



# A Workload Comparison During Anatomical Training with a Physical or Virtual Model

Andrew Wismer<sup>1</sup>(✉), Lauren Reinerman-Jones<sup>1</sup>, Grace Teo<sup>1</sup>,  
Sasha Willis<sup>1</sup>, Kelsey McCracken<sup>1</sup>, and Matthew Hackett<sup>2</sup>

<sup>1</sup> Institute for Simulation and Training, University of Central Florida,  
Orlando, FL, USA

awismer@ist.ucf.edu

<sup>2</sup> Simulation and Training Technology Center, Army Research Lab,  
Orlando, FL, USA

**Abstract.** Recent research argues for the supplementation of traditional anatomical training with emerging three-dimensional visualization technologies (3DVTs); however, little is known regarding the effect these technologies have on learner workload. In this experiment, sixty-one participants studied gross brain anatomy using either a plastic physical model (PM;  $n = 29$ ) or models presented in virtual reality (VR;  $n = 32$ ). Participants were fitted with a functional near-infrared spectroscopy (fNIRS) sensor, worn on the prefrontal cortex. fNIRS measures regional saturation of oxygen ( $RSO_2$ ) and is indicative of workload. Participants then completed a pre-knowledge test on human brain anatomy. Participants were given 10 min to use the provided 3DVT to study 16 anatomical brain structures. Following the study period, participants completed additional surveys measuring workload, newly acquired anatomical knowledge, and cognitive resources used. Overall, anatomical knowledge increased at post-test and the change was no different between PM and VR conditions. Participants in the PM condition reported significantly higher levels of spatial workload, mental demand, and frustration.  $RSO_2$  values suggest left hemispheric increases from baseline during learning for the VR condition, but decreases for the PM condition. No other measures revealed differences between the two conditions. These results provide support for the supplementation of traditional anatomical training techniques with virtual reality technology as a way of alleviating workload. Further research is needed to explain the link between workload and performance in anatomical knowledge acquisition.

**Keywords:** Physiological response · Workload  
Methods and metrics for testing and evaluating augmented cognition system  
Virtual reality · Anatomical training · fNIRS · Physical model  
Visualization technologies

## 1 Introduction

When attempting to learn spatial information from two-dimensional displays, such as when students study gross anatomy using textbook images, a high level of workload is placed on the learner [1]. This workload has been associated with decreases in

knowledge acquisition [2]. Emerging three-dimensional visualization technologies (3DVTs) support a learner's understanding of spatial depth information by providing realistic representations of three-dimensional objects [3]. 3DVTs include physical models, virtual or augmented reality, and holographic displays. Indeed, numerous studies support the utilization of 3DVTs over traditional 2D displays [4–8].

Anatomical science is a domain that requires effective display of spatial information. Digital images are present not only in anatomical training and instruction, but also in medical diagnosis, pre-operative planning, and minimally invasive surgery [9]. Cadavers are commonly held as the gold standard for anatomical training as they enable hands-on experience with actual human tissue [10, 11]. However, cadavers can be costly to maintain, challenging to store, and require extra work for instructors [10–12]. For these reasons, 3DVTs are considered valuable supplements to traditional anatomical training.

A traditional supplement used in anatomical training is a physical model [12, 13]. Physical models replicate an anatomical system/structure using any variety of materials (e.g., plastic, fiberglass, clay). Physical models afford a learner hands-on experience with anatomical structures through rotation and often disassembly to aid in spatial comprehension. In general, physical models are easy to obtain, highly portable, and provide a useful tool to increase a person's base knowledge of anatomy at a low cost [14].

In addition, recent technological advances have increased the use of virtual reality in anatomical training. Virtual reality—the computer-generated simulation of three-dimensional objects/environments—provides capabilities similar to that of physical models for rotation, manipulation, and enhanced spatial understanding [15]. Much of the work to date on the use of virtual reality for knowledge acquisition in anatomical training has involved computer-based applications and modules (i.e., “desktop VR”). Research comparing these computer-based models to physical models has found benefits from using physical models for training in anatomical identification [16, 17]. While some work has been done with respect to more immersive virtual reality technologies (e.g., with head-mounted displays) for procedural training, much less has been done regarding the use of immersive VR for anatomical knowledge acquisition [18–20].

As physical and immersive virtual reality (VR) models share many features (e.g., both present 3D information, allow interaction and study of multiple views), an important criteria for evaluating their use for educational purposes is the level of workload they impose on the learner. Cognitive load theory [21] suggests increased workload is only detrimental if and when it exceeds a learner's working memory capacity. Currently, differences in detrimental workload imposed on a learner by 3DVTs such as physical or virtual reality models are unknown. Some 3DVT types (e.g., monoscopic 3D displays, digital holograms) have shown lower workload compared to 2D displays [22, 23], but this may not be true for physical and VR models.

The present work was designed to address the question: What are the differences in workload between physical and virtual reality models used for supporting knowledge acquisition in gross brain anatomy? This research question was addressed through both physiological and subjective measures of workload, with knowledge gain assessed through pre- and post- brain anatomy tests. The present experiment showed that

workload differences during anatomical knowledge acquisition may stem from limitations in the typical use of physical models compared to models presented in virtual reality.

## 2 Methods

### 2.1 Participants

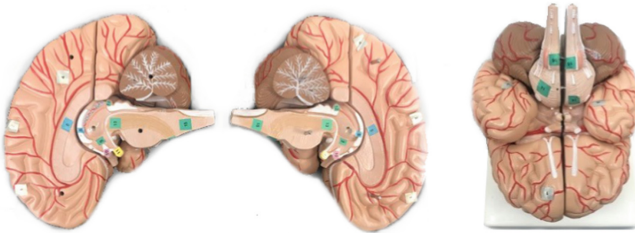
Sixty-one students from the University of Central Florida (29 Males, 32 Females), between the ages of 18 and 28 ( $Mdn = 18$ ,  $IQR = 1$ ), completed the experiment for course credit. All participants provided written informed consent prior to participation and were at least 18 years old with normal, or corrected-to-normal, vision.

### 2.2 Experimental Design

Participants were assigned to the physical model (PM) or virtual reality (VR) learning condition.

### 2.3 Materials

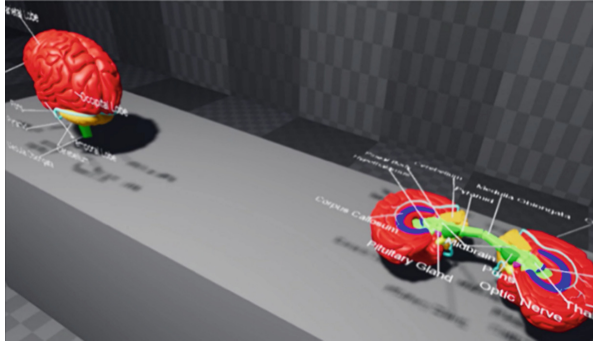
**Physical Model.** The physical model was presented along with a label sheet defining the numbered structures on the model (see Fig. 1). The numbered labels were added by the researchers and color-coded to best match the colored regions of the virtual model in the VR condition. The physical model ( $6 \times 5.5 \times 5.5$  inches) contained eight pieces and weighed 2.5 lb. The model could be examined as a whole or in any combination of its eight pieces.



**Fig. 1.** Plastic physical brain model with eight removable pieces (Axis Scientific) and numbered structures.

**Virtual Models.** The HTC Vive virtual reality (VR) system includes a head-mounted display, two controllers (one for each hand), and two “light house” sensors that track the headset and controllers, and project them into the virtual environment. The Vive connects to a desktop computer and displays a virtual environment through SteamVR.

Within the virtual environment, two brain models were displayed on a table (see Fig. 2). One brain model showed the external view of the brain, with label sets corresponding to a ventral and lateral view, while the second model displayed labeled structures from the medial view of the brain. Label sets could be toggled on or off by the participant using the controllers. The VR system allowed the participant to fully rotate the brain models to study the brain structures and spatial relationships from different viewpoints.



**Fig. 2.** Virtual brain models used in the VR condition.

## 2.4 Measures

### Workload Measures

*NASA-TLX.* The NASA-Task Load Index (NASA-TLX) [24] was administered following the learning task to assess workload. Six items addressing workload (i.e., mental workload, physical workload, temporal workload, effort, frustration, performance) were presented on 100 point scales in 5-point increments. High scores on each scale indicate high workload, with the exception of the performance subscale in which high scores relate to “poor” perceived performance.

*MRQ.* The Multiple Resources Questionnaire (MRQ), developed by Boles and Adair [25], was also administered after the learning task to assess workload in terms of cognitive resources used while studying with the 3DVT. Eleven of the original 17 items were utilized for purposes of this experiment (see Table 1). Responses were measured on a 100-point scale, with 0 indicating no usage and 100 indicating extreme usage.

### Performance Measures

*Spatial Anatomy Test.* A pre- and post-task Spatial Anatomy Test (SAT) was administered to evaluate knowledge gain. Accuracy and completion time were measured on identification (16 questions; one for each labeled brain structure), spatial knowledge (15 multiple-choice questions), and mental rotation questions (4 questions). Identification questions required the participant to select the correct label for a designated structure. Participants chose from a list of 32 brain structures (16 targets, 16 distractors). Four-alternative multiple-choice questions measured an understanding of spatial

**Table 1.** MRQ scores by 3DVT condition

	PM	VR	Test statistic	<i>p</i> -value
	<i>M (SD)</i>	<i>M (SD)</i>		
Manual	51.17 (27.08)	58.56 (22.98)	<i>t</i> (59) = 1.15	.254
Short-term memory	58.55 (32.45)	55.94 (26.36)	<i>t</i> (59) = -0.35	.730
Spatial attentive	82.31 (17.54)	77.94 (17.71)	<i>t</i> (59) = -0.97	.337
Spatial concentrative	74.66 (18.29)	58.47 (23.95)	<i>t</i> (59) = -2.94	.005*
Spatial emergent	66.45 (25.22)	48.88 (25.08)	<i>t</i> (59) = -2.73	.008*
Spatial quantitative	32.48 (29.62)	33.06 (25.32)	<i>t</i> (59) = 0.08	.935
Visual lexical	59.86 (26.23)	71.19 (25.05)	<i>t</i> (59) = 1.72	.090
Visual phonetic	44.00 (34.50)	36.50 (32.04)	<i>t</i> (59) = -0.88	.382
	<i>Mdn (IQR)</i>	<i>Mdn (IQR)</i>		
Spatial categorical	77.00 (36)	69.00 (32)	<i>U</i> = 553.00	.198
Spatial positional	90.00 (36)	69.00 (22)	<i>U</i> = 628.50	.017*
Tactile figural	60.00 (53)	30.50 (76)	<i>U</i> = 574.00	.112

*Note.* Asterisk (\*) represent statistically significant group differences.

relationships between brain structures. In the mental rotation section, a target image of the brain model was provided along with four rotated images (two of which were mirror-images). The participants selected which two of the four new images were simple rotations of the target image. Brain images used for identification and mental rotation test questions were matched to the respective physical or virtual reality model condition.

**Regional Saturation of Oxygen (rSO<sub>2</sub>)**

Changes in regional saturation of oxygen (rSO<sub>2</sub>) in the left and right prefrontal cortex were measured using the Somanetics INVOS Cerebral/Somatic Oximeter through near-infrared light [26]. This non-invasive, indirect neuroimaging measurement, referred to as functional near-infrared spectroscopy (fNIRS), sheds light into cognitive functions such as workload [27].

**2.5 Procedure**

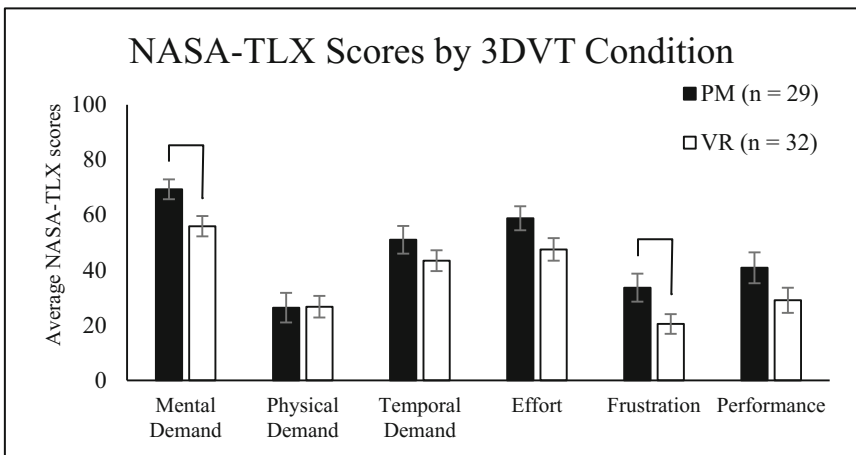
All participants provided written informed consent prior to participation. Each participant then completed a demographics survey and restrictions checklist, along with an Ishahara Color Blindness Test. These items did not serve as exclusion criteria; rather, they served to provide background information for use in later analyses. Next, the researcher fitted the participant with the fNIRS sensors. A five minute resting baseline was conducted as a reference for any changes in oxygenation during the experiment. The participant then completed the pre-task Spatial Anatomy Test (pre-SAT) to assess his or her prior knowledge concerning spatial brain anatomy. The participant was then given their assigned 3DVT (either the physical model or virtual reality system) and had ten minutes to use the technology to study the 16 labeled brain structures. The participant had the option to end the ten-minute study period early if they felt confident. Once the study time was complete, the participant completed a series of post-task

surveys on the computer, including the NASA-TLX, post-task SAT (identical to pre-task SAT but randomized order), and MRQ. Upon completion of the experiment, the fNIRS sensors were removed, and the participant was thanked, granted credit, and dismissed. The experiment took no longer than three hours to complete.

### 3 Results

#### 3.1 Workload

**NASA-TLX.** Independent-samples t-tests were conducted for four of the six subscales of the NASA-TLX (i.e., Mental Demand, Temporal Demand, Effort, and Frustration) to examine the effect of each 3DVT on workload. Nonparametric Mann Whitney *U* tests were conducted for the two other subscales (i.e., Physical Demand and Performance), determined to come from non-normal distributions. Average scores on each subscale can be seen in Fig. 3.



**Fig. 3.** NASA-TLX scores by 3DVT condition. Error bars represent standard error.

There was a significant difference between 3DVTs for Mental Demand scores,  $t(59) = -2.56$ ,  $p = .012$ , with higher Mental Demand in the PM ( $M = 69.31$ ,  $SD = 19.31$ ) than VR condition ( $M = 55.94$ ,  $SD = 20.73$ ). There was also a significant difference between PM and VR conditions for Frustration, with higher Frustration in the PM ( $M = 33.62$ ,  $SD = 27.35$ ) than VR condition ( $M = 20.47$ ,  $SD = 20.26$ ). There was no significant difference between the two 3DVTs with respect to Physical Demand,  $U = 415.50$ ,  $Z = -0.706$ ,  $p = .480$ ,  $r = -0.09$ , Temporal Demand,  $t(59) = -1.22$ ,  $p = .226$ , Effort,  $t(59) = -1.89$ ,  $p = .064$ , or Performance scores,  $U = 569.00$ ,  $Z = 1.53$ ,  $p = .127$ ,  $r = .20$ .

**MRQ.** Independent-samples *t*-tests were conducted for 8 of the 11 included MRQ subscales (Manual, Short-Term Memory, Spatial Attentive, Spatial Concentrative, Spatial Emergent, Spatial Quantitative, Visual Lexical, and Visual Phonetic processes) to examine the effect of 3DVT on workload.

Nonparametric Mann Whitney *U* tests were conducted for the remaining three subscales (i.e., Spatial Categorical, Spatial Positional, and Tactile processes), which were determined to violate normality assumptions. See Table 1 for each of the 11 MRQ scale scores by 3DVT condition.

There was a significant difference between 3DVTs on Spatial Concentrative process scores,  $t(59) = -2.94$ ,  $p = .005$ , with higher scores in the PM ( $M = 74.66$ ,  $SD = 18.29$ ) than VR condition ( $M = 58.47$ ,  $SD = 23.95$ ). There was a significant difference between 3DVTs on Spatial Emergent process scores,  $t(59) = -2.73$ ,  $p = .008$ , with higher scores in the PM ( $M = 66.45$ ,  $SD = 25.22$ ) than VR condition ( $M = 48.88$ ,  $SD = 25.08$ ). There was a significant difference between 3DVTs on Spatial Positional process scores,  $U = 628.50$ ,  $Z = 2.38$ ,  $p = .017$ ,  $r = .31$ , such that the PM condition ( $Mdn = 90.00$ ,  $IQR = 36$ ) had significantly higher scores than the VR condition ( $Mdn = 69.00$ ,  $IQR = 22$ ). No other significant differences were found between 3DVTs for the remaining subscales (all  $p$ 's  $> .089$ ).

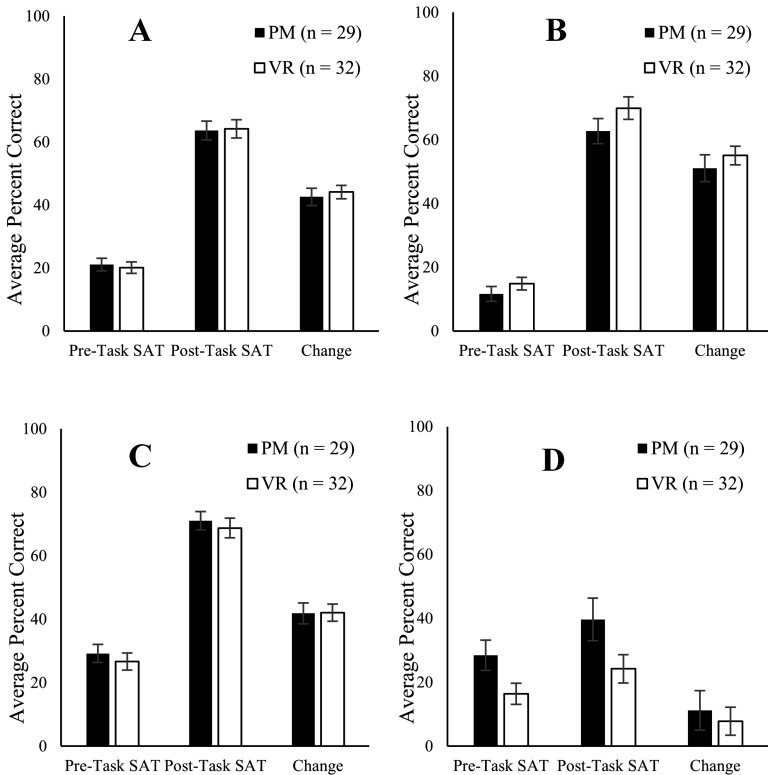
### 3.2 Performance

Independent samples *t*-tests were conducted on pre-task SAT accuracy scores to investigate any differences in prior knowledge between groups. There was no significant difference between PM and VR conditions for overall, identification, or multiple choice SAT accuracy (all  $p$ 's  $> .059$ ). However, there was a significant difference between PM and VR conditions for mental rotation accuracy scores ( $p = .039$ ).

2 (3DVT: PM, VR)  $\times$  2 (Testing time: pre, post) mixed factor ANOVAs were conducted on average SAT accuracy scores and completion times for the overall test and for identification, multiple choice, and mental rotation questions to examine differences in the level of spatial knowledge acquired between PM and VR conditions.

**Level of Spatial Knowledge Acquired.** There was a significant main effect of testing time on each of the following: average overall accuracy,  $F(1, 59) = 629.23$ ,  $p < .001$ ,  $\eta_p^2 = .91$ , average identification accuracy,  $F(1, 59) = 445.67$ ,  $p < .001$ ,  $\eta_p^2 = .88$ , average multiple choice accuracy,  $F(1, 59) = 396.57$ ,  $p < .001$ ,  $\eta_p^2 = .87$ , and average mental rotation accuracy,  $F(1, 59) = 6.49$ ,  $p = .013$ ,  $\eta_p^2 = .10$ . For each measure, post-task SAT scores were significantly higher than pre-task SAT scores (see Fig. 4).

There was no main effect of 3DVT condition on average overall,  $F(1, 59) = 0.01$ ,  $p = .942$ ,  $\eta_p^2 < .01$ , identification,  $F(1, 59) = 2.23$ ,  $p = .141$ ,  $\eta_p^2 = .04$ , or multiple choice accuracy scores,  $F(1, 59) = 0.47$ ,  $p = .494$ ,  $\eta_p^2 = .01$ . There was a significant main effect of 3DVT condition on average mental rotation accuracy,  $F(1, 59) = 5.73$ ,  $p = .020$ ,  $\eta_p^2 = .09$ , with higher accuracy in the PM ( $M = 34.05$ ,  $SD = 22.39$ ) than VR ( $M = 20.31$ ,  $SD = 22.39$ ) condition.



**Fig. 4.** Average overall (A), identification (B), multiple choice (C), and mental rotation (D) accuracy on SAT by 3DVT condition. Error bars represent standard error.

There was no significant interaction between testing time and 3DVT on any accuracy measure (all  $p$ 's > .430).

**Completion Times.** There was a significant main effect of testing time on each of the following: average overall SAT completion time,  $F(1, 59) = 12.06$ ,  $p = .001$ ,  $\eta_p^2 = .17$ , multiple choice completion time,  $F(1, 59) = 66.93$ ,  $p < .001$ ,  $\eta_p^2 = .53$ , and mental rotation completion time,  $F(1, 59) = 15.51$ ,  $p < .001$ ,  $\eta_p^2 = .21$ . For overall and multiple choice questions, participants took significantly longer to complete the post-task SAT than the pre-task SAT. Conversely, participants took significantly longer to complete the pre-task SAT than the post-task SAT for mental rotation questions. There was no main effect of testing time on average identification completion time,  $F(1, 59) = 0.22$ ,  $p = .642$ ,  $\eta_p^2 < .01$  (see Table 2).

There was no main effect of 3DVT condition on average overall,  $F(1, 59) = 0.04$ ,  $p = .836$ ,  $\eta_p^2 < .01$ , identification,  $F(1, 59) = 1.31$ ,  $p = .257$ ,  $\eta_p^2 = .02$ , multiple choice,  $F(1, 59) = 0.02$ ,  $p = .887$ ,  $\eta_p^2 < .01$ , or mental rotation completion times,  $F(1, 59) = 0.87$ ,  $p = .355$ ,  $\eta_p^2 = .02$ .



**Table 2.** Completion times on SAT by 3DVT condition and overall

Question type	Condition	Pre-task SAT	Post-task SAT	Change
		<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Overall	PM	400.32 (151.73)	488.59 (141.33)	88.27 (125.11)
	VR	428.54 (121.99)	447.95 (108.27)	19.41 (117.06)
Identification	PM	167.51 (82.51)	193.47 (59.53)	25.96 (89.40)
	VR	173.86 (65.43)	157.97 (49.66)	-15.90 (78.73)
Multiple choice	PM	134.82 (67.05)	215.22 (95.72)	80.39 (67.69)
	VR	155.90 (58.02)	198.90 (62.02)	43.00 (49.47)
Mental rotation	PM	97.99 (29.93)	79.90 (29.45)	-18.09 (27.67)
	VR	98.77 (23.12)	91.08 (29.64)	-7.69 (23.43)

*Note.* Completion Times are an average summation of all the questions in a category type (e.g., all 15 multiple choice questions).

There was an interaction between testing time and 3DVT for overall,  $F(1, 59) = 4.93, p = .030, \eta_p^2 = .08$ , and multiple choice completion times,  $F(1, 59) = 6.15, p = .016, \eta_p^2 = .09$ . Overall test and multiple choice questions showed significantly longer completion times for post- than pre-test. Average completion times were longer for the VR condition compared to the PM condition at pre-test with the reverse trend at post-test. There was no significant interaction between testing time and 3DVT for identification,  $F(1, 59) = 3.78, p = .057, \eta_p^2 = .06$ , or mental rotation completion times,  $F(1, 59) = 2.52, p = .118, \eta_p^2 = .04$ .

### 3.3 Regional Saturation of Oxygen (rSO<sub>2</sub>)

A 2 (3DVT: PM, VR) × 2 (Testing Time: pre, post) × 2 (Hemisphere: left, right) mixed factor ANOVA was conducted on average rSO<sub>2</sub> values during pre- and post-task SAT. There were no significant main effects or interactions among the included variables (all  $p$ 's > .225). A 2 (3DVT: PM, VR) × 2 (Hemisphere: left, right) mixed factor ANOVA was conducted on average rSO<sub>2</sub> values during the learning task. There was no main effect of hemisphere on average rSO<sub>2</sub> values during the learning task,  $F(1, 57) = 1.88, p = .175, \eta_p^2 = .03$ . There was a main effect of 3DVT condition,  $F(1, 57) = 9.17, p = .004, \eta_p^2 = .14$ , with higher change from baseline in VR ( $M = 1.78, SD = 3.51$ ) than PM ( $M = -0.65, SD = 2.99$ ). Critically, there was a significant interaction of hemisphere and 3DVT condition,  $F(1, 57) = 6.90, p = .011, \eta_p^2 = .11$ . Bonferroni pairwise comparisons revealed a significant hemispheric difference in the PM condition ( $p = .008$ ), but not in the VR condition ( $p = .367$ ). Pairwise comparisons revealed a significant difference in rSO<sub>2</sub> change from baseline in the left hemisphere between 3DVT conditions ( $p = .001$ ), while this difference between 3DVTs did not reach statistical significance for the right hemisphere ( $p = .053$ ). See Table 3 for average rSO<sub>2</sub> change from baseline values by 3DVT condition and hemisphere.

**Table 3.** Average RSO<sub>2</sub> change from baseline values by 3DVT condition and hemisphere

		PM ( <i>n</i> = 27) <i>M</i> ( <i>SD</i> )	VR ( <i>n</i> = 32) <i>M</i> ( <i>SD</i> )
Pre-task	Left	0.63 (2.33)	-0.05 (2.73)
	Right	0.27 (2.93)	0.31 (3.88)
Learning	Left	-1.23 (2.87)	1.97 (3.72)
	Right	-0.06 (3.05)	1.60 (3.35)
Post-task	Left	-0.25 (2.64)	0.24 (4.59)
	Right	0.78 (2.38)	0.04 (2.24)

## 4 Discussion

The present experiment provides a workload comparison between a physical model (PM) and models presented in virtual reality (VR) for supporting knowledge acquisition in anatomical training. While participants in both three-dimensional visualization technology (3DVT) conditions showed similar levels of knowledge gain, the VR condition decreased test completion time (pre- to post-test) to a greater extent than the PM condition. The PM condition was found to impose a higher degree of workload in terms of mental demand, frustration, and spatial processes, while average RSO<sub>2</sub> values suggested higher workload in the VR condition.

These results differ from previous studies comparing physical models to computer-based virtual models which showed benefits for physical models in anatomical knowledge acquisition [16, 17]. Thus, the more immersive virtual reality condition explored here may provide a closer match to the use of tangible, physical models. Benefits of physical models (e.g., tangible, portable) must be weighed against benefit of models presented in virtual reality (e.g., immersive, readily accessible library of models to access online).

The workload differences found in the present experiment may stem from different presentation formats between the physical and virtual reality models. The virtual reality model had structure labels fixed to the models (see Fig. 3) that could be toggled on and off. The physical model differed in that the model was labeled by numbers 1-16 with structure identifiers listed on a sheet of paper next to the model. Thus, studying with the physical model required an extra mental step to connect numbered labels to structure identifiers. It is possible that this difference is responsible for the higher workload seen in the physical model condition. Future work could better match the information presentation formats to provide a more direct comparison between 3DVT types.

Previous studies have shown workload affects a person's ability to learn from a 3D anatomical model [28–30]. The workload differences here were not associated with detriments in knowledge gain, but rather in test completion time. This suggests that the higher workload with the physical model relative to virtual reality models may not have been high enough to hinder overall knowledge gain. In other words, when the workload associated with a knowledge acquisition task is moderate at most, workload differences among 3DVT types may manifest in completion or response times rather than

knowledge gain. When task load increases, the selection of 3DVT may become more important for knowledge acquisition. This question is left open to future work.

In sum, educators and trainers should be aware of the capabilities and limitations of 3DVTs to ensure they do not impose a level of workload that hinders knowledge acquisition. The present experiment suggests that the selection of 3DVT for supporting anatomical knowledge acquisition may be made on factors such as cost, accessibility, and interest since minor differences in workload did not hinder learning. Still, further research is needed to better understand the link between workload and performance in spatial knowledge acquisition.

**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

1. Said, C., Shamsudin, K., Mailok, R., Johan, R., Hanaif, H.: The development and evaluation of a 3D visualization tool in anatomy education. *EDUCATUM J. Sci. Math. Technol.* **2**, 48–56 (2018)
2. Khalil, M., Paas, F., Johnson, T., Payer, A.: Interactive and dynamic visualizations in teaching and learning of anatomy: a cognitive load perspective. *Anat. Rec. Part B: New Anatomist* **286B**, 8–14 (2005). <https://doi.org/10.1002/ar.b.20077>
3. Young, J., Sewell, J.: Applying cognitive load theory to medical education: construct and measurement challenges. *Perspect. Med. Educ.* **4**, 107–109 (2015). <https://doi.org/10.1007/s40037-015-0193-9>
4. Müller-Stich, B., Löb, N., Wald, D., Bruckner, T., Meinzer, H., Kadmon, M., Büchler, M., Fischer, L.: Regular three-dimensional presentations improve in the identification of surgical liver anatomy – a randomized study. *BMC Med. Educ.* **13** (2013). <https://doi.org/10.1186/1472-6920-13-131>
5. Ruisoto, P., Juanes, J., Contador, I., Mayoral, P., Prats-Galino, A.: Experimental evidence for improved neuroimaging interpretation using three-dimensional graphic models. *Anat. Sci. Educ.* **5**, 132–137 (2012). <https://doi.org/10.1002/ase.1275>
6. Petersson, H., Sinkvist, D., Wang, C., Smedby, Ö.: Web-based interactive 3D visualization as a tool for improved anatomy learning. *Anat. Sci. Educ.* **2**, 61–68 (2009). <https://doi.org/10.1002/ase.76>
7. Hilbelink, A.: A measure of the effectiveness of incorporating 3D human anatomy into an online undergraduate laboratory. *Br. J. Educ. Technol.* **40**, 664–672 (2009). <https://doi.org/10.1111/j.1467-8535.2008.00886.x>
8. Hackett, M., Proctor, M.: Three-dimensional display technologies for anatomical education: a literature review. *J. Sci. Educ. Technol.* **25**, 641–654 (2016). <https://doi.org/10.1007/s10956-016-9619-3>
9. Escobar, M., Junke, B., Holub, J., Hisley, K., Eliot, D., Winer, E.: Evaluation of monoscopic and stereoscopic displays for visual–spatial tasks in medical contexts. *Comput. Biol. Med.* **61**, 138–143 (2015). <https://doi.org/10.1016/j.compbimed.2015.03.026>

10. Ghosh, S.: Cadaveric dissection as an educational tool for anatomical sciences in the 21st century. *Anat. Sci. Educ.* **10**, 286–299 (2016). <https://doi.org/10.1002/ase.1649>
11. Habbal, O.: The state of human anatomy teaching in the medical schools of gulf cooperation council countries: present and future perspectives. *Sultan Qaboos Univ. Med. J.* **9**, 24–31 (2018)
12. Baskaran, V., Štrkalj, G., Štrkalj, M., Di Ieva, A.: Current applications and future perspectives of the use of 3D printing in anatomical training and neurosurgery. *Front. Neuroanat.* **10** (2016). <https://doi.org/10.3389/fnana.2016.00069>
13. McMenamin, P., Quayle, M., McHenry, C., Adams, J.: The production of anatomical teaching resources using three-dimensional (3D) printing technology. *Anat. Sci. Educ.* **7**, 479–486 (2014). <https://doi.org/10.1002/ase.1475>
14. Yamine, K., Violato, C.: The effectiveness of physical models in teaching anatomy: a meta-analysis of comparative studies. *Adv. Health Sci. Educ.* **21**, 883–895 (2015). <https://doi.org/10.1007/s10459-015-9644-7>
15. Chavan, S.: Augmented reality vs. virtual reality: what are the differences and similarities. *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)* **5**, 1947–1952 (2018)
16. Khot, Z., Quinlan, K., Norman, G., Wainman, B.: The relative effectiveness of computer-based and traditional resources for education in anatomy. *Anat. Sci. Educ.* **6**, 211–215 (2013). <https://doi.org/10.1002/ase.1355>
17. Preece, D., Williams, S., Lam, R., Weller, R.: “Let’s Get Physical”: advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anat. Sci. Educ.* **6**, 216–224 (2013). <https://doi.org/10.1002/ase.1345>
18. Seo, J.H., Smith, B.M., Cook, M., Malone, E., Pine, M., Leal, S., Bai, Z., Suh, J.: Anatomy builder VR: applying a constructive learning method in the virtual reality canine skeletal system. In: Andre, T. (ed.) *AHFE 2017. AISC*, vol. 596, pp. 245–252. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-60018-5\\_24](https://doi.org/10.1007/978-3-319-60018-5_24)
19. Izard, S., Méndez, J.: Virtual reality medical training system. In: *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM 2016* (2016). <https://doi.org/10.1145/3012430.3012560>
20. Marks, S., White, D., Singh, M.: Getting up your nose. In: *2017 Symposium on Education on SIGGRAPH Asia - SA 2017* (2017). <https://doi.org/10.1145/3134368.3139218>
21. Paas, F., Renkl, A., Sweller, J.: Cognitive load theory and instructional design: recent developments. *Educ. Psychol.* **38**, 1–4 (2003). [https://doi.org/10.1207/S15326985EP3801\\_1](https://doi.org/10.1207/S15326985EP3801_1)
22. Foo, J., Martínez-Escobar, M., Juhnke, B., Cassidy, K., Hisley, K., Lobe, T., Winer, E.: Evaluating mental workload of two-dimensional and three-dimensional visualization for anatomical structure localization. *J. Laparoendosc. Adv. Surg. Tech.* **23**, 65–70 (2013). <https://doi.org/10.1089/lap.2012.0150>
23. Hackett, M.: Medical holography for basic anatomy training. In: *Interservice/Industry Training, Simulation, and Education Conference (IITSEC)*, pp. 1–10 (2013)
24. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* 139–183 (1988). [https://doi.org/10.1016/s0166-4115\(08\)62386-9](https://doi.org/10.1016/s0166-4115(08)62386-9)
25. Boles, D., Adair, L.: The multiple resources questionnaire (MRQ). *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **45**, 1790–1794 (2001). <https://doi.org/10.1177/154193120104502507>
26. León-Carrión, J., León-Domínguez, U.: Functional near-infrared spectroscopy (fNIRS): principles and neuroscientific applications. *Neuroimaging - Methods* (2012). <https://doi.org/10.5772/23146>
27. Peck, E., Afergan, D., Yuksel, B., Lalooses, F., Jacob, R.: Using fNIRS to measure mental workload in the real world. *Hum.–Comput. Interact. Ser.* 117–139 (2014). [https://doi.org/10.1007/978-1-4471-6392-3\\_6](https://doi.org/10.1007/978-1-4471-6392-3_6)

28. Garg, A., Norman, G.R., Spero, L., Maheshwari, P.: Do virtual computer models hinder computer learning? *Acad. Med.* **74**(Suppl 10), S87–S89 (1999). <https://doi.org/10.1097/00001888-199910000-00049>
29. Roach, V.A., Brandt, M.G., Moore, C.C., Wilson, T.D.: Is three-dimensional videography the cutting edge of surgical skill acquisition? *Anat. Sci. Educ.* **5**(3), 138–145 (2012). <https://doi.org/10.1002/ase.1262>
30. Lisk, K., McKee, P., Baskwill, A., Agur, A.M.: Student perceptions and effectiveness of an innovative learning tool: anatomy glove learning system. *Anat. Sci. Educ.* **8**(2), 140–148 (2015). <https://doi.org/10.1002/ase.1459>