Comparison of Augmented and Virtual Reality Training for Spatial Anatomy

Andrew Wismer, Lauren Reinerman-Jones, Grace Teo,
Sasha Willis, Kelsey McCracken
University of Central Florida
Institute for Simulation and Training (UCF IST)
Orlando, FL
awismer@ist.ucf.edu, lreinerm@ist.ucf.edu, gteo@ist.ucf.edu,
swillis@ist.ucf.edu, kmccrack@ist.ucf.edu

Matthew Hackett
Army Research Laboratory
Orlando, FL
matthew.g.hackett.civ@mail.mil

ABSTRACT

On a spectrum from fully real to fully virtual environments, augmented reality (AR) systems allow a user to visualize three-dimensional objects. AR combines elements of real and virtual environments allowing the user to maintain a sense of presence in the real environment while manipulating virtual objects. Virtual reality (VR) immerses the user into a virtual environment, typically through the use of a head-mounted display (HMD). As consumer access to AR/VR has increased over the past few years, so have the potential applications for these systems to enhance training and education. Anatomy is one such field that requires a thorough understanding of difficult, spatial relationships, and serves as the knowledge foundation for medical practitioners. Some work has been done with newer AR or VR systems for anatomical training, but a comparison between the two systems for (i) anatomical knowledge acquisition, and (ii) the workload incurred by the learner is lacking. The present work provides a comparison between AR (i.e., zSpace) and VR (i.e., HTC Vive) for the learning of macroscopic brain anatomy. A suite of objective (pre/post-learning tests), subjective (surveys), and physiological (EEG) measures were used to provide a comprehensive evaluation of AR and VR systems for anatomical training. Results suggest a high degree of similarity between AR and VR, yet hint at some differences in the associated cognitive processes. These results inform selection of the most beneficial training platform for anatomical knowledge acquisition in first-time learners and provide insight into alternative implementations for these emerging technologies.

ABOUT THE AUTHORS

Andrew Wismer, Ph.D., is a Researcher in Experimental Psychology at the University of Central Florida's Institute for Simulation and Training. He is also a Human Factors Researcher at HP Inc. His research involves assessing learning in dynamic environments.

Lauren Reinerman-Jones, Ph.D., is the Director of Prodigy, which is one lab at the University of Central Florida's Institute for Simulation and Training, focusing on assessment for explaining, predicting, and improving human performance and systems.

Grace Teo, Ph.D., is a Research Associate at the University of Central Florida's Institute for Simulation and Training. Her research includes work on decision making, individual differences, human-technology interactions, and various assessments of human performance.

Sasha Willis, B.S., is a former Research Assistant at the University of Central Florida's Institute for Simulation and Training and a HF/IO doctoral student at Wright State University with research interests in visual attention, game-based learning, and three-dimensional imaging technologies.

Kelsey McCracken is an Undergraduate Research Assistant at the University of Central Florida's Institute for Simulation and Training and is currently pursuing higher education in psychology.

Matthew Hackett, Ph.D., is a Science and Technology Manager with the US Army Research Laboratory Medical Simulation and Performance Branch, conducting research in a variety of technologies to improve military medical training.

Comparison of Augmented and Virtual Reality Training for Spatial Anatomy

Andrew Wismer, Lauren Reinerman-Jones, Grace Teo,
Sasha Willis, Kelsey McCracken
University of Central Florida
Institute for Simulation and Training (UCF IST)
Orlando, FL
awismer@ist.ucf.edu, lreinerm@ist.ucf.edu, gteo@ist.ucf.edu,
swillis@ist.ucf.edu, kmccrack@ist.ucf.edu

Matthew Hackett
Army Research Laboratory
Orlando, FL
matthew.g.hackett.civ@mail.mil

INTRODUCTION

Virtual and augmented reality systems have been shown to effectively supplement traditional methods of anatomical instruction, such as dissection (Habbal, 2009; Ghosh, 2017; Brewer, Wilson, Eagleson, & de Ribaupierre, 2012; Moro, Stromberga, Raikos, & Stirling, 2017; Khot, Quinlan, Norman, & Wainman, 2013). With the emergence of commercially available augmented reality (AR) and virtual reality (VR) systems, new methods for supplementing training in fields such as anatomy and physiology become available (Skochelak, 2010). It is important that these new technologies be evaluated with respect to their suitability for training on knowledge acquisition and impact on cognitive load and user satisfaction. However, limited studies have directly compared virtual and augmented reality systems with respect to anatomical knowledge acquisition (Moro et al., 2017). The present work provides a comparison of two commercially-available systems for use in anatomical training.

The purpose of this experiment is to provide insight into the selection of different emerging three-dimensional visualization technologies (3DVTs) for training in medical domains where the visualization of three-dimensional information is of central importance. The AR system used in the present work is the zSpace system, comprising an all-in-one computer with two tracking sensors, 3D glasses, and a stylus with six degrees of freedom for interaction (zSpace, Inc., 2018). The VR system used is the HTC Vive system, which includes two controllers and a head-mounted display for full immersion through SteamVR software (HTC Corporation, 2018). These systems are evaluated with respect to a wide range of outcomes, including knowledge acquisition, cognitive load/workload, simulation sickness, presence, stress, and usability.

VR and AR systems differ on a number of characteristics, such as the method of augmentation, accommodation, stereoscopy, resolution, and latency (Kiyokawa, 2008). These characteristics were not developed to be configurable or controlled by the user. Instead, current emerging technologies were evaluated with respect to their capability of supporting anatomical knowledge acquisition through actions afforded by each system. This design allowed us to ask the following questions: i) Do current VR and AR technologies provide equal support for the acquisition of anatomical knowledge? ii) Do VR and AR systems affect cognitive load and other related learner outcomes differently?

Many similarities and some differences between VR and AR technologies have been found when they were applied towards anatomical knowledge acquisition. Both VR and AR have been associated with increased immersion and engagement, and users of either showed similar levels of knowledge gain as compared to users of traditional learning methods (Hu, Wilson, Ladak, Haase, & Fung, 2009; Keedy, et al., 2011; Moro et al., 2017). However, VR has been associated with increased adverse effects, such as headaches, dizziness, or blurred vision (Treleaven, et al., 2015; Munafo, Diedrick, & Stoffregen, 2017; Moro et al., 2017). The present work provides a replication and extension of work such as that by Moro et al. (2017), through the inclusion of a multi-faceted spatial anatomy test, various subjective measures, and a physiological measure of workload.

METHODS

Participants

Participants were students at the University of Central Florida and received course credit in exchange for their participation. In total, there were 65 participants (30 female, 35 male). All participants provided written consent prior to participation, were at least 18 years of age, and had normal, or corrected-to-normal, vision. There were no significant differences between participants with respect to the demographic variables listed in Table 1.

Table 1. Participant demographics by 3DVT condition

		VR ***n* = 32**	**AR** ***n* = 33**	**Total** ***N* = 65**
	Age *M* (*SD*)	18.59 (1.21)	19.36 (1.90)	18.98 (1.63)
Gender (Count and %)	Males	13 (40.6%)	22 (66.7%)	35 (53.8%)
	Females	19 (59.4%)	11 (33.3%)	30 (46.2%)
Number of Anatomy Courses (Count and %)	0	20 (62.5%)	16 (48.5%)	36 (55.4%)
	1	9 (28.1%)	10 (30.3%)	19 (29.2%)
	2	3 (9.4%)	5 (15.2%)	8 (12.3%)
	3	0 (0.0%)	2 (6.0%)	2 (3.0%)
	*Self-Rated Anatomy Pre-Knowledge *M* (*SD*)	4.28 (1.90)	4.15 (1.50)	4.22 (1.70)

* Note. Participants were asked to rate their knowledge of the parts of the human brain on a scale from 1 (low) to 10 (high).

Experimental Design

Participants were assigned to one of two conditions (virtual reality [VR] or augmented reality [AR]) based on the three-dimensional visualization technology (3DVT) they would interact with during the learning phase. The experimental design consisted of three phases: (i) pre-task surveys, (ii) a guided training session followed by the learning task (with either VR or AR), and (iii) post-task surveys. The only difference between conditions was the 3DVT used to study brain anatomy in the learning task.

Materials

Three-Dimensional Visualization Technologies (3DVTs)

Virtual Reality (VR) system

The virtual reality condition utilized the HTC Vive system (see Figure 1). Equipment required for this system included the following: a headset that the user wore to display the brain models in virtual reality, two controllers (one for each hand), two infrared receivers that were positioned to track the headset and controllers and project the user into the virtual reality, and a computer equipped with a sufficiently powerful graphics card.

Figure 1. Virtual reality system (HTC Vive) and display

Augmented Reality (AR) system
As for the augmented reality condition, the zSpace all-in-one (AIO) system was used. The following items were part of the zSpace system: a zSpace all-in-one computer, a stylus pen that allowed the user to manipulate objects within the program, and glasses for the user to wear in order to see objects displayed in three dimensions. These glasses contain tracking elements that allow the zSpace system to generate the appropriate stereoscopic perspective to the viewer, resulting in the spatially realistic image displayed to the user (see Figure 2 below).

Figure 2. zSpace AIO augmented reality system and materials. Note – the picture shows the typical 2D presentation of the model, whereas the participant wearing the glasses was able to see the model in 3D.

Surveys

Simulation Sickness Questionnaire (SSQ)
Participants completed the Simulation Sickness Questionnaire (Kennedy et al., 1993) at the beginning of the experiment and immediately following the learning task. The SSQ consists of 16 questions relating to three factors of simulation sickness: Nausea, Oculomotor, and Disorientation. The survey asked participants to assess the extent to which different symptoms of simulation sickness were affecting them at the given moment. Ratings were made on a four-point Likert Scale from 0 (not at all) to 3 (very much).

Short Dundee Stress State Questionnaire (DSSQ)
A shortened version of the Dundee Stress Scale Questionnaire (Matthews, Emo, & Funke, 2005; derived from Matthews et al., 2013) was completed in both the pre- and post-task phases. The pre-DSSQ assessed participants' stress level prior to interacting with the 3DVT, while the post-DSSQ assessed any changes in stress levels following usage of the 3DVT to study brain anatomy. Pre- and post-DSSQ surveys were identical other than differences in verb tense. There were 30 questions with responses given on a five-point Likert Scale, from 0 (definitely false) to 4 (definitely true). The DSSQ measures three factors of stress: Distress, Engagement, and Worry.

Spatial Anatomy Test

Spatial knowledge acquisition was assessed through a novel pre-post-Spatial Anatomy Test constructed by the research team. Three facets of knowledge were assessed: (i) factual identification (16 questions), (ii) spatial relations multiple choice (15 questions), and (iii) mental rotation ability (4 questions). Identification questions tested the participant's ability to locate and identify different anatomical structures in the human brain. Participants were asked to label highlighted brain structures by selecting a response from a list comprising 16 targets (items studied in the learning task) and 16 distractors (items not studied in the learning task). Spatial multiple choice questions, each having four options, required the participant to apply their understanding of the location and inter-relationships of anatomical structures. Finally, mental rotation questions assessed the participant's ability to compare four slightly altered images (simply rotated or rotated and mirrored) of brain structures to a target image of the same structure.

Following each mental rotation question, participants were asked to rate their confidence in their answers on a five-point Likert scale (ranging from very low to very high). At the end of the Spatial Anatomy Test, participants were prompted to provide two additional confidence ratings pertaining to their perceived performance on the test, assessing overall confidence in their responses to the identification and spatial multiple choice questions. The time in which it took the participants to complete the Spatial Anatomy Test was also recorded and used for later analyses. Each of these measures – accuracy, confidence, and completion time – was coded by question type (identification, spatial multiple choice, mental rotation), with overall values provided for accuracy and completion time.

System Usability Scale (SUS)

A System Usability Scale (Brooke, 1996) was administered to provide insight into the usability of the given 3DVT. Participants responded to 10 questions regarding the usability of their assigned 3DVT for learning spatial brain anatomy. Participants responded on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), and responses were summed and doubled to provide an overall score out of 100.

NASA Task-Load Index (NASA-TLX)

The participants' perceived workload, measured with a NASA-TLX questionnaire (Hart & Staveland, 1988), was also assessed in the post-task phase. Participants rated their perceived workload (on scales from 0-100, in five-point increments) along six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. An overall average level of workload (i.e., Global Workload) was also calculated from these scores.

Presence Questionnaire (PQ)

A 24-item Presence Questionnaire (Witmer & Singer ,1998) was utilized in the post-task phase to assess a participant's subjective experience with the virtual or augmented reality learning environment with respect to Realism, Possibility to Act, Quality of Interface, Possibility to Examine, Self-Evaluation of Performance, Sounds, and Degree of Haptic Sensation. Scores on each dimension were standardized to facilitate comparison across PQ dimensions.

Electroencephalography (EEG)

Prior to the pre-task phase, participants were fitted with a B-Alert X10 electroencephalograph (EEG) cap, made by Advanced Brain Monitoring to enable the EEG recording. EEG measures voltage fluctuations in the brain and provides a description of patterns of brain activation during learning. The EEG cap had nine sensors which were placed directly on the scalp on the participant. These measured electrical activity in different regions of the brain (i.e., Frontal, Occipital, and Parietal) across three spectral power bands (i.e., Alpha, Beta, and Theta) associated with workload. EEG was measured continuously throughout the experiment.

Procedure

Participants were met in the lobby and directed to the research lab. All participants provided written informed consent prior to participation. They then completed a restrictions checklist and were excluded if they reported having non-normal vision or if they were of a vulnerable population (i.e., pregnant, prisoner, or under 18 years old).

Next, the researcher measured the participant's head to determine the correct size EEG cap to use (small or medium). The participant was then asked to complete a demographics survey while researchers prepared the appropriate EEG cap. Upon completion, the participant was fitted with the EEG cap and an impedance check was conducted to ensure that all of the nine sensors and reference electrodes (affixed to the mastoid bones) had a clear connection. After acquiring and confirming a clear EEG connection, a five-minute sitting baseline measurement was conducted. Participants then completed a series of pre-task surveys (pre-SSQ, pre-DSSQ, pre-Spatial Anatomy Test, CCT).

Following the pre-task surveys, participants were trained on the use of their assigned 3DVT. Participants in the VR condition were fitted with the head-mounted display (HMD), and another impedance check was run to ensure stability of the EEG signals during the set up. In addition, a second baseline was conducted with the participant standing, since they would be standing throughout the learning task. Participants in the AR condition required no extra baseline or impedance check. The 3D glasses could be placed on the participant's face (over prescription glasses if needed) without affecting the EEG cap. A training scenario was provided in both VR and AR conditions to introduce the participant to the technology and ensure they could properly operate the system without assistance for the learning task (see Figure 3 below). The training scenario was constructed with colored 3D shapes to represent the brain models used in the learning task.

Participants then moved on to the learning task and were given up to 10 minutes to study the models using their given technology. The models they studied included 16 different brain structures, demonstrated on the models using both internal and external views of brain. In the VR condition, two models (one external, one internal) were present in the environment and could be viewed individually or together. In the AR condition, three separate models (two external, one internal) were available to view one at a time. In both conditions, labels could be toggled on and off, models could be reset to the starting location, and a timer counting down from 10 minutes was within view. Participants had the option to press a button to quit the learning task before the end of the ten minutes if they felt confident to continue. After the learning task, the VR headset was removed (or AR glasses were removed). In the VR condition only, another impedance check was run to ensure that the EEG sensors were still adequately connected.

Virtual Reality *Augmented Reality*

Figure 3. Training scenarios for VR and AR conditions

Finally, participants completed a series of post-task surveys (post-SSQ, NASA-TLX, post-DSSQ, post-Spatial Anatomy Test, SUS, and PQ). After the participant completed the post-task surveys, the EEG cap was removed, they were thanked for their participation, awarded course credit, and subsequently dismissed from the experiment.

RESULTS

Pre-Task Differences

First, independent-samples *t*-tests were conducted on each measure and question type administered during pre-Spatial Anatomy Test to determine if there were any pre-existing knowledge differences between the AR and VR groups. Indeed, the two groups did not differ on any measure of accuracy (all p's > .192), confidence (all p's > .555), or completion time (all p's > .648). This suggests the two groups were equivalent in terms of pre-existing anatomical knowledge.

Pre-existing differences between conditions with respect to stress and simulation sickness symptoms were assessed. Independent-samples *t*-tests conducted on DSSQ scores for each of the three subscales (i.e., Distress, Engagement, and Worry) revealed no differences between the two groups on pre-task DSSQ scores for Distress, $t(63) = -0.35$, p =

.726, Engagement, $t(63) = 0.10$, $p = .924$, or Worry, $t(63) = 0.04$, $p = .966$. Similarly, independent-samples *t*-tests conducted on each of the three SSQ subscales (i.e., Nausea, Oculomotor, and Disorientation) revealed no significant difference between groups for Nausea, $t(62) = -1.55$, $p = .126$; Oculomotor, $t(62) = -1.22$, $p = .228$; or Disorientation, $t(62) = -1.73$, $p = .089$.

Spatial Anatomy Test

Next, Spatial Anatomy Test results were analyzed using separate 2 (testing phase: pre, post) x 2 (3DVT: VR, AR) mixed ANOVAs, with 3DVT (VR, AR) as the between-subjects variable and testing phase (pre, post) as the within-subjects variable, for each measure of average accuracy, confidence, and completion time. Confidence level for the mental rotation questions was computed as an average, while confidence for the identification and multiple choice questions was computed as an overall total.

Evidence of learning can be seen in an increase in accuracy and confidence at post-task (see main effect of testing phase on confidence in Table 2 and main effect of testing on accuracy in Figure 4). Evidence of learning is also apparent in the change in completion times also seen in Figure 4, suggesting increased deliberation or retrieval times at post-test. Critically, the increase in spatial anatomy knowledge did not vary by condition (i.e., no significant interaction between testing phase and 3DVT condition on accuracy, confidence, or completion time; all p's > .729, all η_p^2's < .01). This suggests VR and AR afforded similar levels of knowledge acquisition.

Table 2. Main effect of testing phase (pre, post) on Spatial Anatomy Test confidence by question type

Measure	Question Type	Pre-Task *M* (*SD*)	Post-Task *M* (*SD*)	Result
Confidence	**Identification**	1.77 (0.93)	3.66 (0.94)	$F(1, 63) = 146.23$, $p < .001$, $\eta_p^2 = .70$
	Multiple choice	1.71 (0.82)	3.55 (0.98)	$F(1, 63) = 165.39$, $p < .001$, $\eta_p^2 = .72$
	Mental rotation	3.13 (0.91)	3.42 (0.83)	$F(1, 63) = 12.68$, $p = .001$, $\eta_p^2 = .17$

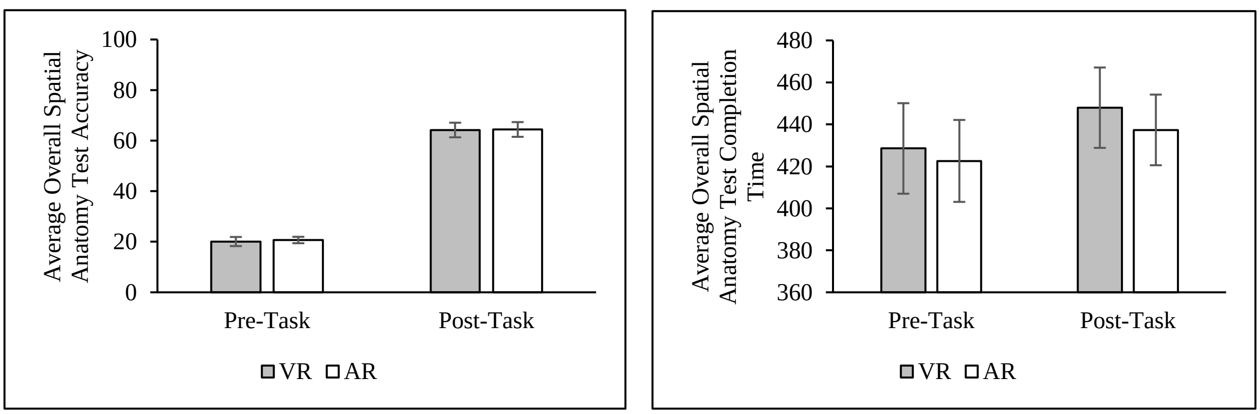

Figure 4. Average overall accuracy and completion time on the Spatial Anatomy Test.

Usability (SUS)

There was no significant difference in average overall usability scores between VR ($M = 81.00$, $SD = 16.30$) and AR ($M = 83.30$, $SD = 14.30$) conditions, $t(63) = -0.61$, $p = .547$. Analysis of individual SUS questions revealed a main effect of 3DVT for question 9 ("I felt very confident using the system") only, $t(51) = -2.44$, $p = .018$, such that participants in the AR condition ($M = 4.76$, $SD = 0.44$) responded with higher confidence in using the system than participants in the VR condition ($M = 4.34$, $SD = 0.70$). While both overall usability and average confidence were high for VR and AR groups, results suggest that VR may benefit from slightly longer training to increase confidence in using the system.

Stress (DSSQ)

Separate 2 (testing phase: pre, post) x 2 (3DVT: VR, AR) mixed ANOVAs were conducted for the Distress, Engagement, and Worry subscales of the DSSQ to analyze possible differences between 3DVT conditions (see Table 3). There was a main effect of testing phase on average Distress scores, $F(1, 63) = 7.31$, $p = .009$, $\eta_p^2 = .10$, such that pre-task distress ($M = 7.95$, $SD = 4.57$) was significantly higher than post-task distress ($M = 6.60$, $SD = 5.20$). Similarly, there was a main effect of testing phase on average Worry scores, $F(1, 63) = 66.23$, $p < .001$, $\eta_p^2 = .51$, such that pre-task Worry ($M = 13.31$, $SD = 6.68$) was significantly higher than post-task Worry ($M = 7.38$, $SD = 5.44$). For Engagement scores, there was also a main effect of testing phase on average Engagement, $F(1, 63) = 63.30$, $p < .001$, $\eta_p^2 = .50$, though an opposite trend was seen such that average post-task Engagement scores ($M = 24.65$, $SD = 4.98$) were significantly greater than average pre-task Engagement scores ($M = 19.88$, $SD = 4.99$). There was no main effect of 3DVT on Distress, $F(1, 63) = 0.09$, $p = .768$, $\eta_p^2 < .01$, Engagement, $F(1, 63) = 0.05$, $p = .817$, $\eta_p^2 < .01$, or Worry scores, $F(1, 63) = 0.09$, $p = .763$, $\eta_p^2 < .01$, nor was there a significant interaction between testing phase and 3DVT on Distress, $F(1, 63) = 0.02$, $p = .887$, $\eta_p^2 < .01$, Engagement, $F(1, 63) = 0.05$, $p = .823$, $\eta_p^2 < .01$, or Worry, $F(1, 63) = 0.43$, $p = .516$, $\eta_p^2 < .01$.

Table 3. Average DSSQ scores by 3DVT condition and testing time

Measures	VR *M* (*SD*)	AR *M* (*SD*)
Pre-Distress	7.75 (4.24)	8.15 (4.93)
Post-Distress	6.47 (5.70)	6.73 (4.75)
Pre-Engagement	19.94 (5.00)	19.82 (5.06)
Post-Engagement	24.84 (4.76)	24.45 (5.26)
Pre-Worry	13.34 (6.69)	13.27 (6.77)
Post-Worry	6.94 (4.94)	7.82 (5.92)

Workload (NASA-TLX)

Separate independent-samples *t*-tests were used to analyze any differences between 3DVT condition for NASA-TLX overall (i.e., global workload) values and for each of the six subscale values. A main effect of 3DVT was found for Effort scores, $t(63) = -2.30$, $p = .025$, such that the VR group ($M = 47.50$, $SD = 23.18$) had significantly lower scores than the AR group ($M = 60.61$, $SD = 22.84$). No other significant differences were found (see Table 4 below). As a whole, workload was found to be low to moderate.

Table 4. Average NASA-TLX scores by 3DVT condition

Measure	VR *M* (*SD*)	AR *M* (*SD*)	*p*-values
Global Workload	37.19 (15.71)	44.12 (16.36)	$p = .087$
Mental Demand	55.94 (20.73)	63.03 (27.61)	$p = .247$
Physical Demand	26.72 (22.17)	28.48 (24.61)	$p = .762$
Temporal Demand	43.44 (21.34)	46.52 (25.91)	$p = .604$
Effort	47.50 (23.18)	60.61 (22.84)	**$p = .025$**
Frustration	20.47 (20.25)	24.24 (23.92)	$p = .496$
Performance	29.06 (25.64)	41.82 (30.69)	$p = .074$

Simulation Sickness (SSQ)

A set of 2 x 2 mixed ANOVAs were conducted on SSQ subscale scores, with testing phase (pre, post) as the within-subjects variable and 3DVT (VR, AR) as the between-subjects variable, to examine differences between groups across the two testing times. There were main effects of testing phase on Oculomotor, $F(1, 62) = 0.89$, $p = .350$, $\eta_p^2 = .01$, and Disorientation, $F(1, 62) = 6.39$, $p = .014$, $\eta_p^2 = .09$, with increased simulation sickness symptoms at post-test. There was no main effect of testing phase on Nausea, $F(1, 62) = 0.89$, $p = .350$, $\eta_p^2 = .01$. Critically, the change in simulation sickness symptoms from pre-to post-test did not differ between the two 3DVT conditions (all p's > .075). Average SSQ scores can be seen by testing phase, factor and 3DVT condition in Table 5 below.

Table 5. Average Simulation Sickness Questionnaire (SSQ) scores by 3DVT condition, testing phase, and SSQ factor

Measures	VR *M* (*SD*)	AR *M* (*SD*)
Pre-Nausea	0.26 (0.44)	0.73 (1.63)
Post-Nausea	0.65 (0.88)	0.61 (0.75)
Pre-Oculomotor	0.77 (1.12)	1.27 (2.00)
Post-Oculomotor	1.90 (1.92)	1.85 (1.89)
Pre-Disorientation	0.10 (0.30)	0.55 (1.42)
Post-Disorientation	0.74 (0.89)	0.73 (1.23)

Presence Questionnaire (PQ)

Independent-samples *t*-tests were conducted on each of the seven standardized subscales (i.e., Realism, Possibility to Act, Quality of Interface, Possibility to Examine, Self-Evaluation of Performance, Sounds, and Degree of Haptic Sensation) of the Presence Questionnaire to reveal any differences between the VR and AR conditions (see Figure 5). There was a main effect of 3DVT condition on Possibility to Act scores, $t(63) = 2.00$, $p = .050$, such that participants in the VR group ($M = 0.86$, $SD = 0.02$) had significantly higher scores than participants in the AR group ($M = 0.81$, $SD = 0.02$). However, there was no main effect of 3DVT condition on any of the other subscales, as follows: Realism, $t(63) = -0.33$, $p = .746$; Quality of Interface, $t(63) = 1.12$, $p = .265$; Possibility to Examine, $t(63) = 0.34$, $p = .733$; Self-Evaluation of Performance, $t(63) = -0.79$, $p = .435$; Sounds, $t(63) = 0.00$, $p = .997$; or Haptic, $t(63) = 1.55$, $p = .126$.

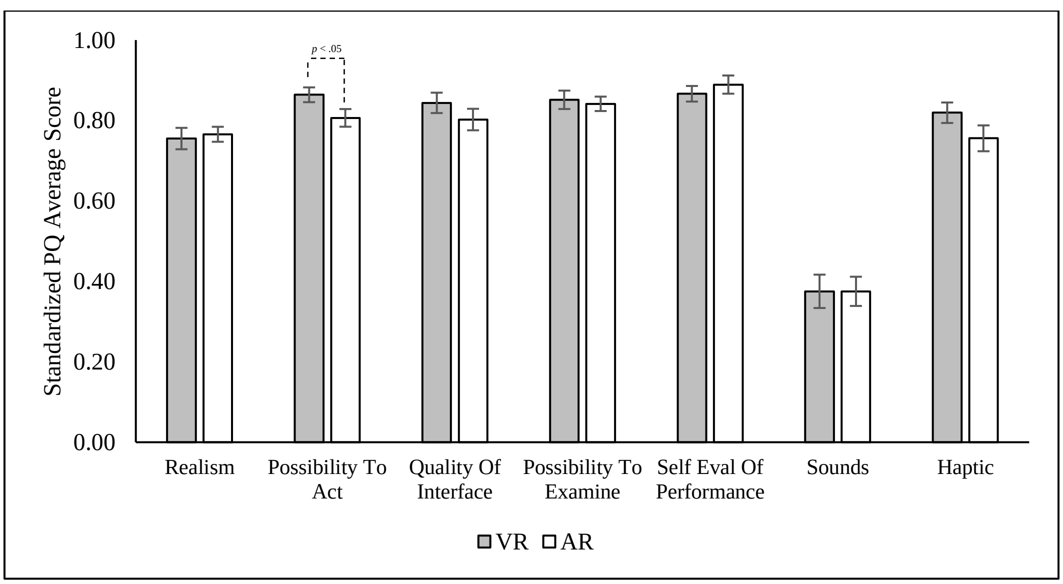

Figure 5. Standardized scores on Presence Questionnaire subscales

Electroencephalography (EEG)

A series of independent samples *t*-tests were conducted on average EEG activity at each combination of brain region (i.e., Frontal, Parietal, and Occipital lobes) and three power spectral bands (i.e., Alpha, Beta, and Theta waves) during the learning task. These results are summarized in Table 6 below. Average Beta activity in the Frontal lobe was significantly higher in VR (M = 4183.22, SD = 5127.53) than AR (M = -72.27, SD = 1413.76), $t(34.27) = 4.46$, $p < .001$. Average Beta activity in the Parietal lobe was also significantly higher in VR (M = 4050.35, SD = 5047.13) than AR (M = -317.66, SD = 1472.51), $t(34.78) = 4.64$, $p < .001$. No other significant differences between VR and AR were found (all *p*'s > .155).

Table 6. Average EEG activity during the learning task by 3DVT condition, brain lobe, and power spectral band

		VR *M* (*SD*)	AR *M* (*SD*)	Result
Alpha	Frontal	-4240.33 (9165.59)	-2290.83 (5410.04)	$t(62) = -1.04$, $p = .301$
	Occipital	-2983.12 (6035.95)	-1252.41 (2924.25)	$t(42.74) = -1.45$, $p = .156$
	Parietal	-3132.42 (7892.41)	-1944.29 (4012.41)	$t(62) = -0.77$, $p = .447$
Beta	Frontal	4183.22 (5127.53)	-72.27 (1413.76)	**$t(34.27) = 4.46$, $p < .001$**
	Occipital	-106.80 (5924.00)	-1053.14 (2408.20)	$t(39.15) = 0.83$, $p = .413$
	Parietal	4050.35 (5047.13)	-317.66 (1472.51)	**$t(34.78) = 4.64$, $p < .001$**
Theta	Frontal	-278.80 (2957.67)	207.48 (2311.81)	$t(62) = -0.74$, $p = .465$
	Occipital	-357.59 (2269.78)	164.02 (908.26)	$t(62) = -1.22$, $p = .227$
	Parietal	109.24 (3025.89)	-523.45 (7416.57)	$t(61) = 0.44$, $p = .661$

DISCUSSION

The present work provides a direct comparison of a commercially available virtual reality system (HTC Vive) and augmented reality system (zSpace AIO) in their ability to support the acquisition of anatomical knowledge. Learners completed a set of pre- and post-task surveys, between which they received training on one of the two technologies and were given ten minutes to study 16 brain structures. EEG was measured throughout the experiment. Results suggest that both VR and AR systems may provide similarly effective methods of gaining anatomical knowledge. Across the suite of measures, conditions differed only on Effort (higher in AR), Possibility to Act (higher in VR), and average Beta activity in Frontal and Parietal lobes (higher in VR).

Results suggest that these VR and AR systems are similarly effective for anatomical knowledge acquisition and have similar effects on learners in terms of stress, workload, and perceived usability. While learners in the AR condition perceived a higher degree of required effort when studying than learners in the VR condition, EEG measurements suggested VR may be associated with a higher level of concentration and engagement. Differences in Beta activity between the conditions may reflect differences in cognitive processes associated with the use of the 3DVTs (Ray & Cole, 1985). This may be explained in part by both the immersiveness of the virtual environment in VR and the inherent challenges with presenting virtual images in the real environment to an AR user.

Increased Possibility to Act ratings in VR compared to AR is not a surprise given the full immersion provided by virtual reality. It is also possible that the VR controls were more intuitive and natural to participants than using a stylus with the AR system. This finding highlights the common trend to focus VR/AR systems on supporting procedural-based tasks such as training surgical skills (de Visser et al., 2011; Muller-Stich et al., 2013; Baskaran et al., 2016). In this case, selection of the appropriate training system may require consideration of the device used to manipulate virtual objects. The HTC Vive controllers may provide a better sense of natural interaction by hands, whereas a stylus device might be more appropriate for certain surgical or pointing tasks that typically involve use of a peripheral tool.

The present findings mirror that of Moro et al. (2017), who found comparable benefits from using AR and VR in anatomical knowledge acquisition. However, the present work provides a finer level of discrimination between the two systems through inclusion of a wider range of measures. Future work should consider exploring the utility of

VR/AR systems for more intricate anatomical systems, tasks requiring retention of information, and knowledge vs. skill-based training.

Finally, the potential effect of spatial ability should be considered in future work. One measure of spatial ability was collected in the present study (the Cube Comparison Test; Ekstrom, French, Harman, & Dermen, 1976); however, it had no bearing on the pattern or significance of results and thus was left out. Still, spatial ability is a commonly cited factor in learning with three-dimensional visualization technologies, and more work needs to be done to understand the extent of its influence in anatomical training. Nonetheless, this experiment provides further evidence for the effectiveness of emerging visualization technologies for supplementing traditional anatomical training.

ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Army Research Laboratory (ARL) and was accomplished under Cooperative Agreement Number W911NF-15-2-0011. The views and conclusions contained in this document are those of the author's and should not be interpreted as representing the official policies, either expressed or implied, of ARL or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

REFERENCES

Baskaran, V., Štrkalj, G., Štrkalj, M., & Di Ieva, A. (2016). Current applications and future perspectives of the use of 3D printing in anatomical training and neurosurgery. *Frontiers in neuroanatomy*, *10*.

Brewer, D. N., Wilson, T. D., Eagleson, R., & de Ribaupierre, S. (2012). Evaluation of neuroanatomical training using a 3D visual reality model. In *MMVR* (pp. 85-91).

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

De Visser, H., Watson, M. O., Salvado, O., & Passenger, J. D. (2011). Progress in virtual reality simulators for surgical training and certification. *Medical Journal of Australia*, *194*(4), S38.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational testing service*.

Ghosh, S. K. (2017). Cadaveric dissection as an educational tool for anatomical sciences in the 21st century. *Anatomical sciences education*, *10*(3), 286-299.

Habbal, O. (2009). The state of human anatomy teaching in the medical schools of Gulf Cooperation Council countries: Present and future perspectives. *Sultan Qaboos University Medical Journal*, *9*(1), 24.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

HTC Corporation. (2018). *Vive VR System*. Retrieved from https://www.vive.com/us/product/vive-virtual-reality-system/

Hu, A., Wilson, T., Ladak, H., Haase, P., & Fung, K. (2009). Three-dimensional educational computer model of the larynx: voicing a new direction. *Archives of Otolaryngology–Head & Neck Surgery*, *135*(7), 677-681.

Keedy, A. W., Durack, J. C., Sandhu, P., Chen, E. M., O'Sullivan, P. S., & Breiman, R. S. (2011). Comparison of traditional methods with 3D computer models in the instruction of hepatobiliary anatomy. *Anatomical sciences education*, *4*(2), 84-91.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203-220.

Khot, Z., Quinlan, K., Norman, G. R., & Wainman, B. (2013). The relative effectiveness of computer-based and traditional resources for education in anatomy. *Anatomical sciences education*, *6*(4), 211-215.

Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). Profiling task stress with the Dundee Stress State Questionnaire. *Psychology of stress: New research*, 49-90.

Matthews, G., Emo, A. K., & Funke, G. J. (2005, July). A short version of the Dundee Stress State Questionnaire. In *Twelfth Meeting of the International Society for the Study of Individual Differences, Adelaide, Australia*.

Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical sciences education*, *10*(6), 549-559.

Müller-Stich, B. P., Löb, N., Wald, D., Bruckner, T., Meinzer, H. P., Kadmon, M., ... & Fischer, L. (2013). Regular three-dimensional presentations improve in the identification of surgical liver anatomy–a randomized study. *BMC medical education*, *13*(1), 131.

Munafo, J., Diedrick, M., & Stoffregen, T. A. (2017). The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. Experimental Brain Research, 235(3), 889-901.

Ray, W. J., & Cole, H. W. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*, 228(4700), 750-752.

Skochelak, S. E. (2010). A decade of reports calling for change in medical education: What do they say? *Academic Medicine, 85*(9), S26-S33.

Treleaven, J., Battershill, J., Cole, D., Fadelli, C., Freestone, S., Lang, K., & Sarig-Bahat, H. (2015). Simulator sickness incidence and susceptibility during neck motion-controlled virtual reality tasks. *Virtual Reality, 19*(3-4), 267-275.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, *7*(3), 225-240.

zSpace, Inc. (2018). *Behind the Magic of zSpace*. Retrieved from http://zspace.com/technology/: http://zspace.com/technology/