system. In this condition, the participants used a mouse and computer keyboard to navigate and interact with

Measures

Demographic data was collected via a researcher-developed online questionnaire administered to the participants prior to the start of the study. The data included: participants age, sex, ethnicity, country of origin, first language, degree program, experience with VR, experience with maintenance, and gaming experience and frequency.

Knowledge acquisition, the measure of learning performance, was assessed utilizing a researcher-developed pre and post multiple-choice knowledge quiz. The pre knowledge quiz was used prior to exposure to the simulation to establish initial group equivalency. The post-task quiz was used as an objective measure of learning performance in terms of knowledge acquisition. The knowledge assessment consisted of eight questions related to the maintenance task. For example, one question asked: "The small object located underneath the hub cap of the wheel is known as what?" and the participants were given three multiple choice options to choose from. Knowledge acquisition has been used as a measure of learner performance in previous immersion studies, (Guiterrez, et al., 2008).

The Cognitive, Affective, and Psychomotor (CAP) Perceived Learning Scale (Rovai et al., 2009) was used to measure the participants' perceived learning in each environment. The CAP Perceived Learning Scale has been utilized in several research studies (Maffei, 2020; Kizil, 2020; Aranyossy, & Kulcsar, 2020; Li, 2019). Rovai, Wighting, Baker and Grooms (2009) provide evidence of the instrument's validity and reliability to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual higher education classroom settings. The CAP Perceived Learning scale features three distinct subscales: cognitive, psychomotor learning, and affective. The instrument includes nine statements utilizing a seven-point Likert scale. For example, it asks students to rate statements such as: "I am able to use physical skills learned in this course outside of class" on the Likert scale ((1) Not at all to Very much so (7)).

Learner engagement was measured utilizing the User Engagement Scale (UES) short form. The UES is reliable measurement tool and has been utilized in many previous studies to measure engagement levels associated with various environments, some of which provide direct evidence to support the reliability and construct validity of the UES across multiple studies (Wiebe, et al, 2014; O'Brien, Cairns, & Hall, 2018; Holdener, Gut, & Angerer, 2020). The instrument includes nine statements, which are rated on a Likert scale from one to five, with one being "Strongly disagree", and five being "Strongly agree". The instruments include statements such as: "The time I spent in the VR simulation/Desktop simulation just slipped away".

Memory retention was measured by asking the participants to recall the initial steps of their first task. This was assessed using a researcher developed seven-questions order-of-operations quiz, requiring the participant to place the first seven steps of the virtual task in the correct sequential order. The participants gained a point for each step correctly placed. Participants also received partial credit for placing steps in correct sequential order, even if it was not in the correct linear sequence (e.g., if the answer was ABCD and they reported BCDA, points were still awarded for placing BCD sequentially).

PROCEDURES

Participants arrived at the experimental site, completed an informed consent and were administered the pre-study surveys which included the demographic questionnaire and the knowledge assessment. Participants then entered into a simulation familiarization session in their respective condition (i.e., HiVR or LiVR). The session was embedded within the simulation and explained how to utilize the controls to interact with objects and move around the virtual environment. The session lasted approximately five minutes. Each group then performed the same experimental task, immersed with the simulation using the technology associated with their assigned condition. The participants were instructed to carry out specific steps, using specific tools to complete an aircraft tire change. The instruction was provided to the participants in the simulation by a virtual instruction sheet with step-by-step guidance. The simulation scenario took approximately 10-15 minutes to complete. Following the simulation scenario, the participants were then given the post-task multiple-choice knowledge assessment, the memory retention questionnaire, the CAP Perceived Learning scale, and the UES student engagement questionnaire. Finally, the participants completed the task a second time in the simulation, with less instruction to aid them. This was achieved by removing a function of the simulation which highlighted certain objects needed to complete the task. The instruction sheet also omitted the specific name of the tool that was utilized in completing the task prior. For example: "Please remove the hub-cap bolts using the *ratchet* vs. Please remove the hub-cap bolts".

RESULTS

The following section presents the results of the study. Descriptive statistics are presented in this section for video game frequency, virtual reality experience, and maintenance familiarity. Descriptive statistics are also reported for the pre- and post-test knowledge assessment scores, engagement scores, perceived learning CAP scores, and memory retention scores.

Additionally, several inferential statistics are reported in this section. First, a MANOVA was conducted on several individual difference variables to ensure group equivalency on factors such as experience and familiarity with the maintenance task. Next, a repeated measures analysis of variance (ANOVA) was conducted on the pre- and post-training knowledge assessment to ensure group knowledge equivalency and evaluate the impacts of the two conditions on knowledge acquisition. Finally, a multivariate analysis of variance (MANOVA) was performed on the post-training user engagement scores, CAP scores, and memory retention scores to assess the impact of condition.

Demographics and Individual Differences

Participants video game frequency was distributed with: 12% playing daily, 24% playing weekly, 20% playing monthly, 20% playing yearly, and 24% never playing. Participants experience with VR was distributed with: 40% no experience, 16% less than six months experience, 4% one year to less than three years' experience, 28% three years to less than five years' experience, and 12% with five years or more experience. Maintenance familiarity was distributed with: 4% being very familiar, 48% being somewhat familiar, 24% being not very familiar, and 24% being not familiar at all. Twenty-four participants reported being either very proficient, or proficient in English, with one participant being somewhat proficient

A MANOVA was conducted with condition as a between group factor for individual difference factors of maintenance familiarity, video game frequency, and VR experience to ensure group equivalency. At the multivariate level, there were no main effects of condition, $F(6, 18) = 2.41$, $p > .05$, $\eta^2 = .45$. As a result, group equivalency was assumed.

Learning Performance (Knowledge Acquisition)

Descriptive statistics for the pre and post multiple-choice knowledge assessment scores by condition are included in Table 1. Three participants were missing data from the knowledge assessments due to incomplete assessment results. The descriptive statistics for the knowledge acquisition assessment showed very similar mean scores for the two conditions on the pre-training assessment, suggesting group equivalency (See Table 1). There was a substantial improvement in scores for both groups with the HiVR condition having a slightly higher post-training score. A repeated measures ANOVA was performed on the Pre/Post knowledge assessment scores, with condition as a between-groups variable and maintenance familiarity as a covariate. The results of the repeated measures ANOVA revealed a significant main effect of trial, $F(1, 20) = 44.49$, $p < .001$, $\eta^2 = .69$. There was not, however, a statistically significant interaction between trial and condition, $F(1, 20) = .435$, $p = .52$, $\eta^2 = .02$

Table 1. Pre/Post Knowledge Assessment Descriptive Statistics

	HiVR (VR)			LiVR (Desktop)		
	N	*M*	*SD*	*N*	*M*	*SD*
Pre-Knowledge Assessment	10	27.20	12.97	12	28.41	16.47
Post-Knowledge Assessment	10	64.70	16.42	12	59.16	19.33

Post Training Perceived Learning, Engagement Memory Retention

Descriptive statistics for the dependent variables of perceived learning, engagement, and memory retention are presented in Table 2. Two participants were removed because of missing data due to incomplete survey results. To determine if any of these differences were statistically significant, A MANOVA was conducted to examine the impact of condition on the three dependent variables of perceived learning, engagement, and memory retention. The multivariate results revealed a significant effect of condition, $F(3,19) = 6.46$, $p <. 01$, $\eta^2 = .51$. The univariate results indicated that there was no significant main effect of condition on the perceived learning scores $F(1, 21) = .048$, $p = .81$, $\eta^2 = .002$, but there was a statistically significant main effect of condition on engagement levels, $F(1, 21) = 9.27$, $p = .01$, $\eta^2 = .31$. There was a trend observed with respect to memory retention and HiVR, but it did not reach statistical significance, $F(1, 21) = 3.66$, $p = .07$, $\eta^2 = .15$.

Table 2. Post- Task Variable Descriptive Statistics

	HiVR (VR)			LiVR (Desktop)		
	N	*M*	*SD*	*N*	*M*	*SD*
CAP Perceived Learning Score	11	32.09	7.36	12	31.25	10.60
UES Engagement Score	11	49.73	4.49	12	42.50	6.58
Memory Retention	11	4.36	.80	12	3.25	1.76

DISCUSSION

The following section addresses and discuss the results presented and implications of these findings. Practical and theoretical implications are delineated for each analyzed variable. Limitations, recommendations, and domains of future research are addressed and discussed.

Knowledge Acquisition

The results of the repeated measures ANOVA on the pre and post training knowledge acquisition scores indicated that in both conditions, VR and Desktop, significant learning occurred. This pattern of results is consistent with the previously presented literature, in that both groups benefited from the immersive simulation (Guiterrez et al., 2008; Qian et al. 2020). However, participants in both conditions improved their scores from pre- to post-test and there was not a statistically significant interaction between trial and condition. This does not align with previous literature that has shown that HiVR leads to higher levels of knowledge acquisition. However, our findings could have been the result of a shorter time spent learning in the simulation than in prior studies, and the focus on a more spatial task.

Although these results do not have statistical significance, the findings could suggest practical significance when observing the 6-point difference between post scores. The average score of the desktop condition was a 59% whereas the VR condition rendered an average of 64%. Although not statistically significant, these two scores are the difference between passing and failing a course, an exam, or an assignment. This practical implication may suggest that for hands-on vocational maintenance training, using the highest levels of immersion may still yield beneficial results in terms of knowledge acquisition. However, further research is required to identify if this is a repeatable practical trend

Perceived Learning

The CAP perceived learning scale was used to subjectively assess participants' perceived learning following training in either the VR or Desktop condition. The results of the MANOVA unveiled that both the VR and Desktop conditions perceived a moderate amount of learning during their time in the simulation, with both conditions hovering around a score of 32 out of 54 (higher scores indicate higher perceived learning). Although both groups perceived to have learned a moderate amount, there was effectively no difference between the two conditions.

Literature on perceived learning in solely HiVR compared to LiVR is scarcely available and is a new area of investigation empirically (Zhang et al. 2017). Many studies have cited using AR or computer systems compared to traditional training methods to evaluate participants' perceived learning and found significantly higher perceived learning in the more immersive conditions (Rockinson-Szapkiw, et al. 2013; Zhang, 2017). Whereas the available research has found that higher levels of immersion have led to higher levels of perceived learning, the present study has shown no significant differences among the Desktop and VR conditions. In the absence of significance between conditions, it is important to note that both conditions yielded moderate amounts of perceived learning in the maintenance training environment

Engagement

The MANOVA conducted on the engagement scores yielded results that are consistent with prior findings, in which higher levels of immersion rendered higher levels of engagement (Allcoat & Muhlenen, 2018; Allcoat, et al., 2021). The participants in the HiVR condition reported significantly higher engagement levels than participants in the LiVR condition.

These results are a direct demonstration of an important benefit in utilizing such devices for learning and training. The outcome of this study aligns with current research conclusions in closely related hands-on training (Flavian, Sanchez & Orus, 2020). This increased engagement is cited in part due to the heightened integration between the simulation and human through higher levels of immersion, and increased presence in the environment (Flavian, Sanchez & Orus, 2020). From a practical standpoint, as more studies achieve similar findings, VR testbeds and environments should be viewed as viable options to provide engaging learning experiences to students and trainees in an aviation-based maintenance field, as well as other fields that require hands-on training. These results represent an initial step towards facilitating virtual maintenance training in the future and can be viewed as an economically efficient asset in promoting engagement in specific maintenance training procedures

Memory Retention

The results of the MANOVA revealed that the difference between participants' ability to recall the correct order of steps after training in the Desktop condition and the VR condition was not significance at the $p = .05$ level. The effect size associated with this analysis was $\eta^2 = .15$, indicating a large effect and that lack of significance might be due to the small number of participants. The results might have achieved significance had there been a larger number of participants. These findings do not align with current research conclusions, in which many studies have identified that higher levels of immersion equate to significantly higher memory retention levels. (Krokos, Plaisant, & Varshney, 2018; Babu, et al., 2020).

Anecdotally, the spatial aspects of the task demonstrated to be an area that aided the participants in memory recall. Many participants, while performing the memory retention test, verbally stated that they remembered moving their head or hand in a certain direction during a specific part of the task in the VR condition, which alluded to the importance of spatial awareness in remembering the steps (Krokos, Plaisant, & Varshney, 2018). A testbed that renders higher levels of immersion has been shown by previously mentioned studies to increase subject's memory and should be considered as an effective option for maintenance training testbeds in regard to hands-on spatial training methods (Krokos, Plaisant, & Varshney, 2018; Pauch, Proffit & Williams, 1997).

LIMITATIONS AND FUTURE RESEARCH

This study evaluated the effects of two different levels of immersion on specific learning constructs in an aviation maintenance training task. The study is one of the few studies that compares the highest level of immersion currently available to a lower level of desktop 2D immersion. The study focused on hands-on aviation maintenance training and learning variables which is a grouping that is not widely found in the existing literature.

The study faced limitations that effected the results. First, multiple limitations resulted from the simulation development timeline. The environment was developed from scratch, and the development timeline bled into the data collection timeline, limiting the time available to collect data and the number of participants. Further, because the development timeline had to be cut short in order to start data collection, there were some simulation glitches and errors that were not fixed prior to data collection. These issues presented during the training transfer task in which the participants had to perform the task with limited instructions. As a result, the learner performance data, in the form of skill acquisition resulting from this portion of the experiment, could not be utilized. When the study was proposed, learner performance was to be measured utilizing two variables: Knowledge acquisition and Skill acquisition. Skill acquisition was going to be measured by comparing the errors and time to completion within subjects during their first and second run through. Due to simulation malfunctions occurring more often than not in the transfer task, no viable data could be collected to analyze.

Future research would benefit from eliminating the previously mentioned limitations, which would provide more generalizable results and the opportunity for further exploration of learner performance in terms of skill acquisition in both LiVR and HiVR. It would also be useful to extend the current findings by examining the prior metrics and variables across three dimensions of immersion: Traditional, LiVR, and HiVR in a hands-on maintenance training setting within either aviation, industrial or mechanical industries. Future research should also consider conducting follow-up research on transfer of training to a real-world training event or scenario.

ACKNOWLEDGEMENTS

We would like to thank the Air Force Research Lab's Gaming Research Integration for Learning Laboratory (GRILL) for developing the aviation maintenance simulator testbed that was utilized to complete this research.

REFERENCES

Allcoat, D., Hatchard, T., & Azmat, F. (2021). Education in the digital age:learning experience in virtual and mixed realities. *Journal of Educational Computing Research*. https://doi.org/10.1177/0735633120985120

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

Babu, S., Krishna, S., Unnikrishnan, R., & Bhavani, R. (2018). Virtual reality learning environments for vocational education: A comparison study with conventional instructional media on knowledge retention. *IEEE 18th International Conference on Advanced Learning Technologies*.

Chapman, P. (1997). Models of engagement: Intrinsically motivated interaction with multimedia learning software. Unpublished master's thesis, University of Waterloo, Waterloo, Canada.

Domingo, J., & Bradley, E. (2017). Education student perceptions of virtual reality as a learning tool. *Journal of Educational Technology Systems.* https://doi.org/10.1177/0047239517736873

Flavian, C., Ibanez-Sanchez, S., & Orus, C. (2021). Impacts of technological embodiment through potential guests' emotions and engagement. *Journal of Hospitality Marketing & Management*, 30(1).

Gavish, N., Gutierrez, T., Webel, S., Rodriguez, J., Peveri, M., Bockholt, U., Tecchia, F. (2013). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments, 23*(6). https://doi.org/10.1080/10494820.2013.815221

Girvan, C. (2018). What is a virtual world? Definition and classification. *Educational Technology Research and Development*, 66, 1087-1100.

Hassenzahl, M., & Tractinsky, N. (2006). User experience- A research agenda. *Behaviour and Information Technology*, 25(2), 91-97.

Holdener, M., Gut, A., & Angerer, A. (2020). Applicability of the user engagement scale to mobile health: A survey-based quantitative study. *National Library of Medicine*, 3(8).

Jaynes, J., (1976). The origin of consciousness in the breakdown of the bicameral mind. *Houghton Mifflin Company*.

Rakowski, R., & Gruber, A. (2019). Low-Immersion versus High-Immersion Virtual Reality: Definitions, Classifications, and Examples with a Foreign Language Focus. *12th International Conference Innovation in Language Learning*, 552-555. DOI: 10.26352/DY14_2384-9509

Rockinson-Szapkiw, A., Holmes, J., & Stephen, J. (2019). Identifying Significant Personal and Program Factors that Predict Online EdD Students' Program Integration. *Online Learning*, *23*. https://doi.org/10.24059/olj.v23i4.1579

Krokos, E., Plaisant, C., & Varshney A. (2019). Virtual memory palaces: immersion aids recall. *Virtual Reality*, 23, 1-15.

Laurel, B. (1993). Computers as theatre. Reading, MA: Addison-Wesley.

Lee, E., Wong, K., & Fung, C. (2009). Learning effectiveness in a desktop virtual reality-based learning environment. *Proceedings of the 17th International Conference on Computers in Education*, 832-839.

Maffei, J. (2020). The effect of augmented reality on learning in the mathematics classroom. *Liberty University.*

Mandal, S. (2013). Brief Introduction of Virtual Reality & its Challenges. International Journal of Scientific & Engineering Research, 4(4), 304-309.

Mao, C.C., & Chen, F. Y. (2019). Augmented reality and 3-D visual-ization effects to enhance battlefield situational awareness. *Human interaction and emerging technologies, systems and computing*, 1018. 300–303.

Markansky, G., Terkildsen, T. S., & Mayer, R. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 225-236.

O'Brien, H., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112, 28-39.

O'Brien, H., & Toms, E., (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938-955. https://doi.org/10.1002/asi.20801

Pan, X., & Hamilton, A. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(9).

Pausch, R., Proffit, D., & Williams, G. (1997). Quantifying immersion in virtual reality. *Proceedings of the 24th annual conference on Computer Graphics and Interactive Techniques*, 13-18. https://doi.org/10.1145/258734.258744

Qian, J., Ma, Y., Pan, Z., & Yang, X. (2020). Effects of virtual-real fusion on immersion, presence and learning performance in laboratory education. *Virtual Reality & Intelligent Hardware*, 2(6), 569-584.

Renganayagalu, S., Mallam, S., & Nazir, S. (2021). Effectiveness of VR head mounted displays in professional training a systematic review. *Technology, Knowledge, and Learning*, 26, 999-1041.

Speicher, M., Hall, B., & Nebeling, M. (2019). What is Mixed Reality? *Proceedings of the 2019 CHI Conference on Human Factors in Computing*, 537, 1-15.

Trindade, J., Fiolhais, C., & Almeida, L. (2002). Science learning in virtual environments: A descriptive study. *British Journal of Educational Technology*, 33(4). 471-488. https://doi.org/10.1111/1467-8535.00283

Wiebe, E., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, 32, 123-132.

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F., Li, C., Li, W.,Huang, H., Akdere, M., Mousas C., Yu, L. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*.

Yates, F. (1992) The art of memory. *Random House,* 64. New York

Zhang, X., Jiang, S, Pablos, P., Lytras, M., & Sun, Y. (2017). How virtual reality affects perceived learning effectiveness: a task-technology fit perspective. *Behaviour andInformation Technology*, 36(5), 1-9. DOI: 10.1080/0144929X.2016.1268647