ANATOMY EDUCATION AND VIRTUAL REALITY

THE EFFECT OF VARIED LEARNING ENVIRONMENTS AND MODALITIES ON ANATOMICAL KNOWLEDGE ACQUISITION, PERCEIVED WORKLOAD, CYBERSICKNESS, AND LEARNER ENGAGEMENT

By FARAH ZAREEN HASAN, Hons. B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2024) Hamilton, Ontario (Health Science Education)

TITLE: The Effect of Varied Learning Environments and Modalities on Anatomical Knowledge Acquisition, Perceived Workload, Cybersickness, and Learner Engagement

AUTHOR: Farah Zareen Hasan, Hons. B.Sc. (McMaster University)

SUPERVISOR: Dr. Ranil Sonnadara

PAGES: xii, 88

Lay Abstract

Institutions are looking to find the best learning technologies to deliver anatomy curricula to diverse student populations, often working with financial and time-based constraints. Visualization techniques have been at the forefront of this innovation, and the widespread use of virtual reality headsets has made once-impossible learning experiences achievable. This thesis explores the effect of different learning modalities and environments on learning with a pelvic floor anatomy module. We investigated how these factors, along with mental rotation ability and stereoacuity impact test performance and the perception of workload, cybersickness, and engagement. The results emphasize the importance of curricular design over the implementation of new technologies and the need to be critical of the impression that a one-size-fits-all solution exists.

Abstract

Institutions are looking to find the best learning technologies to deliver anatomy curricula to diverse student populations, often working with financial and time-based constraints. Visualization techniques, particularly the widespread use of virtual reality headsets, have made once-impossible learning experiences possible. This thesis explores the effect of different learning modalities (virtual reality headset, computer screen, and 3D-printed models) and environments (clinical context or context-free) on knowledge acquisition and learning experiences for a pelvic floor anatomy module. We investigated how these factors, along with mental rotation ability and stereoacuity, impact various aspects of learning, including performance on anatomy tests, perceived workload (measured using the NASA Task Load Index), cybersickness (measured using the Simulator Sickness Questionnaire), and engagement during learning (measured using the User Engagement Scale). Significant interactions were found between modality and environment for test scores and workload, a significant main effect of modality and environment for cybersickness, and a significant main effect of modality for engagement. Importantly, though significant differences were found between modalities and environments, participants reported concerningly high levels of workload and cybersickness across all conditions. High levels of engagement were also reported across all learning conditions. The lack of meaningful differences between intervention groups emphasizes the importance of curricular design over the implementation of new technologies and the need to be critical of the impression that a one-size-fits-all solution exists. Theories of cognitive load, constructivism, syncretion, visuospatial ability, cybersickness, and embodied learning in the context of technology-enhanced anatomy education are discussed as the foundation upon which design decisions should be made. A multi-faceted approach focused on aligning learning objectives with learning activities is outlined as a means of driving more impactful research and improving anatomy education.

Acknowledgements

I would like to thank Dr. Ranil Sonnadara and Dr. Bruce Wainman for their constant encouragement, guidance, patience, and support in the supervision of this thesis and for indulging my curiosity and constant questions. My thanks to Dr. Ruth Chen for her insight and support as a committee member and to Dr. Jennifer McBride for creating the opportunity to collaborate on this project. A heartfelt thanks to Josh Mitchell: colleague and co-conspirator extraordinaire. I am grateful for all the time, energy, and creativity you invested into this project. I would also like to thank Dr. Danielle Brewer-Deluce for generously sharing her time and expertise in statistics.

To my students Zoe, Athena, Aida, Julia, Alyssandra, and Nancy: you are the reason that this project was possible. I am immensely grateful for the enthusiasm, dedication, energy, and excitement that you brought not only to the VR Clinic, but to my Master's experience as a whole. Thank you to the members of the Sonnadara Lab for their ongoing support, especially to Kathleen Howcroft and Deewa Anwarzi who worked tirelessly to set me up for success, and to Ana Maria Ilicic for her thorough review of this thesis. Thanks to the members of the Anatomy Lab for your friendship and for making the lab feel like home, to Dr. Yasmeen Mezil for being a source of inspiration and encouragement throughout, and to Sakshi Sinha for helping me immerse myself in the world of anatomy education and virtual reality.

Thank you to Thorel Beckett for the time and energy you put into helping us set up our experiment, and to professor David Mazierski for working months on end to paint the beautiful models that we were able to use for our experiments.

Thank you to MERIT for financially supporting me in my studies and to the HSED program, instructors, and staff for this experience. I would also like to thank the Cleveland Clinic Foundation for financially supporting this project and the team at Zygote for this collaboration. A big thank you to all participants who enrolled in our study and to the beta testers for volunteering their time to shape our work.

To Mary and Kyle: thank you for endless emotional support. Thank you from the bottom of my heart to Jay, Candy, Karishini, Smruthi, Avika, Safat, Colin, and Zeeshan for carrying me through this degree, for sharing in my moments of excitement, nerdiness, and chaos, for inspiring me to pursue this work in the first place, and for helping me to get through the final stretch.

Finally, to my family: Mom and Dad for your support and love throughout, and to my sister Zahra, without whom this thesis and this degree simply would not have been possible. Thank you for everything.

Table of Contents

List of Figures and Tables

List of Abbreviations

3DPM: 3-dimensional printed model

ANCOVA: Analysis of covariance

ANOVA: Analysis of variance

CL: Cognitive load

CLT: Cognitive load theory

CS: Cybersickness

ECL: Extraneous cognitive load

GCL: Germane cognitive load

ICL: Intrinsic cognitive load

IPD: Interpupillary distance

IQR: Interquartile range

MRT: Mental rotations test

NASA-TLX: NASA task load index

SSQ: Simulator sickness questionnaire

UES: User engagement scale

VIMS: Visually induced motion sickness

VR: Virtual reality

VRLE: Virtual reality learning environment

WMC: Working memory capacity

Declaration of Academic Achievement

The work outlined in this thesis was completed by Farah Zareen Hasan, supervised by Dr. Ranil Sonnadara and Dr. Bruce Wainman. The design and development of equipment required for experiments was done with support from Joshua Mitchell, a Developer for the Education Program in Anatomy; Dr. Jennifer McBride, Medical Director for Anatomy at Baylor College of Medicine; Professor David Mazierski, Associate Professor for the Biomedical Communications program at the University of Toronto Mississauga; Jared Churchill developer at Zygote; and Ryan Klatte, Research Engineer at Cleveland Clinic. Experiment set up and data collection were supported by undergraduate students Zhiyu (Zoe) Wu, Athena Li, Aida Esmaelbeigi, Julia Issa, Alyssandra Mammoliti, and Nancy Paris Rosen. Statistical analysis was completed with the support of Dr. Danielle Brewer-Deluce.

1.0 Introduction

1.1 Teaching Tools for Anatomy Education

Anatomy education, a discipline with roots stretching back millennia, has undergone profound transformations in recent history. Within the past century alone, there has been a significant shift in perspectives regarding best teaching practices. Once confined to lecture-based instruction, today's educators widely embrace self-guided, technology-enhanced, exploratory learning as an essential part of the curriculum.

In medical education, however, there has been a concerning decline in the number of hours dedicated to anatomy education, particularly with the push for integrated curricula (Drake et al., 2014; McBride & Drake, 2018; Rockarts et al., 2020). This trend has resulted in graduates with limited knowledge of the fundamental aspects of human anatomy, highlighting the critical need for ongoing evaluation and improvement of anatomy curricula and instructional design.

Cadavers have been considered the gold standard for anatomy education for centuries (Estai & Bunt, 2016; Ghosh, 2015; Varner et al., 2021). Through cadaveric dissection, learners not only gain an understanding of anatomical structures but also develop the values, principles, and ethics which go on to shape how they practice medicine. However, the use of cadavers presents limitations, including visual differences between embalmed tissue and live tissue (Darras et al., 2018), lack of anatomical structural variability (Cheung et al., 2021), and limited representation of diverse body types (Finn et al., 2022). Additionally, wet laboratory instruction requires large sums of money to maintain specimens and facilities and the hiring of instructors with expertise to teach and supervise learners (Yammine, 2014). Advancements in digital learning technologies, such as virtual simulations, offer promising alternatives to traditional instruction with cadavers. These digital tools address issues of visibility, anatomical variability, making anatomy education more accessible and inclusive.

Among these digital tools, virtual reality (VR) stands out for its immersive and interactive nature (Pottle, 2019). VR head-mounted displays (HMDs) provide users with complete visual immersion, that can transport learners into any conceivable virtual reality learning environments (VRLE). Unlike many other digital technologies, its benefits are not limited to the visual senses: VR allows a multisensory immersion which can include auditory, haptic, and proprioceptive engagement. Uniquely interactive experiences are possible where the user can engage with the virtual environment, avatars, equipment and tools, simulations, and demonstrations that may be physically impossible in the real world, or difficult to access in the real world due to time, budget, safety, and legal constraints.

However, while VR holds immense potential as a learning tool, its adoption in academic settings, requires careful evaluation to ensure its efficacy and sustainability. Because of the realism and immersion, these VRLEs often elicit high sensory stimulation and can be physically and cognitively taxing in ways that may not be immediately evident to the user. Collaboration between education scientists, cognitive and behavioural psychologists, and experts is needed for the development of evidence-based design VR use tailored to educational objectives.

In health professions, VR offers exciting opportunities for trainees to explore realistic clinical environments and engage in uniquely interactive experiences. Particularly in anatomical education, VR allows for visualization and manipulation of anatomical structures in ways that cannot be achieved in a traditional laboratory setting, offering a promising impact on teaching and learning.

1.2 Key considerations for the design and implementation of VR learning tools

1.2.1 Cognitive load

Cognitive load theory (CLT) can be used to create the foundation for designing VR tools for anatomy education. CLT is based on the understanding that we have a limited working memory capacity (WMC), which allows us to process only a finite amount of information simultaneously (Sweller, 1988). When the information load surpasses our capacity, cognitive overload occurs, leading to the loss of incoming information that cannot be processed.

CLT encompasses three components of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load (ICL) reflects the resources required to complete the task given its inherent difficulty, while extraneous cognitive load (ECL) encompasses the resources spent in processing and making sense of the task. Germane cognitive load (GCL) pertains to the resources allocated to connecting new information with prior knowledge for memory consolidation.

To understand the role of each of these contributors, consider asking a learner to solve the equation: $x = 2 + 2$. The inherent complexity of the task remains unchanged (ICL), but variations in its presentation, such as on a blank piece of paper, on a cluttered blackboard, or demonstrating using marbles, affect ECL. Simplifying the presentation of the task frees up resources to focus on the task itself and on how to process the outcome in a way that promotes deeper learning. GCL allows new information to be connected to existing knowledge, enhancing understanding. For example, it may not matter to a toddler that two plus two equals four, but if the child has two crayons and is given another two crayons to make a total of four, there is context now that provides meaning to the information.

The goal in designing educational content and resources is to ensure that the ICL is appropriate for the intended learners, to minimize the ECL that could hinder learning by causing cognitive overload, and to optimize the GCL to facilitate deep learning (Young et al., 2014). Simple VRLEs can facilitative information reception, processing, and encoding, thus enhancing learning outcomes (Chen et al., 2015), but striking the right balance is key: reducing ECL may limit opportunities for optimizing GCL, potentially leading to cognitive underload, decreased engagement, and reduced knowledge acquisition (Paas et al., 2004).

Optimizing GCL involves giving meaning and personal value to new information, aiding memory retention. Visual cues with semantic meaning in the learning environment can foster critical connections and allow for the integration of new information into existing schemas (Ertmer & Newby, 2013). Instructors can accomplish this by providing priming cues, using appropriate language, metaphors, analogies, contextual environments (Ertmer & Newby, 2013).

Historically, in anatomical education, approaches often lack contextualization, presenting anatomical specimens and structures in isolation. Theorists (Perkins & Salomon, 1992) suggest that the more similarities exist between the learning and performance environments, the greater the transfer of learning. However, in anatomy education, there is a discrepancy between learning and real-world environments, posing a challenge for learners, particularly on a novice learner, leading to greater cognitive overload.

1.2.2 Constructivism and Syncretion

The constructivist philosophy of learning is based on the principle that learners actively construct knowledge by integrating new information with their existing understanding, rather than passively receiving knowledge from teachers (Seo et al., 2021). This process involves continually updating mental frameworks, referred to as schemata, to support existing knowledge.

Central to constructivism is learner-centred and learner-driven activity, with instructors serving as facilitators rather than transmitters of knowledge. Traditionally, anatomy has been taught in a "top-down" fashion, often through cadaver dissection, where students learn from the outside to the inside ("skin-to-bone"). While constructivist practices can still be applied in this context, a syncretic approach to learning anatomy is much better aligned. Syncretion involves systematically building up gross anatomy from smaller units (Miller, 2000), starting with the simplest layers and working up to a macro scale. Miller (2000) describes how the syncretic approach is preferable on a perceptual level, an important consideration for integrating with the foundations of CLT (Miller, 2000).

In order to facilitate learners' comprehension of the bigger anatomical context, instructors must provide scaffolding for students to build upon foundational knowledge. With a syncretic approach, this means focusing on subsystems rather than full systems (e.g., pelvic floor anatomy rather than all the musculoskeletal structures and organs within the pelvic cavity) and building from the innermost structures to the outermost structures (e.g., starting with the bones of the pelvic bowl and layering on muscles, tendons, and ligaments).

While some tools have been developed to explore and compare syncretion to dissection (Gangata, 2008; Sergovich et al., 2010), the literature remains limited, requiring further investigation. Nevertheless, the underlying cognitive principles of the theory remain strong, offering a promising approach for enhancing anatomy education.

1.2.3 Visuospatial Ability

The syncretic approach offers perceptual advantages by facilitating the understanding of larger anatomical structures from smaller components. According to Miller (2000), this method can reduce confusion, particularly for learners with low visuospatial ability and low mental rotation abilities (Miller, 2000). The relationship between visuospatial ability and test performance in anatomy education has been well established (Nguyen et al., 2014).

In anatomy courses, learners engage with various visualization tools and media, including textbook diagrams, cadavers, plastinated specimens, computer animations, and increasingly, VR headsets. This requires learners to be able to translate between two-dimensional representations and three-dimensional representations of anatomical structures. Importantly, when presented with a visual representation in a two-dimensional modality (such as a textbook illustration or a computer rendering), learners must translate the 2D representation into a 3D mental representation which they can then manipulate to understand the relationship between various anatomical structures and their relative positions to one another (Hegarty & Kriz, 2008). Interestingly, when learning from a 3D representation, research shows that our mental representations actually consist primarily of a few key views rather than the full range of views possible (Garg et al., 1999). The ease with which these mental representations are constructed by a learner varies within the population, and in the context of a classroom, instructors must be mindful of the tools selected for teaching as those with low visuospatial and mental rotation ability are disproportionately disadvantaged (Rochford, 1985). The use of physical models has been shown to facilitate this translation (Stull et al., 2012).

Stereoscopy has also been shown to play a key role in anatomy learning (Wainman et al., 2018, 2020). There is a distinct difference between learning tools which appear to offer 3D visualizations on 2D planar screens, such as computer applications, and those that utilize stereoscopic vision and are truly three-dimensional, such as VR HMDs or physical models. Several studies have compared the effectiveness of different modalities for anatomy learning and concluded that physical models are more effective than computer-based models due to their true three-dimensionality (Khot et al., 2013; Preece et al., 2013; Wainman et al., 2018, 2020), however, with the advancement of technology, there is research to suggest that VR HMDs may be just as effective as physical models (Brewer-Deluce et al., 2021; Wainman et al., 2021).

1.2.4 Cybersickness

The immersive nature of VR HMDs creates a unique sensory challenge: cybersickness. Cybersickness (CS) includes symptoms of discomfort ranging from nausea and dizziness to headache and eyestrain that is accompanied by the use of electronic visual displays like VR HMDs, computer screens, and training simulators (LaViola, 2000). CS is thought to be caused by sensory mismatch, much like motion sickness, where conflicting multisensory percepts contribute to physiological stress. This physical strain can jeopardize learning by causing fatigue, discomfort, and unwillingness to engage in the activity, whether consciously or unconsciously. CS can be assessed subjectively through biological and physiological responses and through self-reported measures in questionnaires. Autonomic changes in skin conductance (Gavgani et al., 2017), gastric activity, and blinking are among the measurements used (Dennison et al., 2016; Y. Y. Kim et al., 2005).

The Simulator Sickness Questionnaire (SSQ), developed by Kennedy et al. (1993), is widely used to evaluate simulator sickness including in VR settings (Kennedy et al., 1993). Consisting of 16 items, grouped into 3 subscales, (Nausea (N), Oculomotor disturbances (O), and

Disorientation (D)), the SSQ provides a comprehensive assessment of symptoms and their severity. Despite the development of alternative tools (Keshavarz et al., 2019; H. K. Kim et al., 2018), the SSQ remains the most widely used and validated tool, due to its applicability across various simulation technologies.

Reports of cybersickness with the use of VR in educational settings are frequent, with symptoms ranging from mild to severe. These symptoms can be attributed to both the hardware and software characteristics, and fortunately, strategies exist to mitigate them. Software modifications often focus on reducing vection (the perception of motion in the absence of physical motion) (Hettinger et al., 1990), viewpoint snapping (Farmani & Teather, 2020), field of view reduction, and peripheral blurring (Groth et al., 2021), while hardware improvements including high refresh rates, high display resolution, and using appropriate interpupillary distance (IPD) settings aim to reduce CS. Content complexity also influences levels of CS, with visually complex scenes and high interactive scenes (such as those encountered in games) inducing more symptoms (Saredakis et al., 2020). For applications involving locomotion, high resolution, elimination of lag, and tracking the user's movements with precision once again reduce the mismatch and can reduce CS (NATO Science and Technology Office, 2021).

1.2.5 Embodied learning and engagement

Tension between the philosophies of René Descartes and Immanuel Kant have driven the debate between mentalistic and embodied views of learning for centuries (Fincher-Kiefer, 2019). Descartes proposed the existence of a mind-body dualism, regarding learning as a purely mental process which occurs in isolation of the body, the environment, or emotions, while Kant argued that the interplay between mental and physical experiences give rise to knowledge. Modern educational perspectives have generally leaned toward Kant's perspective, recognizing that learning encompasses not only cognitive processes but also the body's engagement and the learning environment itself – referred to embodied learning (Skulmowski & Rey, 2018).

Two main principles of embodied learning are particularly important for the use of VR for anatomy education: the impact of physical engagement and the impact of the learning environment (Skulmowski & Rey, 2018). Learning involves the perception of novel stimuli received through multiple sensory modalities which are then processed and stored though a complex series of neural functions. Importantly, perception depends not only on stimuli, but also on physical state, expectations, experiences, and prior knowledge (Fincher-Kiefer, 2019).

Multisensory learning experiences inherent in embodied learning offer multiple pathways for encoding new information and optimizing the finite cognitive resources available to learners (Baddeley, 1992). Gestures play a vital role in embodied learning, aiding communication and freeing cognitive resources otherwise required for verbal tasks (Goldin-Meadow, 2011).

The physical environment significantly influences emotional, physiological, behavioural, and cognitive aspects (Choi et al., 2014). Environmental cues shape our behaviour and affect cognitive processes. While novel environments can be beneficial for stimulating attention and engagement, there exists an optimal range in which performance peaks – too much novelty can deplete cognitive resources in an attempt to process the environment, leaving little capacity to carry out learning tasks (Kalet et al., 2012). Finding a balance is essential to maximizing learning benefits while avoiding cognitive overload or underload.

Interest in the field of environmental psychology has been growing, with evidence-based design emerging as a prominent approach to building and designing physical spaces. Studies, such as

Ulrich's (1984) seminal work comparing postsurgical outcomes based on patient's room views, exhibit the profound impact of environment factors on learning outcomes (Ulrich, 1984).

Engaging learners in today's classroom poses significant challenges for educators, with prominent distractions in the digital age. The concept of "edutainment" is a contentious one, with some educators arguing that learning must be entertaining for success (Aksakal, 2015; Okan, 2003). Leveraging the unique qualities of VR for embodied learning naturally fosters engagement through increased presence, immersion, and interactivity (Johnson-Glenberg, 2018), and when designed thoughtfully, VR learning tools can support both education and entertainment.

1.3 Purpose of this work

The aim of this thesis was to evaluate the impact of various learning modalities and environments on knowledge acquisition, workload, cybersickness, engagement, and learner perceptions, using a pelvic floor learning module called *cAnatomy*.

Designed by the Cleveland Clinic and produced by Zygote, the *cAnatomy* Pelvic Floor module details a clinical case of a 24-year-old woman experiencing postpartum urinary and flatal incontinence following a natural delivery. Interactive features allow users to adjust magnification and rotate anatomical structures for enhanced viewing. Two versions of the module were created:

a clinical version where the pelvic models are presented in a hospital room, and a blackout version where the pelvic models are presented against a featureless, black background (Figure 1).

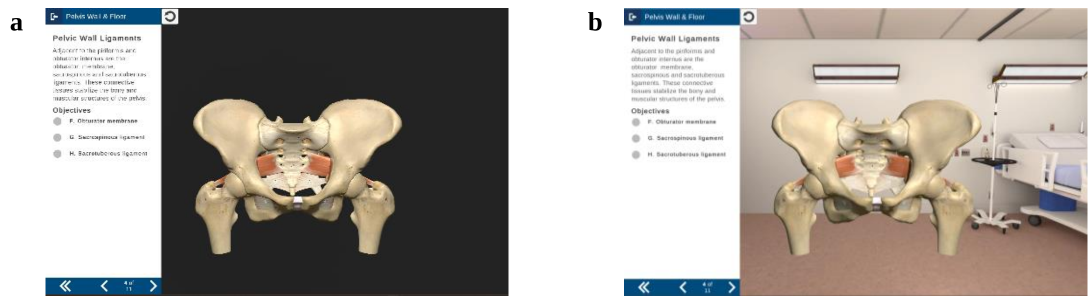


Figure 1. cAnatomy pelvic floor learning module in (a) blackout and (b) clinical environments.

This study used a 3x2 experimental design to explore the impact of modality (3D printed models, computer, or virtual reality) and environment (enriched clinical or contextless blackout) on test scores, perceived workload, cybersickness, and user engagement when using a pelvic anatomy learning module.

The goal with this work was to gain a better understanding of the factors influencing knowledge acquisition and learning experiences in anatomy education. By identifying areas of concern and offering solutions, direction can be offered for the selection of appropriate tools for impactful learning experiences.

1.3.1 Research questions and hypotheses

Aligned with the objectives, the following five research questions were addressed:

RQ1. Is there a relationship between test scores, perceived workload, cybersickness, user engagement, mental rotation ability, and stereoscopic discriminability?

RQ2. Do learning modality and learning environment affect test scores?

RQ3. Do learning modality and learning environment affect perceived workload?

RQ4. Do learning modality and learning environment affect cybersickness?

RQ5. Do learning modality and learning environment affect user engagement?

It was hypothesized that test scores would be highest and workload lowest when using 3D printed models (3DPMs). Additionally, the enriched clinical environment was expected to increase test scores but also to increase workload.

Cybersickness was expected to be highest in the VR groups and lowest in the 3DPM groups due to the digital nature and extent of sensory mismatch that can occur with the use of VR. Furthermore, greater presence and severity of cybersickness was predicted for the clinical group due to additional visual stimuli in the clinical environment.

Regarding engagement, it was hypothesized that VR would yield the highest engagement levels and 2D the lowest. Greater engagement was also expected in a visually complex clinical environment compared to a contextless blackout space, due to its immersive quality.

2.0 Materials and Methods

2.1 Recruitment

One hundred and twenty undergraduate students at McMaster University were recruited to participate in this study, with 40 participants in each of the 6 experimental groups (VR clinical, VR blackout, computer clinical, computer blackout, 3DPM clinical, 3DPM blackout). Eligible participants were at least 18 years of age, had not taken any university-level gross anatomy or physiology courses, and did not have known discomfort when using VR (e.g., nausea, dizziness, anxiety). The study was approved by the Hamilton Integrated Research Ethics Board under Project Number 14210.

Participants were recruited primarily through the use of an online portal (Sona Systems, Ltd., n.d.) where researchers in the Department of Psychology, Neuroscience and Behaviour at McMaster University post brief descriptions of studies to recruit participants (Appendix A). A non-probability, convenience sampling strategy was used as first year psychology courses offer extra credit to students for enrolling in research studies. Additional recruitment was also done through advertisement posters around the McMaster University campus, word of mouth, and social media posts on Instagram (Appendix B).

Upon enrolling in the study, participants received an email with a link to a screening survey (Appendix C), hosted on LimeSurvey v5.6.53+240131 (Schmitz, 2003). Survey responses were reviewed, and eligible participants were emailed instructions on how to find the Anatomy Lab along with the Letter of Information for the study. All experiments were carried out in an office space within the Anatomy Laboratory at the McMaster University Health Sciences Centre.

2.2 Experiment Set Up

2.2.1 Learning Module

The cAnatomy Pelvic Floor module v2.49.1 (Churchill, 2022) was used for the learning task in this study. The module featured eleven scenes revolving around the clinical case of a 24-year-old woman experiencing postpartum urinary and flatal incontinence after a natural delivery (Figure 2). The first and last scenes introduced and summarized the case study, while the other nine scenes depicted pelvic floor anatomy using a syncretic approach (i.e., scene two displayed only the bony pelvis, scene three displayed the bony pelvis and pelvic wall muscles, and scene four displayed the bony pelvis, pelvic wall muscles, and pelvic wall ligaments). Key structures were labeled with capital letters corresponding to a legend with accompanying text to explaining the function of the structure, its relationship to other structures, and relevance to the case study. Users were able to adjust the magnification of the models and rotate them in order to view anatomical structures from different angles. The module content was identical for all participants; only the modality used to display the module varied.

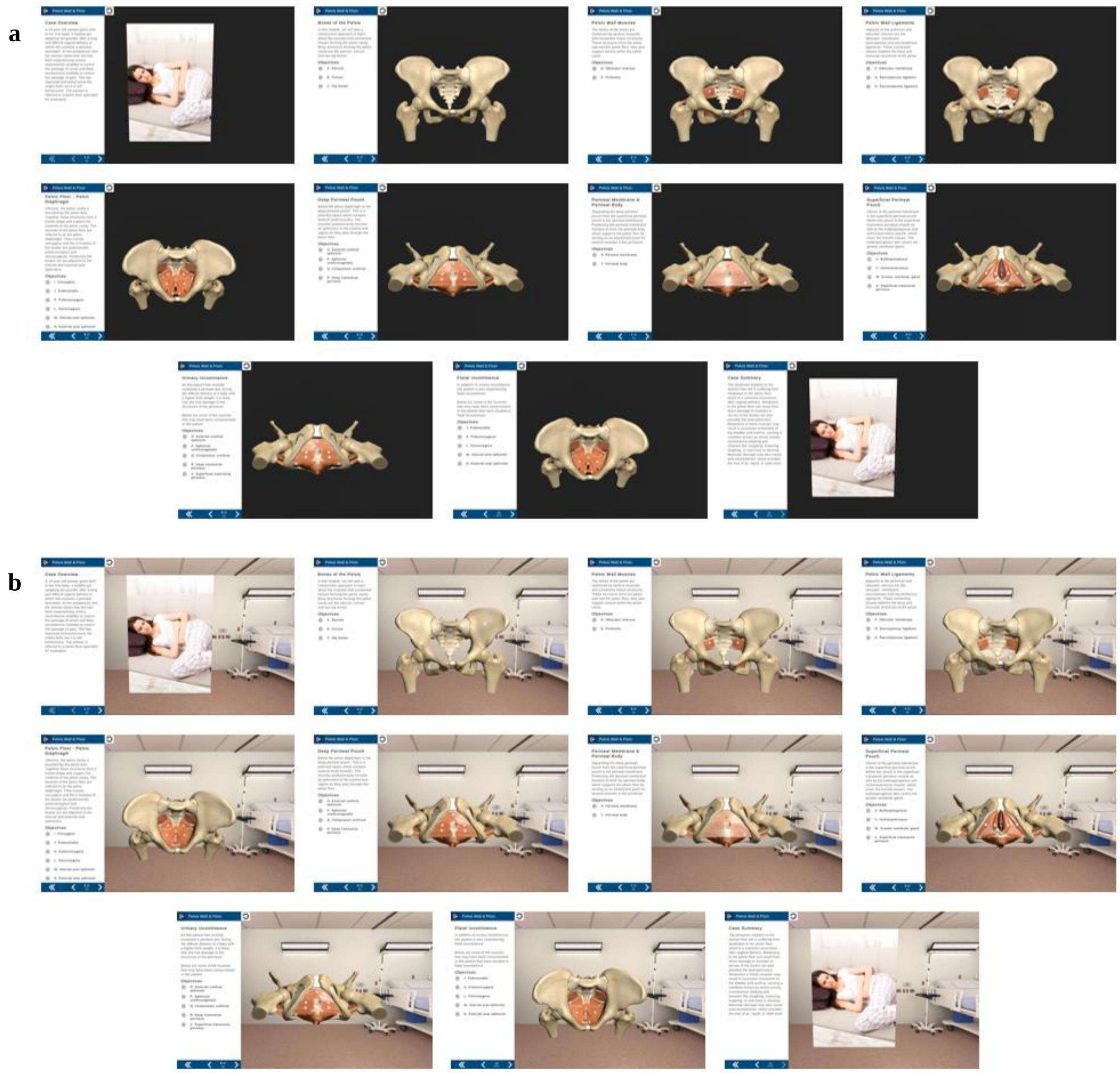


Figure 2. Scenes one through eleven of the cAnatomy module in the (a) blackout environment and (b) clinical environments.

2.2.2 VR

For the VR condition, the module was run on a Razer Blade 15 laptop (Razer, Irvine, California, USA) and displayed on a Meta Quest 2 VR headset (Meta Platform Technologies LLC, Menlo Park, California, USA).

In the development of the standard operating procedures used for our experiment, a number of equity-related challenges came to light that afforded the opportunity for thoughtful design and important discussion around making VR accessible for diverse learners. Preliminary beta testing revealed that that some participants had a flatter, less pronounced nose bridge which created a large gap between the face and the bottom of the headset. This resulted in a significantly less immersive experience as the floor of the room and the participant's feet were easily seen, and also an increased risk of distraction while viewing the learning module. While those with a more pronounced nose bridge did have some view of the external environment, it was significantly reduced, creating a gap in the experience depending on facial features. A modified silicone light blocking facial interface (Petiarkit) was tested in place of the original facial interface included with the Meta Quest 2 to block peripheral light and to increase immersion (Figure 3). Beta testers with flatter nose bridges found this to be a significant improvement and testers with pronounced bridges also noted improvements. Based on this feedback, the light-blocking facial interface on the headset was adopted for all participants.

Headset fitting also required thoughtful planning with consideration for varied head size and shape, hair types, and head coverings. It was important to ensure that the headset fitting process was as quick, comfortable, and unobtrusive as possible. Individuals with long hair were encouraged to either wear their hair down or in a low bun so that the straps could be securely tightened across the back of the head without interference. The protocol was also tested to determine the best ways to fit the headset for individuals who observe religious head coverings such as hijab, turban, and patka. For head coverings which were raised and were directly on top of the head, it was possible to remove the top strap in the sagittal plane and still keep a secure fit with only the horizontal strap. Participants were emailed a photo of a member of the research team wearing the headset with hair tied in a low bun and recommendations were offered for how participants might wish to wear their hair in order to feel most comfortable with the headset on (Appendix D). Participants were also encouraged to reach out to the research team if they required any further accommodation for the headset fitting so that we could ensure a comfortable and respectful experience for all participants.

The interpupillary distance (IPD) was also measured by the researcher using a plastic optical vernier pupillary distance ruler (Uoeo) and the IPD on the headset was adjusted accordingly. IPD is a known correlate of cybersickness, therefore it was important to control for inappropriate fitting as a potential source of error (Stanney et al., 2020). Participants were given either the left- or right-handed Meta Quest 2 controller, depending on their handedness, to navigate through the scenes of the module and to rotate and magnify the pelvis.

Figure 3. Meta Quest 2 headset with original facial interface (top) and modified silicone light blocking facial interface (bottom).

2.2.3 Computer

For the computer condition, the module was run and displayed on the Razer Blade 15 laptop. Participants were given a standard computer mouse with a scroll wheel (HP Inc.) to navigate through the scenes of the module and to rotate and magnify the pelvis (Figure 4). It should be noted that unlike in the VR modality, the entire clinical environment was not visible as a 360-degree view of the room was not possible with the computer application.

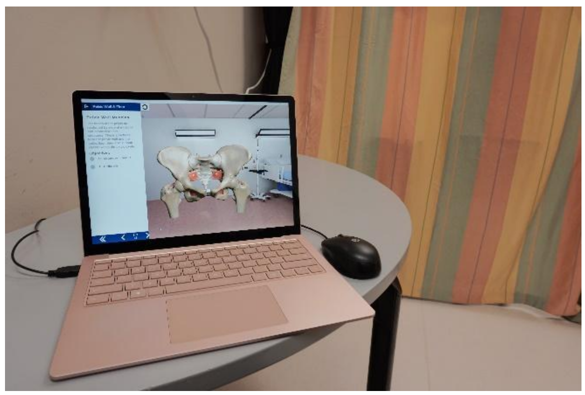


Figure 4. Computer condition set up with computer mouse. N.B. Razer Blade 15 laptop is not pictured here.

2.2.4 3DPM

For the 3DPM condition, the pelvis models from the cAnatomy module were exported as 3D-printable files and printed on a ColorJet powder bed fusion printer. Due to visual inconsistencies in the colour of different structures across the 3D prints, and to ensure consistent colouring between presentation modalities, all models were hand painted by Professor David Mazierski, a medical illustrator and Associate Professor for the Biomedical Communications program at the University of Toronto Mississauga (Figure 5). To prepare the models for painting, grey Flex Fill Multi-Purpose Filler auto body primer was used (Dominion Sure Seal Ltd., 24002). Satin finish commercial house paints were chosen for the model base coats. Bony structures were painted with Canadian Tire Premier Paint “Sugar Cookie” (PR20D17-2); muscles were painted with Behr Premium Plus “Tiki Torch” (M180-6, Home Depot); and cartilage and tendons were painted with Behr Ultra “Falling Snow” (18-07, Home Depot). The VR models included colour texture maps which needed to be recreated on the 3D-printed models. This was accomplished using a Holbein Hohmi-Dash Y2D airbrush attached to an Iwata Power Jet Pro IS975 compressor; pressure used was 20 PSI. Vallejo Model Color acrylic hobby paints were diluted to 25% using distilled water. Muscle striations were painted with Cavalry Brown (982), tendon and ligament texture were suggested with White (951), and random markings on the bones to suggest foramina and rugosity were painted with Flat Brown (984).

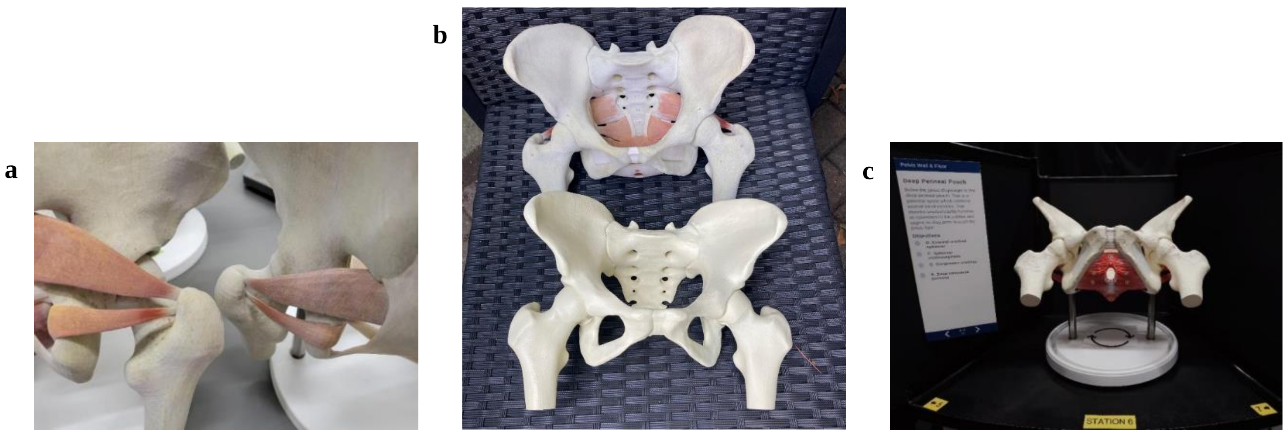


Figure 5. 3D-printed models in various stages of preparation. (a) Original models prior to painting demonstrating colour variation between models. (b) Comparison between unpainted model (top) and model with base coat of paint (bottom). (c) Completed model with airbrushed details.

To mimic the visual presentation of the pelvis models in the VR and computer conditions, a custom table was constructed to allow participants to view one model at a time in such a way that they could mimic the rotation and magnification functions of the module when displayed in a digital format. A large turntable with a 6-foot diameter was constructed using plywood and a ball bearing. The large turntable was secured on top of a stationary table to elevate its surface which would serve as the viewing platform for the 3DPMs. Each of the 3DPMs were placed upon a 10-inch turntable (Pretireno). This set up allowed participants to rotate the large turntable to navigate between the scenes of the module, and to rotate the smaller turntables to rotate individual pelvis models (Figure 6).

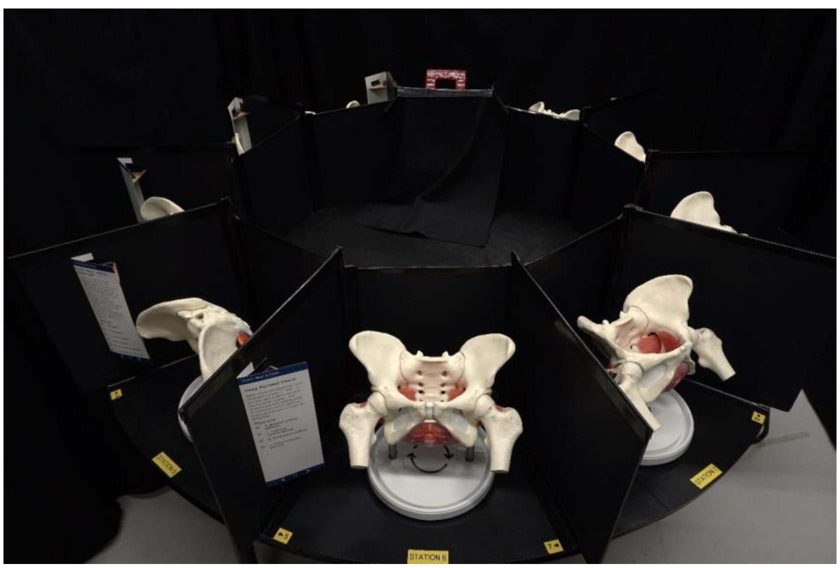


Figure 6. 3DPM turntable setup in blackout environment. Stations were spatially arranged to mimic the setup within the VR and computer presentations of the module. Stations were divided with black walls so that only one scene was visible at a time.

Structures of the 3DPMs were labelled using small, adhesive labels. The accompanying text and legend for each scene was laid out and formatted identically to that of the digital modalities and was printed and kept beside the 3DPMs.

2.2.5 Learning Environments

To mimic the clinical environment of the cAnatomy module, a simulation space was designed with props including an examination bed, desk with folders and a computer, sharps bin, vitals monitor, and various wall-mounted items including boxes of gloves and masks, clinical posters, a whiteboard, and x-ray films. Participants assigned to the clinical learning environment in all 3 learning modalities were seated in the simulation area while the researcher explained the study and obtained written consent from the participants. In the 3DPM condition, this clinical space was visible to the participant from their seat at the large turntable and served as a backdrop to the physical learning module (Figure 7a).

For the blackout condition, black curtains were used to conceal the simulation space and were also drawn across the walls to mimic the VR and computer learning environments (Figure 7b).

a **b**

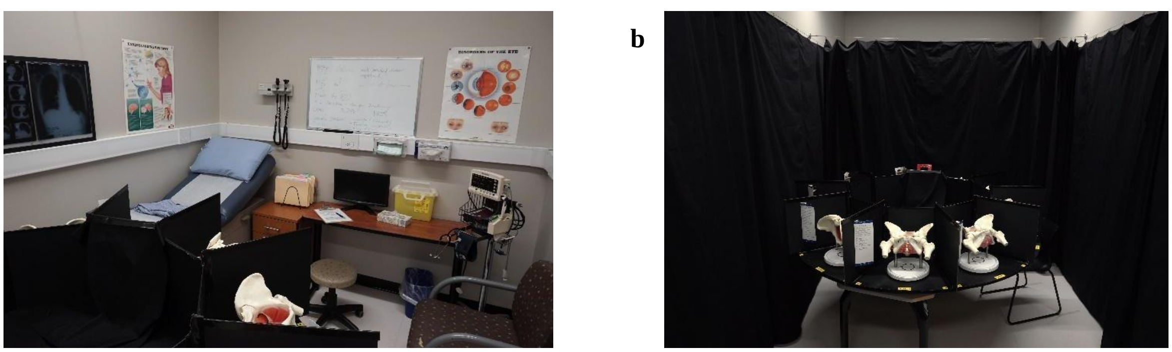


Figure 7. Physical layout for the (a) clinical learning environment and (b) blackout learning environment.

2.3 Data Collection and Experimental Protocol

Due to time constraints and issues with equipment availability during the COVID-19 pandemic, data collection was carried out in blocks based on learning modality. All VR data were collected prior to the start of computer data collection and 3DPM data were collected last. Eligible participants were randomly assigned to either the clinical or blackout condition for the learning modality being recruited for at the time. Participants were not aware of which learning modality nor which learning environment they were assigned to prior to participation since it was not advertised on the recruitment portal. A summary of the experimental protocol is shown in Figure 8.

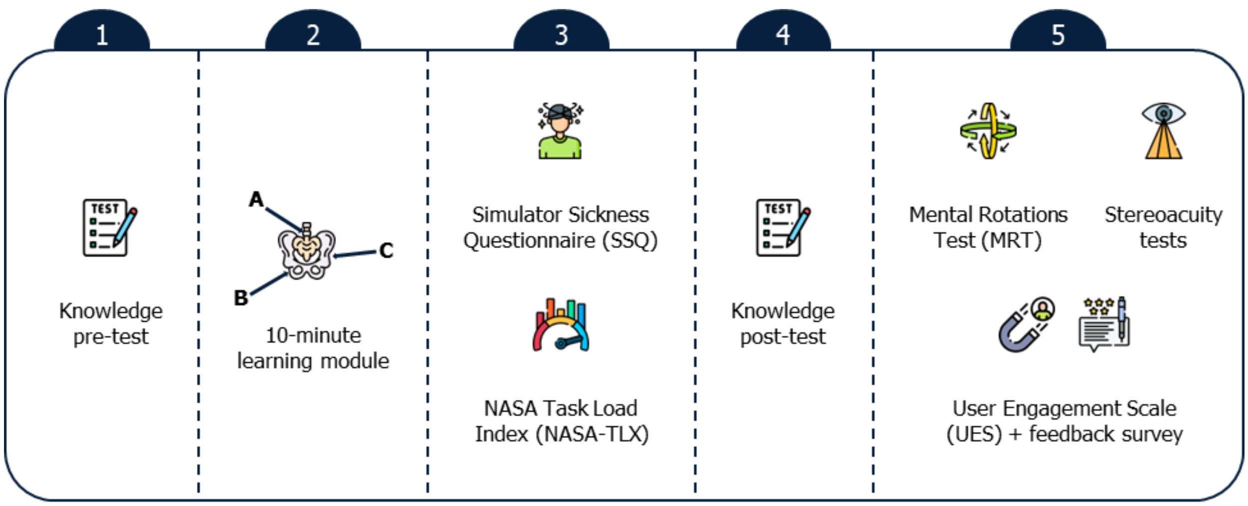


Figure 8. Summary of experimental protocol. 1. Participants completed a knowledge pre-test assessment. 2. Participants were given 10 minutes to navigate the pelvic anatomy learning module in their assigned learning condition. 3. The Simulator Sickness Questionnaire and NASA Task Load Index were administered in counterbalanced order. 4. Participants completed a knowledge post-test assessment. 5. The Mental Rotations Test was administered followed by tests of stereoacuity. Participants then completed the User Engagement Scale and an open-ended feedback survey.

2.3.1 Informed Consent

Upon arrival at the Anatomy Laboratory, participants signed in at the main office and were then brought into the experiment room which was set up according to the randomly assigned condition. Participants were seated either in the simulation clinical space or blackout space while the researcher reviewed the Letter of Information and consent form. Following the verbal overview, participants were given time to review the documents independently and to ask questions before signing the consent form.

2.3.2 Pre-test Assessment

Participants completed a 15-question pre-test assessment on paper (Appendix E). Two reference sheets were provided for use during the assessment: a word bank containing structures of the female pelvis (Appendix G) and a list of anatomical orientation terms (Appendix H). For some questions, participants were required to identify structure labeled on two plastic models of the pelvis (Candent, 12 Parts Female Pelvic Floor Model; SOMSO, Female Pelvis Floor Model) (Figure 9). Participants were given 15 minutes to complete the assessment with a 5-minute warning, and were instructed to complete all 15 questions, even if guessing was required.

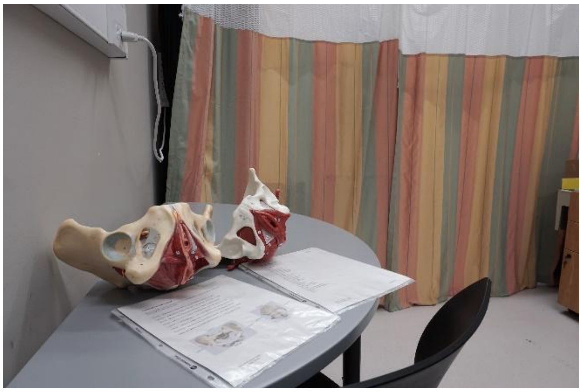


Figure 9. Pre-test and post-test assessment setup with reference sheets and testing pelvis models.

2.3.3 Learning Task

Participants were briefed on the learning task and were told that they would need to memorize the information presented in the learning module on which they would later be tested. The researcher explained the layout of the module, referring to a printed screenshot as a visual guide. Instructions on how to navigate through the eleven scenes, and how to rotate and magnify the pelvis models were also provided. Participants were given ten minutes to navigate the module, with a five-minute warning, and were free to move through the scenes at their own pace.

2.3.4 SSQ and NASA-TLX

Following the learning task, the SSQ and NASA-TLX were administered, counterbalancing the order within each of the six experimental groups. The SSQ (Kennedy et al., 1993) was administered digitally using a LimeSurvey questionnaire (Appendix I) which was completed on the Razer Blade 15 laptop for the VR and computer conditions, and on a Toshiba Chromebook 2 (Toshiba Corp., Tokyo, Japan) for the 3DPM condition. Participants were instructed to read the prompts carefully and consider how they felt while using their respective learning modality to view the cAnatomy module.

For the NASA-TLX, participants were provided definitions for the 6 factors contributing to workload (Appendix J) and were told to reflect on the task of memorizing the cAnatomy module content presented in their assigned modality when rating their experiences (Hart & Staveland, 1988; *NASA Task Load Index (TLX) v 1.0 Paper and Pencil Package*, n.d.). For Part 1, participants were provided a scoring sheet and asked to mark their responses on the sheet. For Part 2, participants were provided with a deck of cards containing the factor comparison pairs and asked to circle their responses.

2.3.5 Post-test Assessment

The post-test assessment was then administered using the same method as the pre-test assessment (Appendix F). It consisted of 15 questions which were similar in nature to those on the pre-test assessment and were balanced for level of difficulty and type of question.

2.3.6 MRT and Stereoacuity Tests

To assess mental rotation ability, the Vandenberg and Kuze MRT was administered on paper (Vandenberg & Kuse, 1978). Participants were given time to review the instructions and complete a sample problem. They were then given five minutes for each of the two parts of the test (Appendix K).

Stereoacuity was evaluated using the Stereo Fly Test (Stereo Optical Company Inc., Chicago, IL) (Appendix L). Participants were asked to put on a pair of polarized glasses and were seated so that the distance from their eyes to the stereo test booklet was measured at 40.6cm, as is indicated for the Stereo Fly Test.

2.3.7 UES and Feedback Survey

Participants were asked to complete the UES (O’Brien et al., 2018) and a feedback survey on the laptop (Appendix M). The feedback survey included three items:

> 1) What facilitated or limited your ability to understand the educational content presented in the pelvic anatomy learning module?
>
> 2) What recommendations do you have to improve the user experience in the pelvic anatomy learning module?
>
> 3) [Do you have any] Additional comments for the investigator[?].

2.3.8 Debrief and Optional VR Demo

Once the final survey was completed, participants were informed that the experiment had concluded and were provided a debrief sheet. Extra credit was assigned to the participants through the online recruitment portal. In the computer and 3DPM groups, participants were offered the chance to explore a five-minute VR demo if they wished, but were informed that it was not required for the study and was therefore entirely optional.

2.3.9 Data Analysis

Quantitative data from all assessments and surveys were analyzed using Pearson's bivariate correlations and ANCOVAs in SPSS v28.0.1.0 (IBM Corp., Armonk, NY). Where significance was observed, post hoc comparisons were carried out using Fisher's least significant difference (LSD).

3.0 Results

Data were collected from 119 participants across the 6 learning conditions with approximately equal numbers of participants assigned to each condition (Table 1). Outliers were identified based on box plots generated in SPSS (Figure 10) and then trimmed from the data set using the Outlier Labelling Rule at 2.2 times the interquartile range (IQR) as this is considered to yield a more accurate representation of the sample than trimming at 1.5 times the IQR (Hoaglin & Iglewicz, 1987).

Table 1. Number of participants recruited for each of six learning conditions. Sample sizes are shown prior to trimming outliers (total n = 119).

		Environment (n)	
		Clinical	Blackout
Modality (n)	3D Printed Models	20	20
	Computer	20	19
	Virtual Reality	20	20

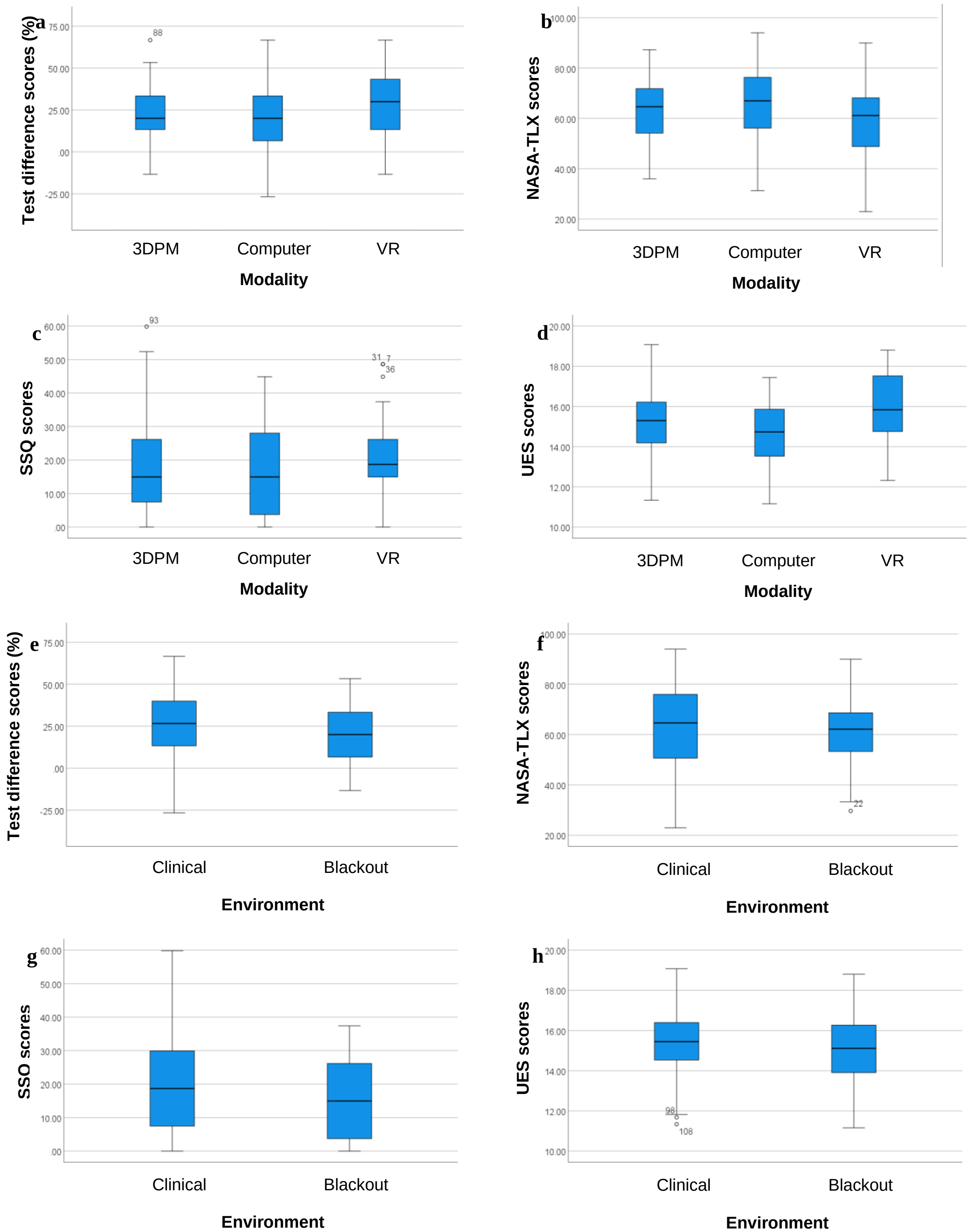


Figure 10. Box plots showing distributions by modality and environment for test difference scores (a, e), NASA Task Load Index (NASA-TLX) scores (b, f), Simulator Sickness Questionnaire (SSQ) scores (c, g), and User Engagement Scale (UES) scores (d, h) (n = 119).

Listwise deletion was carried out and only cases without any missing values were included in the analysis (n = 111). Demographic data were collected in the screening survey and are reported in Table 2. Of the 111 participants, 83 self-identified as female (75%) and 28 as male (25%). The majority of participants identified as first-year (n = 98, 88%), undergraduate n =110, (99%) students aged 18-20 years (n = 101, 91%). All faculties were represented in the final participant pool, with most participants coming from the Faculty of Science (n = 53, 48%) and the Faculty of Social Sciences (n = 24, 22%).

Table 2. Demographic data for participants (n = 111).

	n (%)
Sex	
Female	83 (75%)
Male	28 (25%)
Age (years)	
18-20	101 (91%)
21-23	8 (7%)
24-26	2 (2%)
Current degree	
Bachelors	110 (99%)
Masters	1 (1%)
Year of study	
Year 1	98 (88%)
Year 2	5 (5%)
Year 3	5 (5%)
Year 4	3 (2%)
Faculty/program	
Arts and Science	3 (2%)
Faculty of Science	53 (48%)
Faculty of Health Science	13 (12%)
Faculty of Social Sciences	24 (22%)
Faculty of Humanities	4 (4%)
DeGroote School of Business	3 (2%)
Faculty of Engineering	11 (10%)

Two-way ANCOVAs were conducted to evaluate differences in outcome variables (test difference scores, workload, cybersickness, and user engagement) based on learning modality (VR, 3DPM, computer) and learning environment (clinical, blackout).

The use of ANOVA-based statistics generally assumes independent observations, normally distributed dependent variables, and homogenous population variances (Howell, 2010). Data distributions were evaluated using histograms generated in SPSS (Appendix N, Figure 11). Assumptions of homogenous population variances were confirmed based on Levene's test ($p > 0.05$; Appendix N, Table 3). Assumptions of normality were violated for some of the data based on Shapiro-Wilk's test ($p < 0.05$; Appendix N, Table 4), however, ANOVA-based statistics are known to be robust against these violations when sample size is approximately equal between groups (Mardia, 1971). Sample sizes for the experimental groups are reported in Table 5 and descriptive statistics for the outcome variables are reported in Table 6.

Table 5. Sample sizes for the 6 learning conditions (total n = 111).

		Environment (n)		
		Clinical	Blackout	**Total**
Modality (n)	3D Printed Models	18 (32%)	18 (33%)	36 (32%)
	Computer	20 (35%)	19 (35%)	39 (36%)
	Virtual Reality	19 (33%)	17 (32%)	36 (32%)
	Total	57 (51%)	54 (49%)	

Table 6. Means and standard deviations for test difference scores and NASA Task Load Index (NASA-TLX), Simulator Sickness Questionnaire (SSQ), User Engagement Scale (UES), and Mental Rotations Test (MRT) scores (total n = 111). Covariates appearing in the model are evaluated at the following values: MRT = 9.82, NASA-TLX = 61.83, SSQ = 17.89, UES = 15.21.

Variable	Learning Condition		Mean (SD)	n
Test difference score	Modality	3DPM	22.22 (17.16)	36
		Computer	19.15 (19.24)	39
		VR	28.33 (20.12)	36
	Environment	Clinical	25.96 (20.65)	57
		Blackout	20.12 (17.01)	54
NASA-TLX	Modality	3DPM	62.83 (13.95)	36
		Computer	64.97 (15.05)	39
		VR	57.43 (17.00)	36
	Environment	Clinical	62.50 (17.94)	57
		Blackout	61.13 (12.72)	54
SSQ	Modality	3DPM	17.56 (15.79)	36
		Computer	15.92 (13.23)	39
		VR	20.36 (12.53)	36
	Environment	Clinical	20.73 (15.09)	57
		Blackout	14.89 (11.93)	54
UES	Modality	3DPM	15.22 (1.80)	36
		Computer	14.57 (1.59)	39
		VR	15.88 (1.73)	36
	Environment	Clinical	15.30 (1.78)	57
		Blackout	15.11 (1.78)	54

3.1 Pearson bivariate correlations

Pearson's bivariate correlation was used to determine if any relationships existed between MRT, test scores, NASA-TLX, SSQ, UES, and the Stereo Fly Test scores (Table 7). The analysis revealed the following significant correlations: a moderate, positive correlation between MRT and test scores ($r = 0.309$, $p = 0.001$); a weak, positive correlation between test scores and stereoacuity ($r = 0.236$, $p = 0.013$); and a weak, negative correlation between UES and SSQ scores ($r = -0.250$, $p = 0.008$) (Figure 12).

Table 7. Pearson correlation values and significance for Mental Rotations Test (MRT), test score, NASA Task Load Index (NASA-TLX), Simulator Sickness Questionnaire (SSQ), User Engagement Scale (UES), and stereoacuity (n = 111). *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed). [a]Stereoacuity is an inverse measure (i.e., lower values indicate better stereoscopic acuity) therefore this is a positive correlation.

Variable		Test Score	MRT	NASA-TLX	SSQ	UES	Stereoacuity
MRT	Pearson Correlation	.309**	1	-0.176	0.097	-0.088	-0.166
	Sig. (2-tailed)	0.001		0.064	0.312	0.357	0.082
	n	111	111	111	111	111	111
Test Score	Pearson Correlation	1	.309**	-0.089	-0.027	0.093	-.236*[a]
	Sig. (2-tailed)		0.001	0.354	0.778	0.330	0.013
	n	111	111	111	111	111	111
NASA-TLX	Pearson Correlation	-0.089	-0.176	1	0.149	-0.169	-0.126
	Sig. (2-tailed)	0.354	0.064		0.120	0.076	0.189
	n	111	111	111	111	111	111
SSQ	Pearson Correlation	-0.027	0.097	0.149	1	-.250**	0.032
	Sig. (2-tailed)	0.778	0.312	0.120		0.008	0.742
	n	111	111	111	111	111	111
UES	Pearson Correlation	0.093	-0.088	-0.169	-.250**	1	0.044
	Sig. (2-tailed)	0.330	0.357	0.076	0.008		0.646
	n	111	111	111	111	111	111
Stereoacuity	Pearson Correlation	-.236*[a]	-0.166	-0.126	0.032	0.044	1
	Sig. (2-tailed)	0.013	0.082	0.189	0.742	0.646	
	n	111	111	111	111	111	111

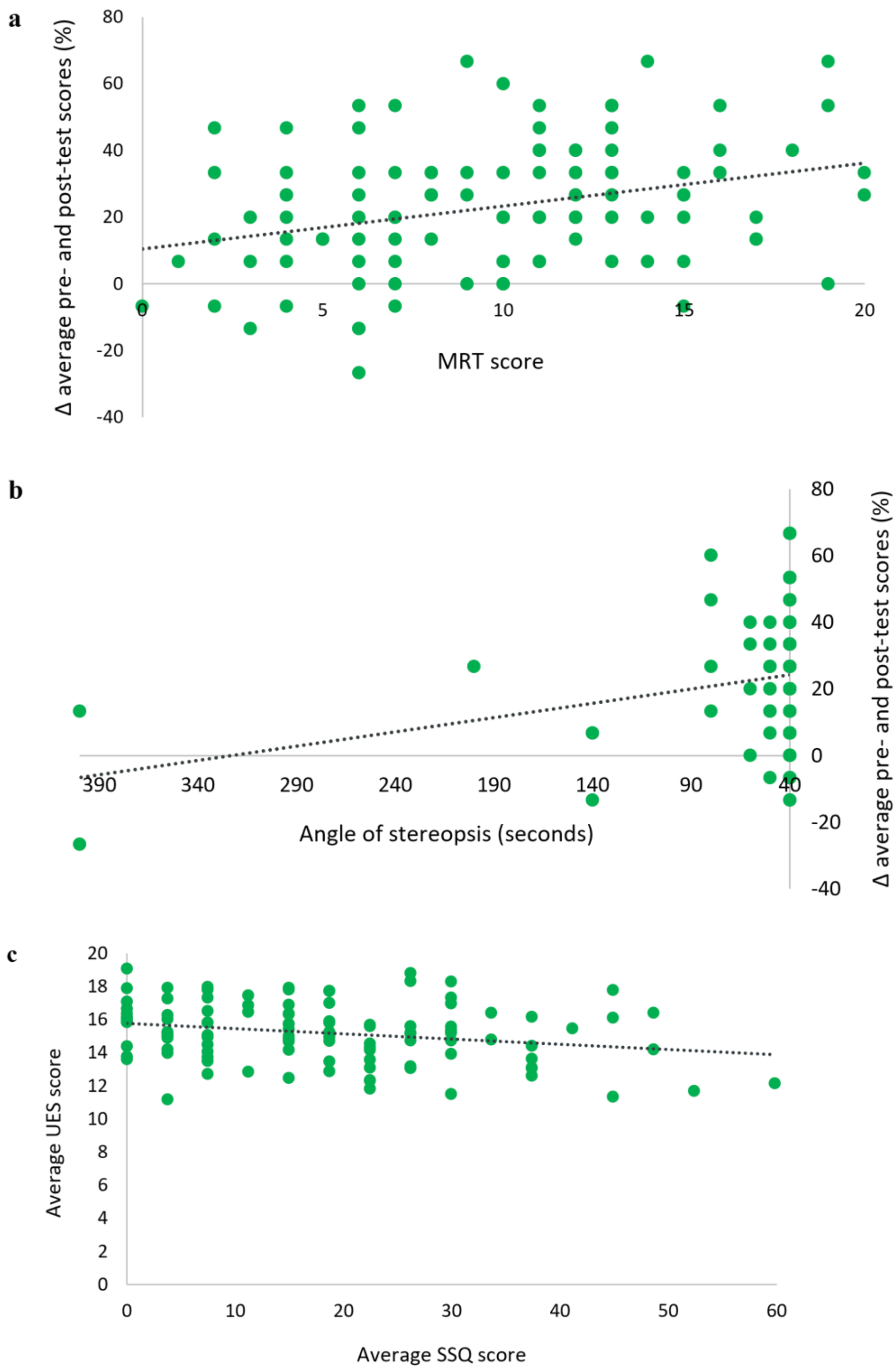


Figure 12. Scatterplots showing correlations between (a) Mental Rotations Test (MRT) score and test score (r = 0.309, p < 0.01), (b) stereoacuity and test score (r = 0.236, p < 0.05), and (c) Simulator Sickness (SSQ) score and User Engagement Scale (UES) score (r = -0.250, p < 0.01). n = 111 for all scatterplots. N.B. Stereoacuity is an inverse measure (i.e., lower values indicate better stereoscopic acuity) therefore this is a positive correlation.

3.2 Test Scores

Test scores for the 15-item pre-test and post-test assessments are reported as a percentage out of 100. As there was not a specific interest in how much learning occurred in each experimental group, but rather the variations in learning between groups, pre-test scores were subtracted from post-test scores, and these difference scores (as percentages) were used in the analysis.

Test difference scores were compared by modality and environment in a 3x2 ANCOVA where MRT, NASA-TLX, SSQ, and UES scores were used as covariates. Adjusted means (±SD) for test difference scores by modality were 23.42% (17.2), 18.91% (19.2), and 27.00% (20.1) for 3DPM, computer, and VR, respectively (Table 8). Adjusted means (±SD) for test difference scores by environment were 26.79% (20.7) and 19.42% (17.0) for clinical and blackout, respectively. Two-way ANCOVA results revealed a significant interaction between modality and environment [$F(2,101) = 4.08$, $p < 0.05$, $\eta^2 = 0.03$] (Appendix O, Table 9; Figure 13).

A simple main effect of environment was found within the VR groups – test difference scores for the clinical environment were significantly higher than for the blackout environment (mean difference = 17.27%, $p < 0.01$). No significant differences were found within the computer or 3DPM modalities.

A simple main effect of modality was found within the clinical environment: 3DPM test difference scores were significantly higher than computer test scores (mean difference = 12.98%, $p < 0.05$). Additionally, VR test scores were significantly higher than computer test scores (mean difference = 19.75%, $p < 0.01$). There were no significant differences within the blackout environment.

Table 8. Estimated marginal means and standard deviations for test difference scores. Covariates appearing in the model are evaluated at the following values: MRT = 9.82, NASA-TLX = 61.83, SSQ = 17.89, UES = 15.21.

	Mean	SD	n
Modality (n = 111)			
3D Printed Models	23.42	17.2	36
Computer	18.91	19.2	39
Virtual Reality	27.00	20.1	36
Environment (n = 111)			
Clinical	26.79	20.7	57
Blackout	19.42	17.0	54

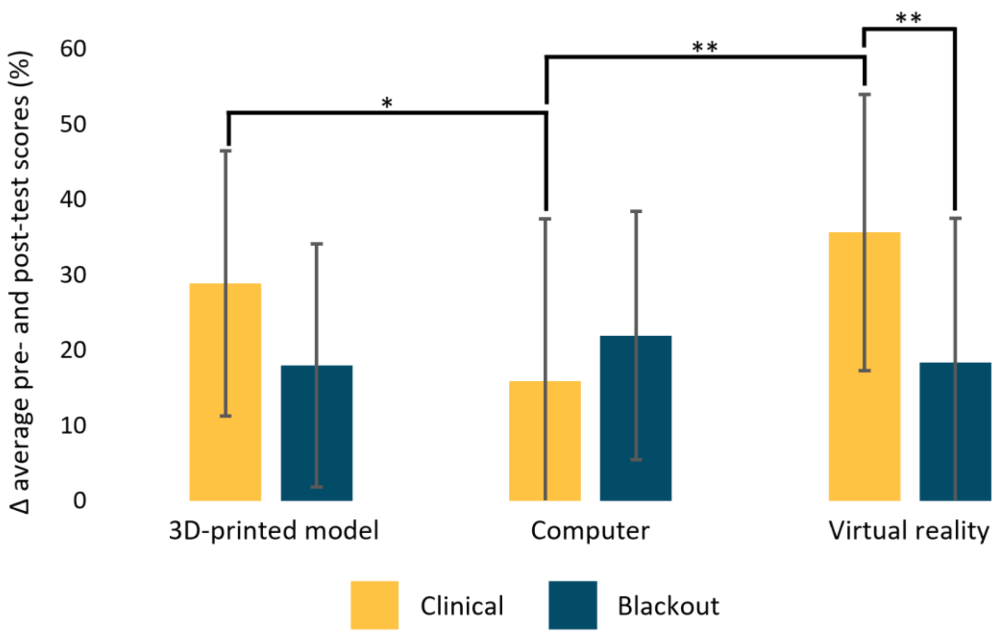


Figure 13. Mean test difference scores by modality and environment. Error bars show SD. $^{*}p < 0.05$. $^{**}p < 0.01$.

<u>3.3 Workload</u>

NASA-TLX scores were compared by modality and environment in a 3x2 ANCOVA where MRT, SSQ, and UES scores were used as covariates. Adjusted means (±SD) for NASA-TLX scores by modality were 62.32 (13.9), 64.69 (15.0), and 58.39 (17.0) for 3DPM, computer, and

VR, respectively (Table 10). Adjusted means (±SD) for NASA-TLX scores by environment were 61.82 (17.9) and 61.78 (12.7) for clinical and blackout, respectively. Two-way ANCOVA results revealed a significant interaction between modality and environment [$F(2,102) = 3.89$, $p < 0.05$, $\eta^2 = 0.0037$] (Appendix O, Table 11; Figure 14).

A simple main effect of environment was found within the computer modality: NASA-TLX scores were significantly higher for the clinical environment than the blackout environment (mean difference = 9.51, $p < 0.05$). No significant differences were found for the other two modalities.

A simple main effect of modality was also found within the clinical environment: NASA-TLX scores were significantly higher for computer than for VR (mean difference = 15.83, $p < 0.01$). No significant differences were found within the blackout environment.

Table 10. Estimated marginal means and standard deviations for NASA-TLX. Covariates appearing in the model are evaluated at the following values: MRT = 9.82, SSQ = 17.89, UES = 15.21.

	Mean	SD	n
Modality (n = 111)			
3D Printed Models	62.32	13.9	36
Computer	64.69	15.0	39
Virtual Reality	58.39	17.0	36
Environment (n = 111)			
Clinical	61.82	17.9	57
Blackout	61.78	12.7	54

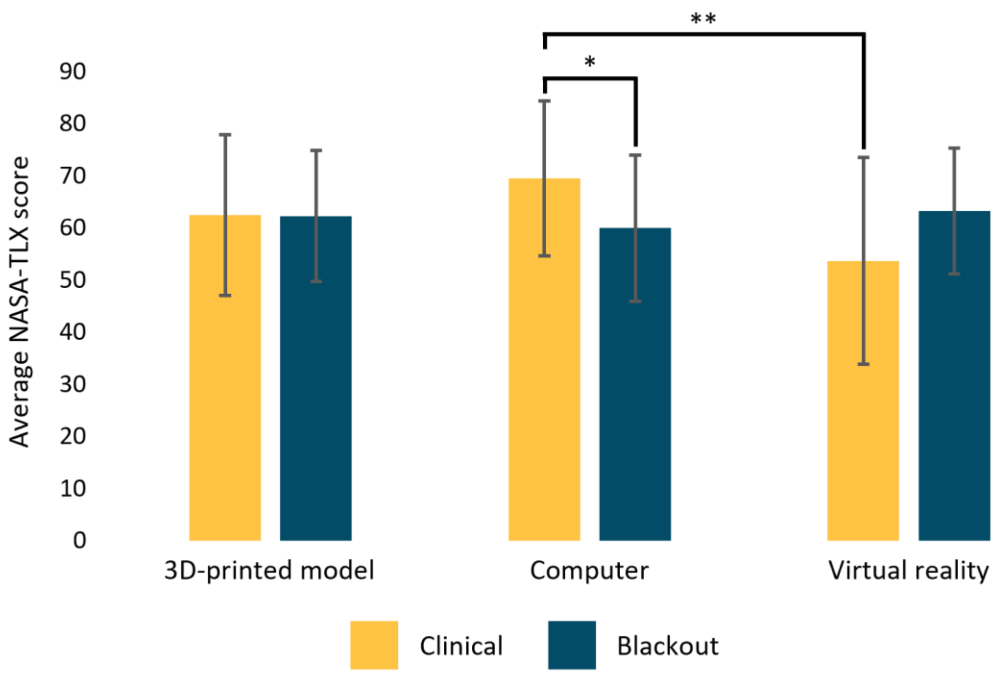


Figure 14. Mean NASA-TLX scores by modality and environment. Error bars show SD. $^{*}p < 0.05$. $^{**}p < 0.01$.

3.4 Cybersickness

SSQ scores were compared by modality and environment in a 3x2 ANCOVA where MRT, NASA-TLX, and UES scores were used as covariates. Adjusted means (±SD) for SSQ scores by modality were 17.67 (15.8), 14.05 (13.3), and 22.07 (12.5) for 3DPM, computer, and VR, respectively (Table 12). Adjusted means (±SD) for SSQ scores by environment were 21.01 (15.1) and 14.85 (11.9) for clinical and blackout, respectively. Two-way ANCOVA results revealed no significant interaction between modality and environment ($p > 0.05$) (Figure 15a; Appendix O, Table 13). Significant main effects of modality [$F(2,101) = 3.181$, $p < 0.05$, $\eta^2 = 0.018$] and of environment [$F(1,102) = 6.34$, $p < 0.05$, $\eta^2 = 0.018$] were observed.

Fisher's LSD post-hoc test revealed that SSQ scores were significantly higher for VR than for computer (mean difference = 8.02, $p < 0.05$) (Figure 15b) and significantly higher for clinical than for blackout (mean difference = 6.16, $p < 0.05$) (Figure 15c).

Table 12. Estimated marginal means and standard deviations for SSQ. Covariates appearing in the model are evaluated at the following values: MRT = 9.82, UES = 15.21, NASA-TLX = 61.83.

	Mean	SD	n
Modality (n = 111)			
3D Printed Models	17.67	15.8	36
Computer	14.05	13.2	39
Virtual Reality	22.07	12.5	36
Environment (n = 111)			
Clinical	21.01	15.1	57
Blackout	14.85	11.9	54

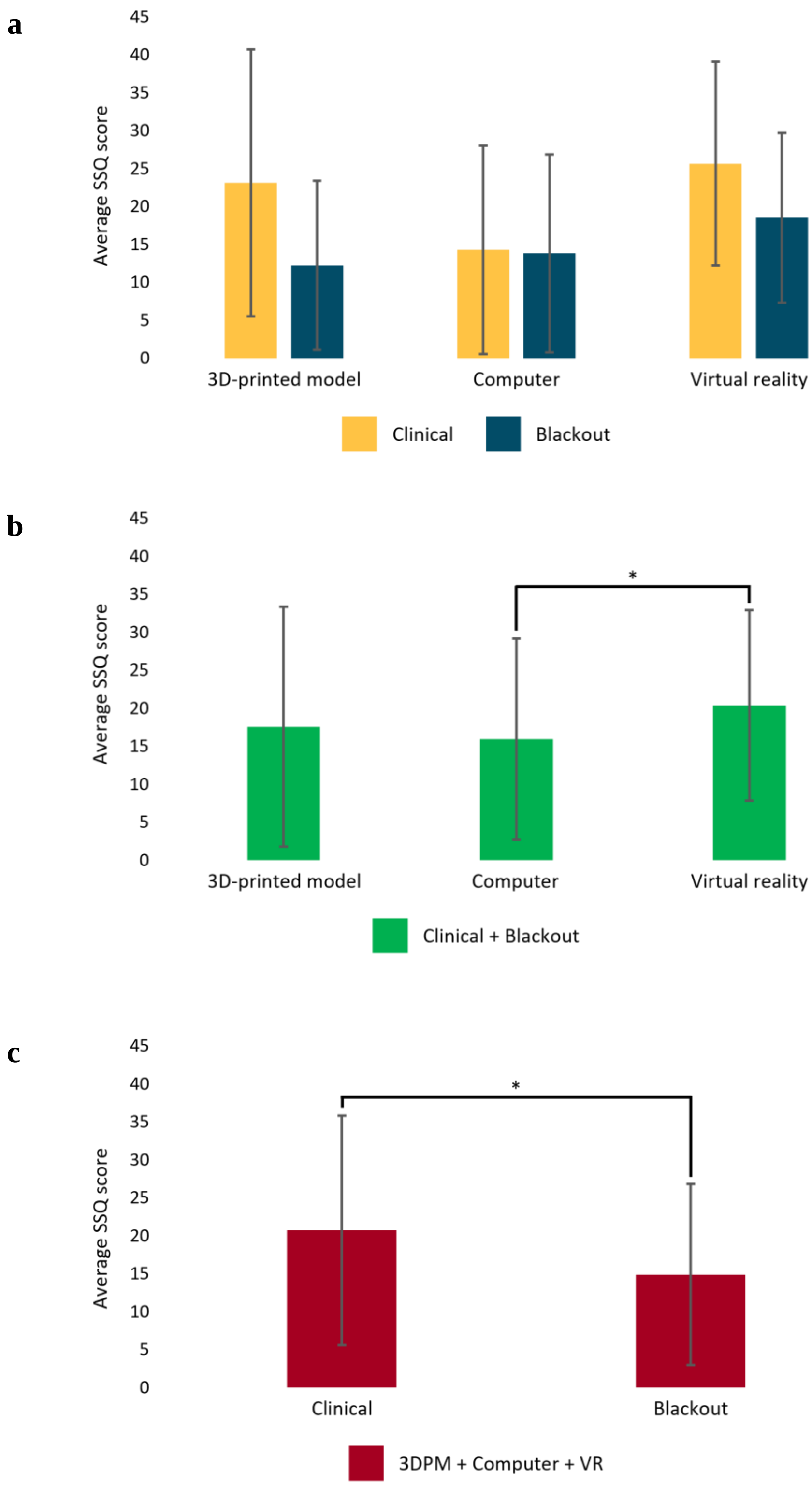


Figure 15. Mean SSQ scores by (a) modality and environment, (b) modality across both environments, (c) environment across all modalities. Error bars show SD. $^{*}p < 0.05$.

3.5 User Engagement

UES scores were compared by modality and environment in a 3x2 ANCOVA where MRT, NASA-TLX, and SSQ scores were used as covariates. Adjusted means (±SD) for UES scores by modality were 15.17 (1.80), 14.53 (1.59), and 15.96 (1.73) for 3DPM, computer, and VR, respectively (Table 14). Adjusted means (±SD) for UES scores by environment were 15.41 (1.78) and 15.03 (1.78) for clinical and blackout, respectively. Two-way ANCOVA results revealed no significant interaction between modality and environment ($p > 0.05$) (Appendix O, Table 15). A significant main effect of modality [$F(2,102) = 6.67$, $p < 0.01$, $\eta^2 = 0.0014$] was observed.

Fisher's LSD post-hoc test revealed that UES scores were significantly higher for VR than for computer (mean difference = 1.43, $p < 0.001$) (Figure 16).

Table 14. Estimated marginal means and standard deviations for UES. Covariates appearing in the model are evaluated at the following values: MRT = 9.82, NASA-TLX = 61.83, SSQ = 17.89.

	Mean	SD	n
Modality (n = 111)			
3D Printed Models	15.17	1.80	36
Computer	14.53	1.59	39
Virtual Reality	15.96	1.73	36
Environment (n = 111)			
Clinical	15.41	1.78	57
Blackout	15.03	1.78	54

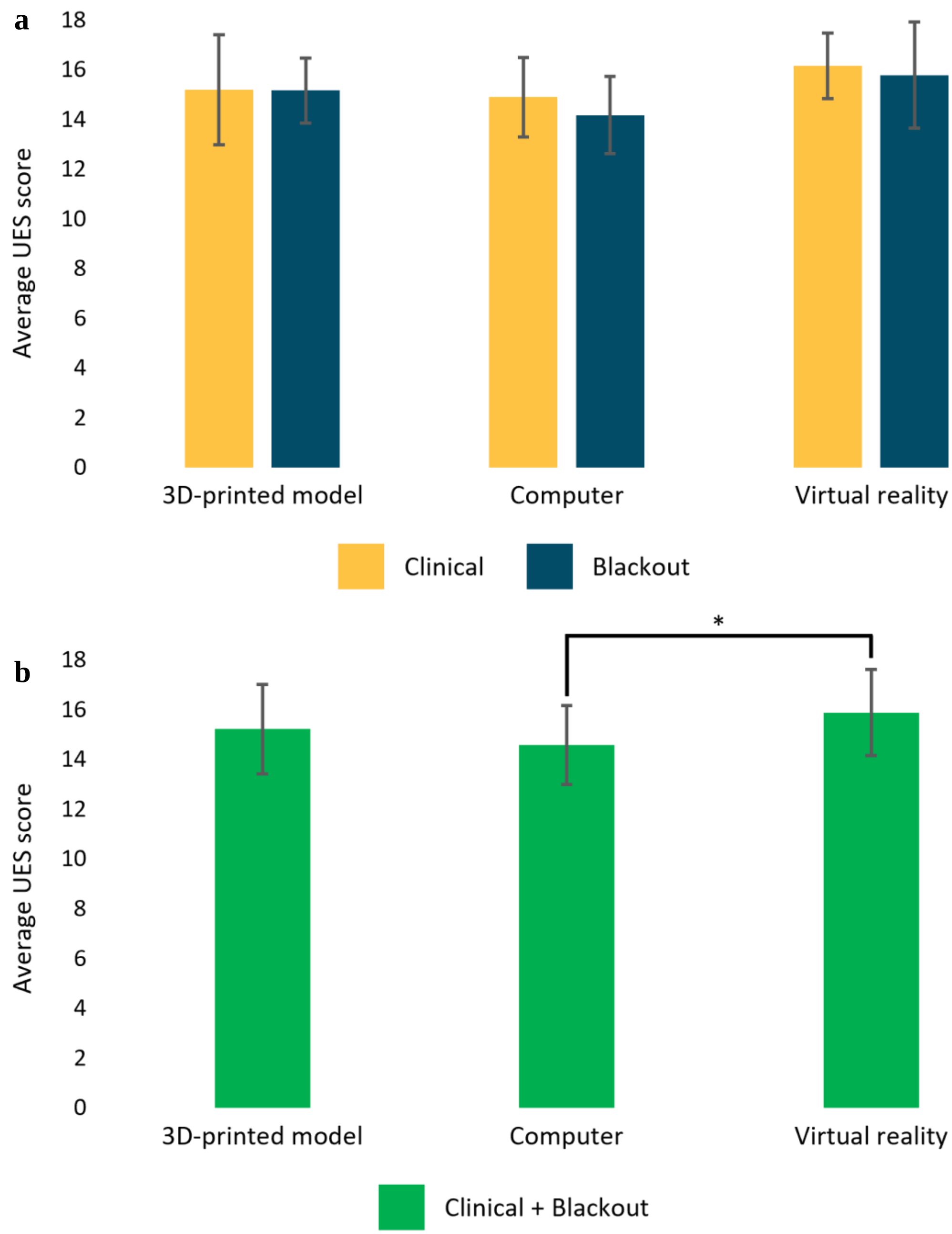


Figure 16. Mean UES scores by (a) modality and environment, (b) modality across both environments. Error bars show SD. $^{*}p < 0.001$.

4.0 Discussion

4.1 Relationships between test scores, perceived workload, cybersickness, user engagement, mental rotation ability, and stereoscopic discriminability

We aimed to explore and understand the relationships between test performance, workload, cybersickness, engagement, mental rotation ability, and stereoscopic discriminability, considering their known correlations (Birbara et al., 2020; Kim et al., 2022; Sepich et al., 2022; Servotte et al., 2020; Skulmowski & Rey, 2017; Weech et al., 2019). This study integrated the findings from a correlation analysis into broader analyses.

A moderate, positive correlation between MRT and test scores was observed and consistent with existing literature on anatomy education (Brewer et al., 2012). While MRT can be improved through training (Hoyek et al., 2009), it remains an inherent trait of the learner, requiring consideration of individual differences when selecting teaching tools and modalities.

Stereoacuity exhibited a weak, positive correlation with test scores. While there is significant evidence to support the importance of stereoscopy for learning (Bogomolova et al., 2023; Brewer et al., 2012; Khot et al., 2013; Wainman et al., 2020), less research has been conducted on specifically exploring the impact of stereoacuity on test performance. In surgical education, it has been found that low stereoacuity is correlated with inferior performance of surgical tasks (Burgess et al., 2021), and more broadly, research has shown that children with poor stereoacuity often have poor literacy skills (Ponsonby et al., 2013). As is the case with mental rotation ability, stereoacuity can be improved through training to some extent (Ponsonby et al., 2009; Xi et al., 2021). The implications of these findings support the need for further research to better understand the link between stereoacuity, learning, and performance.

Our analysis also revealed a weak, negative correlation between UES and SSQ scores, suggesting a potential inverse relationship between engagement and cybersickness. While the constructs of engagement and presence are different, they are often encompassed more broadly within the context of embodied learning experiences in VR (Johnson-Glenberg, 2018; Lindgren et al., 2016; Skulmowski & Rey, 2018). Cybersickness and sense of presence have been shown to be negatively correlated (Weech et al., 2019).

Interestingly, test scores did not significantly correlate with workload, cybersickness, or engagement. However, experiences causing discomfort and high workload can still influence learners' interest and motivation, highlighting the importance of addressing these factors in educational design.

4.2 Test scores

Variability in test scores across modalities was influenced by the learning environment. Participants in the VR clinical environment achieved higher test scores than those in the blackout environment. The differences in test scores between the clinical and blackout environments could be attributed to two potential factors: the clinical environment having a beneficial impact on learning, or the blackout environment having a negative impact. Our initial hypothesis proposed that the enriched clinical environment would enhance GCL, thereby increasing test performance within the VR modality. However, it is possible that the absence of any environment in the blackout scenario decreased the GCL, or perhaps even increased the ECL. As humans, we naturally expect certain environmental cues, which can be manipulated in VR and could potentially affect performance. The emptiness of the blackout environment is much like being in a vacuum, with no spatial references. This may have influenced cognitive processes.

Future studies should be done to determine whether higher test scores can be attributed to contextually relevant environments or if simpler settings that maintain spatial references, like an empty classroom, are sufficient. Furthermore, exploring different settings for VRLEs may also provide insight into their impact on cognitive load and workload.

While the physical environment also provided immersion, it lacked the specific design intent of the VR environment. In fact, participants in the clinical condition often perceived the setting as ordinary, indicating a subconscious acknowledgement of the tailored VR environment's relevance to the task.

Additionally, a number of participants scored lower on the post-test compared to the pre-test, despite careful protocol adherence. External influences which are beyond the researcher's control such as the motivation, attention, mood, and interest of participants, may have affected test scores. Given the participants' lack of prior anatomy training, it is reasonable to assume that pre-test scores should have been lower than post-test scores. It is also worth noting that undergraduate students in the Sciences at McMaster frequently take courses in medical terminology which cover Greek and Latin roots and may have impacted test performance due to familiarity with the terminology. While these factors may have skewed the results and diminished certain differences, they also offer ecological validity to our findings. External influences like motivation, attention, mood, and interest are inherent to the learning experience, challenging instructors' ability to manage them effectively.

4.3 Workload

Our study revealed that perceived workload varied depending on both the modality and environment. We initially predicted that workload would be higher for the clinical environment compared to the blackout environment due to the presence of additional visual stimuli. However, this only proved to be true in the computer modality.

Furthermore, within the clinical context, workload was significantly lower with the use of the VR module than with the computer module. This difference may be explained by the lack of stereoscopic input with the 2D computer presentation, which would require additional visuospatial resources to create mental three-dimensional representations of the pelvic models.

Despite variations between experimental groups, it is important to assess where the average NASA-TLX scores fall within the range of possible scores. Understanding acceptability of workload is a critical consideration for designing educational technologies. Grier (2015) conducted a meta-analysis of 237 studies which employed the NASA-TLX, with the aim of creating guidelines on how to interpret scores across disciplines and types of tasks (Grier, 2015). The types of tasks from the studies were categorized and data were analyzed to determine the minimum and maximum scores reported in the literature, along with benchmarks for each quartile. The task of learning from the cAnatomy module could be considered to fall into the "cognitive tasks" category defined by Grier. In this category, the minimum NASA-TLX score reported was 13.08 and the maximum 64.90. The 75th percentile was demarcated by a score of 54.66.

Most of the averages observed in this study were close to the maximum score within the cognitive tasks category. This raises major concerns regarding the use of the cAnatomy module, irrespective of the modality or environment. However, it is essential to consider the background of participants and whether the perceived workload can be considered acceptable. Our participants were predominantly first-year undergraduate students with no formal anatomy education. While such a high level of perceived workload may pose challenges for novice learners, it may be acceptable for more experienced learners. Interestingly, we found no significant correlation between NASA-TLX scores and test scores.

4.4 Cybersickness

Our results show a significant main effect of both modality and environment. As expected, SSQ scores were significantly higher in the VR condition compared to the computer condition. SSQ scores higher than 20 are considered sufficient cause for redesigning simulations (Kennedy et al., 2003). With a mean SSQ score of 22.07, the VR group falls into this problematic range. The computer condition had an average SSQ score of 14.05, indicating significant symptoms of cybersickness (Kennedy et al., 2003).

Surprisingly, the VR and computer SSQ scores were not significantly higher than the 3DPM scores. While we expected no cybersickness in the 3DPM condition, participants may have experienced baseline discomfort unrelated to the learning module. Because they were being explicitly asked to comment on symptoms, it may have drawn more attention to the baseline experiences, inflating the average SSQ score.

Another possible explanation is that in our attempts to recreate the VR version of the module, there may have been unforeseen physical triggers for symptoms of motion sickness rather than cybersickness. In the VR module, the learner was able to stand in place and use the handheld controller to click back and forth between scenes. In the 3DPM condition, we attempted to mimic this experience as closely as possible with the use of turntables so that participants could rotate forward or backward between scenes and view structures from different angles. Though the speed of rotation was relatively slow and the participant had agency and control over the rotation, it is possible that this caused visually induced motion sickness (VIMS) which shares symptoms of cybersickness that the SSQ would have been sensitive to (Kennedy et al., 2010).

A significant main effect of environment was also found, with SSQ scores significantly higher in the clinical environment than in the blackout environment, as hypothesized. Weech and colleagues reported a positive correlation between workload and cybersickness (Weech et al., 2019), forming the basis for our hypothesis that the clinical environment would induce stronger cybersickness due to increased workload. However, we did not observe a significant main effect of environment for NASA-TLX scores. Notably, Saredakis and colleagues found that scenic VR content, comparable to our clinical environment, resulted in lower SSQ scores compared to minimalist content, comparable to our blackout environment (Saredakis et al., 2020).

The standardized phrasing of items on validated test instruments ensures consistency, but this also means that when complex or ambiguous phrasing is used, researchers must intervene to answer questions from participants. This must be done mindfully as the researcher's explanations may influence participant responses. We found this to be particularly true with the SSQ, NASA-TLX, and UES. Our study encountered participant questions regarding terms such as "stomach awareness", and "vertigo", indicating the need for commonplace language in questionnaire items. Additionally, distinctions between terms like "difficulty focusing" and "difficulty concentrating" required clarification. Challenges in interpreting SSQ symptoms may explain some of the unexpected results.

4.5 User engagement

Our results showed that UES scores were significantly higher for VR than for computer, however, the effect size was negligible. We expected that VR and 3DPMs would have been significantly more engaging than the computer condition due to the greater level of immersion, presence, and interactivity. That there was no meaningful difference when taking effect size into account suggests that the module design itself may have been the predominant contributor to high engagement. This discrepancy aligns with findings by Johnson-Glenberg and colleagues who observed similar misalignments between hypotheses and results regarding the effectiveness of different technology modalities (Johnson-Glenberg et al., 2021).

The lack of difference in UES scores between the clinical and blackout environments also suggests that visually complex and contextually relevant environments may not significantly enhance learner engagement. In fact, knowing that the clinical environment increased workload when the module was delivered via computer highlights the need to reconsider the use of highly visually stimulating virtual environments when the primary objective is knowledge acquisition.

It should also be taken into consideration that despite efforts to match the physical environment to the virtual environment as closely as possible, they were not identical. Further exploration with more similar environments is warranted.

4.6 Limitations

4.6.1 Sampling biases and generalizability

There were a few sources of sampling bias that must be considered when interpreting the results of the study. First, a volunteer bias was present as there was some level of self-selection with participants. Participants who signed up for our study would have at the very least skimmed our study description and wished to participate based on what they read. Because there was no cash incentive for our study, it is reasonable to expect that those who chose to sign up for our study over other actively recruiting studies did so because of personal interest in either the content or the methods described in our study description.

Though the sample of participants was meant to reflect the average university student population, given that recruitment was conducted almost entirely through the online recruitment portal, nearly all participants were students enrolled in an undergraduate psychology course. This meant that most participants were first year, undergraduate students from the Faculty of Science. While having previously taken anatomy courses was an exclusion criterion, it is likely that a significant portion of the students had some experience with basic biological sciences.

Another consideration for sampling bias pertains to the timeline of data collection. Due to supply chain issues during the COVID-19 pandemic, our access to experimental equipment was delayed. As a result, we used a block randomization technique, but were not able to randomly assign all participants to all six experimental groups right at the start of data collection. Instead, the first set of participants were all assigned to the VR modality, but were randomly assigned to either the clinical or blackout conditions. We then moved on to the computer modality, assigning randomly to environments in the same way, and finally ended with the 3DPM modality. In total, data collection occurred over a period of approximately 1.5 years. The development and rigorous testing of our standard operating procedures for the experiment helped to ensure consistency in the way experiments were carried out over time and also consistency between experimenters, however, invariably with such a long data collection period, some error is difficult to avoid.

Since we began data collection in the midst of the COVID-19 pandemic, mask mandates were in place at the University. Our research team was trained to conduct the experiments and fit VR headsets so that the participant's mask was as unobtrusive as possible. Despite these efforts, there may have been some effect of discomfort with the use of masks under a VR headset.

Our sample consisted of a significant discrepancy in the number of male participants (n = 28) and female participants (n = 83). There are many reports in the literature of sex differences in mental rotation ability, visuospatial ability, experience and comfort with VR, susceptibility to cybersickness/simulator sickness/motion sickness, ocular qualities, strategies for taking tests, and responses to high workload and cognitive stimulation (Goldstein et al., 1990; Saredakis et al., 2020; Stanney et al., 2020; Vandenberg & Kuse, 1978; Voyer, 1997). Because of the large disparity, we did not use sex as a covariate in our analyses, but it is quite possible that there was an impact on the outcome variables investigated in this study.

Given the sampling biases outlined, the generalizability of our results may be reduced. Further studies should be conducted, ensuring that there are equal numbers of male and female participants, and that data are collected on a shorter timeline to see if there are marked differences in the outcome variables or if the effect sizes are larger than reported in this thesis.

4.6.2 Adoption of measurement tools for multiple modalities

The use of a full factorial experimental design allowed us to evaluate the effects of both the learning environment and modality with greater efficiency than addressing individual factors. However, a drawback to this design is that the appropriateness of the measures selected to evaluate the outcome variables may have been compromised in an attempt to standardize measurement across all three learning modalities. For example, the outcome variable of workload needed to be measured using a tool that was appropriate for all 3 modalities. The research team had considered using the Simulation Task Load Index (SIM-TLX), however, this would only have been appropriate for the VR modality. Since we were also working with the computer and 3DPMs, we selected the NASA-TLX, despite the fact that the SIM-TLX may have been a more suitable option for the VR modality.

Similarly, the SSQ is best suited to measure cybersickness in VR modalities and was also considered suitable for the computer modality. Unexpectedly, our results indicate that participants in the 3DPM groups reported cybersickness scores that were comparable to those in the VR and computer groups. This discrepancy may point to the presence of VIMS with our 3DPM setup, or it may simply emphasize that the SSQ is not appropriate for use with non-digital modalities.

The UES may also have not been an appropriate selection for measuring user engagement across all modalities. While originally designed to evaluate user engagement in various technologies, its application in the 3DPM condition required participants to adapt the scale to their context. Despite this challenge, adjustments to the scale’s phrasing of the original 31 items were not considered as this would have put the validity of the tool into question.

5.0 Conclusion

In determining the appropriate learning tools for classrooms and labs, it is crucial to select tools that align with the learning objectives rather than forcing learning objectives to fit the tools. Even the best technology in the hands of the best students will yield inadequate results if not implemented in accordance with pedagogical principles. Learning anatomy is a complex process, encompassing various facets, including declarative knowledge of anatomical structures, implicit knowledge of studying and applying new information, and tacit knowledge of ethical conduct and appreciation for the human body. Consequently, no single tool can fulfill every learning outcome. While this understanding is prevalent in the literature (Estai & Bunt, 2016; Kerby et al., 2011; Yammine, 2014), there remains a significant gap between intention and action. So long as an adversarial perspective persists in the selection of tools for curriculum design, anatomy education will continue to fall short of its potential.

To address this challenge, reframing our approach to researching these technologies is key. Rather than questioning whether a tool is suitable for teaching anatomy, or if it surpasses others, we should inquire whether it effectively achieves specific learning objectives.

Bloom's Revised Taxonomy provides a framework for creating learning objectives and determining the best instructional methods to achieve them, thereby informing effective instructional design (Anderson et al., 2001). Within the cognitive, affective, and psychomotor domains, various instructional approaches contribute to learning anatomy. Didactic teaching and textbook diagrams are effective for foundational knowledge, while self-guided tools such as VR applications enhance understanding and facilitate visualization. Learning through cadaveric dissection remains essential for promoting respect, reflection, and ethical conduct, though different technologies may better support foundational anatomy learning.

This study explored the effects of learning modalities and environments on test score, workload, cybersickness, and engagement, considering the influence of mental rotation ability and stereoacuity. A moderate correlation between MRT and test scores, and weak correlations between test scores and stereoacuity, and between engagement and cybersickness were observed. Significant interactions between modality and environment were noted for both test scores and workload, but with small effect sizes. While significant main effects of modality and environment were evident for cybersickness, and a significant main effect of modality for engagement, the effect sizes remained small.

Implications of these findings reveal that workload and cybersickness scores exceeded acceptable ranges across all learning conditions, while engagement scores remained consistently high. The lack of meaningful differences suggests that the factors we prioritize may not be as crucial as previously thought. Instead, emphasis should be placed on the design and development of content itself rather than solely on its presentation. While the overall findings may seem somewhat unremarkable, they are undoubtedly liberating. There is no imperative to adhere to any specific technology, and the pressure to find a singular solution can be alleviated. Embracing a multifaceted approach to curriculum design, where learning objectives and activities are aligned, appears key to improving anatomy education.

References

Aksakal, N. (2015). Theoretical View to The Approach of The Edutainment. *Procedia - Social and Behavioral Sciences*, *186*, 1232–1239. https://doi.org/10.1016/j.sbspro.2015.04.081

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556–559.

Birbara, N. S., Sammut, C., & Pather, N. (2020). Virtual Reality in Anatomy: A Pilot Study Evaluating Different Delivery Modalities. *Anatomical Sciences Education*, *13*(4), 445–457. https://doi.org/10.1002/ase.1921

Bogomolova, K., Vorstenbosch, M. A. T. M., El Messaoudi, I., Holla, M., Hovius, S. E. R., van der Hage, J. A., & Hierck, B. P. (2023). Effect of binocular disparity on learning anatomy with stereoscopic augmented reality visualization: A double center randomized controlled trial. *Anatomical Sciences Education*, *16*(1), 87–98. https://doi.org/10.1002/ase.2164

Brewer, D. N., Wilson, T. D., Eagleson, R., & de, S. (2012). Evaluation of Neuroanatomical Training using a 3D Visual Reality Model. In *Medicine Meets Virtual Reality 19* (Vol. 173, pp. 85–91).

Brewer-Deluce, D., Bak, A. B., Simms, A. J., Sinha, S., Mitchell, J. P., Shin, D., Saraco, A. N., & Wainman, B. C. (2021). Virtual Reality Bell-Ringer: The Development and Testing of a Stereoscopic Application for Human Gross Anatomy. *Anatomical Sciences Education*, *14*(3), 330–341. https://doi.org/10.1002/ase.2074

Burgess, S., Kousha, O., Khalil, M., Gilmour, C., MacEwen, C. J., & Gillan, S. N. (2021). Impact of stereoacuity on simulated cataract surgery ability. *Eye*, *35*(11), Article 11. https://doi.org/10.1038/s41433-020-01346-4

Chen, R., Grierson, L. E., & Norman, G. R. (2015). Evaluating the impact of high- and low-fidelity instruction in the development of auscultation skills. *Medical Education*, *49*(3), 276–285. https://doi.org/10.1111/medu.12653

Cheung, C. C., Bridges, S. M., & Tipoe, G. L. (2021). Why is Anatomy Difficult to Learn? The Implications for Undergraduate Medical Curricula. *Anatomical Sciences Education*, *14*(6), 752–763. https://doi.org/10.1002/ase.2071

Choi, H.-H., van Merrienboer, J. J. G., & Paas, F. (2014). Effects of the Physical Environment on Cognitive Load and Learning: Towards a New Model of Cognitive Load. *Educational Psychology Review*, *26*(2), 225–245. https://doi.org/10.1007/s10648-014-9262-6

Churchill, J. (2022). *cAnatomy Female Pelvic Floor* (2.49.1) [Computer software]. Zygote.

Darras, K. E., de Bruin, A. B. H., Nicolaou, S., Dahlström, N., Persson, A., van Merriënboer, J., & Forster, B. B. (2018). Is there a superior simulator for human anatomy education? How virtual dissection can overcome the anatomic and pedagogic limitations of cadaveric dissection. *Medical Teacher*, *40*(7), 752–753. https://doi.org/10.1080/0142159X.2018.1451629

Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective. *Performance Improvement Quarterly*, *26*(2), 43–71. https://doi.org/10.1002/piq.21143

Estai, M., & Bunt, S. (2016). Best teaching practices in anatomy education: A critical review. *Annals of Anatomy - Anatomischer Anzeiger*, *208*, 151–157. https://doi.org/10.1016/j.aanat.2016.02.010

Farmani, Y., & Teather, R. J. (2020). Evaluating discrete viewpoint control to reduce cybersickness in virtual reality. *Virtual Reality*, *24*(4), 645–664. https://doi.org/10.1007/s10055-020-00425-x

Fincher-Kiefer, R. (2019). An Introduction to the Theory of Embodied Cognition. In *How the body shapes knowledge: Empirical support for embodied cognition.* American Psychological Association. https://doi.org/10.1037/0000136-000

Finn, G. M., Danquah, A., & Matthan, J. (2022). Colonization, cadavers, and color: Considering decolonization of anatomy curricula. *Anatomical Record (Hoboken, N.j. : 2007)*, *305*(4), 938–951. https://doi.org/10.1002/ar.24855

Garg, A., Norman, G. R., Spero, L., & Maheshwari, P. (1999). Do virtual computer models hinder anatomy learning? *Academic Medicine*, *74*(10), S87-9. https://doi.org/10.1097/00001888-199910000-00049

Goldin-Meadow, S. (2011). Learning through gesture. *WIREs Cognitive Science*, *2*(6), 595–607. https://doi.org/10.1002/wcs.132

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition*, *18*(5), 546–550. https://doi.org/10.3758/BF03198487

Groth, C., Tauscher, J.-P., Heesen, N., Castillo, S., & Magnor, M. (2021). Visual Techniques to Reduce Cybersickness in Virtual Reality. *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 486–487. https://doi.org/10.1109/VRW52623.2021.00125

Hegarty, M., & Kriz, S. (2008). Effects of knowledge and spatial ability on learning from animation. In *Learning with Animation: Research Implications for Design* (pp. 3–29). Cambridge University Press.

Hettinger, L. J., Berbaum, K. S., Kennedy, R. S., Dunlap, W. P., & Nolan, M. D. (1990). Vection and Simulator Sickness. *Military Psychology*, *2*(3), 171–181. https://doi.org/10.1207/s15327876mp0203_4

Howell, D. C. (2010). *Statistical Methods for Psychology* (7th ed.). Wadsworth, Cengage Learning.

Hoyek, N., Collet, C., Rastello, O., Fargier, P., Thiriet, P., & Guillot, A. (2009). Enhancement of Mental Rotation Abilities and Its Effect on Anatomy Learning. *Teaching and Learning in Medicine*, *21*(3), 201–206. https://doi.org/10.1080/10401330903014178

Johnson-Glenberg, M. C. (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. *Frontiers in Robotics and AI*, *5*, 81. https://doi.org/10.3389/frobt.2018.00081

Johnson-Glenberg, M. C., Bartolomea, H., & Kalina, E. (2021). Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *Journal of Computer Assisted Learning*, *37*(5), 1263–1284. https://doi.org/10.1111/jcal.12567

Kalet, A., Song, H., Sarpel, U., Schwartz, R., Brenner, J., Ark, T., & Plass, J. (2012). Just enough, but not too much interactivity leads to better clinical skills performance after a computer assisted learning module. *Medical Teacher*, *34*(10), 833–839. https://doi.org/10.3109/0142159X.2012.706727

Kennedy, R. S., Drexler, J., & Kennedy, R. C. (2010). Research in visually induced motion sickness. *Applied Ergonomics*, *41*(4), 494–503. https://doi.org/10.1016/j.apergo.2009.11.006

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. https://doi.org/10.1207/s15327108ijap0303_3

Kennedy, R. S., Stanney, K. M., & Lanham, D. S. (2003). Configural Scoring of Simulator Sickness, Cybersickness and Space Adaptation Syndrome: Similarities and Differences? In *Virtual and adaptive environments: Applications, implications, and human performance issues* (pp. 247–278). CRC Press.

Kerby, J., Shukur, Z. N., & Shalhoub, J. (2011). The relationships between learning outcomes and methods of teaching anatomy as perceived by medical students. *Clinical Anatomy*, *24*(4), 489–497. https://doi.org/10.1002/ca.21059

Khot, Z., Quinlan, K., Norman, G. R., & Wainman, B. (2013). The relative effectiveness of computer-based and traditional resources for education in anatomy: Virtual Reality Versus Reality in Anatomy. *Anatomical Sciences Education*, *6*(4), 211–215. https://doi.org/10.1002/ase.1355

Kim, S. Y., Park, H., Kim, H., Kim, J., & Seo, K. (2022). Technostress causes cognitive overload in high-stress people: Eye tracking analysis in a virtual kiosk test. *Information Processing & Management*, *59*(6), 103093. https://doi.org/10.1016/j.ipm.2022.103093

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education*, *95*, 174–187. https://doi.org/10.1016/j.compedu.2016.01.001

NATO Science and Technology Office. (2021). *Guidelines for Mitigating Cybersickness in Virtual Reality Systems* (Technical Report NATO STO-TR-HFM-MSG-323). Science and Technology Organization, North Atlantic Treaty Organization.

Nguyen, N., Mulla, A., Nelson, A. J., & Wilson, T. D. (2014). Visuospatial anatomy comprehension: The role of spatial visualization ability and problem-solving strategies: Spatial Anatomy Task Performance. *Anatomical Sciences Education*, *7*(4), 280–288. https://doi.org/10.1002/ase.1415

O'Brien, H. L., 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. https://doi.org/10.1016/j.ijhcs.2018.01.004

Okan, Z. (2003). Edutainment: Is learning at risk? *British Journal of Educational Technology*, *34*(3), 255–264. https://doi.org/10.1111/1467-8535.00325

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, *32*(1/2), 1–8.

Perkins, D., & Salomon, G. (1992). Transfer Of Learning. In *International Encyclopedia of Education* (Second Edition, Vol. 11). Pergamon Press.

Ponsonby, A.-L., Smith, K., Williamson, E., Bridge, D., Carmichael, A., Dwyer, T., Jacobs, A., & Keeffe, J. (2013). Poor Stereoacuity Among Children With Poor Literacy: Prevalence and Associated Factors. *Optometry and Vision Science*, *90*(1), 75–83. https://doi.org/10.1097/OPX.0b013e3182780dd0

Ponsonby, A.-L., Williamson, E., Smith, K., Bridge, D., Carmichael, A., Jacobs, A., Burrill, J., Ollington, N., Keeffe, J., & Dwyer, T. (2009). Children with Low Literacy and Poor Stereoacuity: An Evaluation of Complex Interventions in a Community-Based Randomized Trial. *Ophthalmic Epidemiology*, *16*(5), 311–321. https://doi.org/10.1080/09286580903144712

Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future Healthcare Journal*, *6*(3), 181–185.

Preece, D., Williams, S. B., Lam, R., & Weller, R. (2013). "Let's Get Physical": Advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education*, *6*(4), 216–224. https://doi.org/10.1002/ase.1345

Rochford, K. (1985). Spatial learning disabilities and underachievement among university anatomy students. *Medical Education*, *19*(1), 13–26. https://doi.org/10.1111/j.1365-2923.1985.tb01134.x

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience*, *14*, 96. https://doi.org/10.3389/fnhum.2020.00096

Schmitz, C. (2003). *LimeSurvey* (Version 5.6.53+240131) [Computer software].

Sepich, N. C., Jasper, A., Fieffer, S., Gilbert, S. B., Dorneich, M. C., & Kelly, J. W. (2022). The impact of task workload on cybersickness. *Frontiers in Virtual Reality*, *3*, 943409. https://doi.org/10.3389/frvir.2022.943409

Servotte, J.-C., Goosse, M., Campbell, S. H., Dardenne, N., Pilote, B., Simoneau, I. L., Guillaume, M., Bragard, I., & Ghuysen, A. (2020). Virtual Reality Experience: Immersion, Sense of Presence, and Cybersickness. *Clinical Simulation in Nursing*, *38*, 35–43. https://doi.org/10.1016/j.ecns.2019.09.006

Skulmowski, A., & Rey, G. D. (2017). Measuring Cognitive Load in Embodied Learning Settings. *Frontiers in Psychology*, *8*, 1191. https://doi.org/10.3389/fpsyg.2017.01191

Skulmowski, A., & Rey, G. D. (2018). Embodied learning: Introducing a taxonomy based on bodily engagement and task integration. *Cognitive Research: Principles and Implications*, *3*(1), 6. https://doi.org/10.1186/s41235-018-0092-9

Sona Systems, Ltd. (n.d.). *Sona Systems: Cloud-based Participant Management Software* [Computer software]. Sona Systems, Ltd. https://www.sona-systems.com/

Stanney, K., Fidopiastis, C., & Foster, L. (2020). Virtual Reality Is Sexist: But It Does Not Have to Be. *Frontiers in Robotics and AI*, *7*, 4. https://doi.org/10.3389/frobt.2020.00004

Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational Translation With Concrete Models in Organic Chemistry. *Cognition and Instruction*, *30*(4), 404–434. https://doi.org/10.1080/07370008.2012.719956

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Vandenberg, S. G., & Kuse, A. R. (1978). Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Voyer, D. (1997). Scoring Procedure, Performance Factors, and Magnitude of Sex Differences in Spatial Performance. *The American Journal of Psychology*, *110*(2), 259. https://doi.org/10.2307/1423717

Wainman, B., Aggarwal, A., Birk, S. K., Gill, J. S., Hass, K. S., & Fenesi, B. (2021). Virtual Dissection: An Interactive Anatomy Learning Tool. *Anatomical Sciences Education*, *14*(6), 788–798. https://doi.org/10.1002/ase.2035

Wainman, B., Pukas, G., Wolak, L., Mohanraj, S., Lamb, J., & Norman, G. R. (2020). The Critical Role of Stereopsis in Virtual and Mixed Reality Learning Environments. *Anatomical Sciences Education*, *13*(3), 401–412. https://doi.org/10.1002/ase.1928

Wainman, B., Wolak, L., Pukas, G., Zheng, E., & Norman, G. R. (2018). The superiority of three-dimensional physical models to two-dimensional computer presentations in anatomy learning. *Medical Education*, *52*(11), 1138–1146. https://doi.org/10.1111/medu.13683

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in Psychology*, *10*, 158. https://doi.org/10.3389/fpsyg.2019.00158

Xi, J., Wang, G.-T., Zhao, J., & Huang, C.-B. (2021). General and Specific Effects of Stereo Learning. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnhum.2021.535512

Yammine, K. (2014). The Current Status of Anatomy Knowledge: Where Are We Now? Where Do We Need to Go and How Do We Get There? *Teaching and Learning in Medicine*, *26*(2), 184–188. https://doi.org/10.1080/10401334.2014.883985

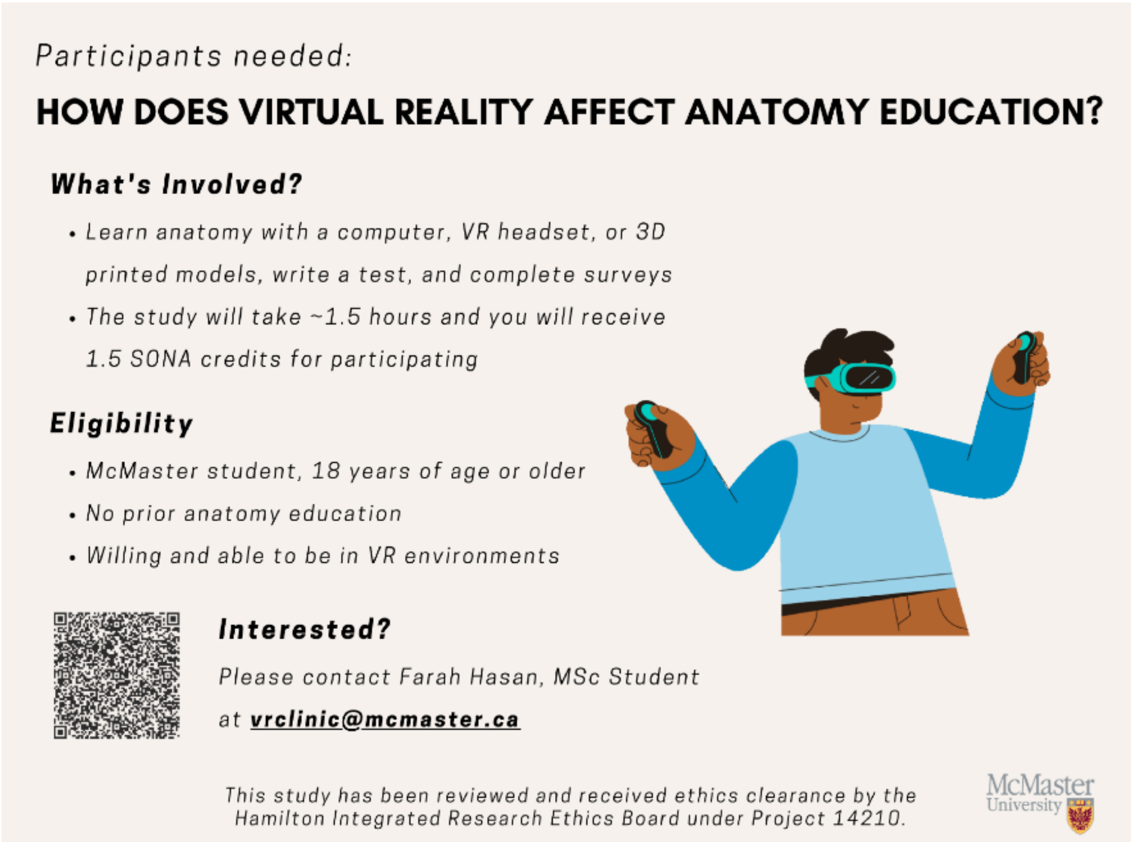
Young, J. Q., Van Merrienboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86. *Medical Teacher*, *36*(5), 371–384. https://doi.org/10.3109/0142159X.2014.889290

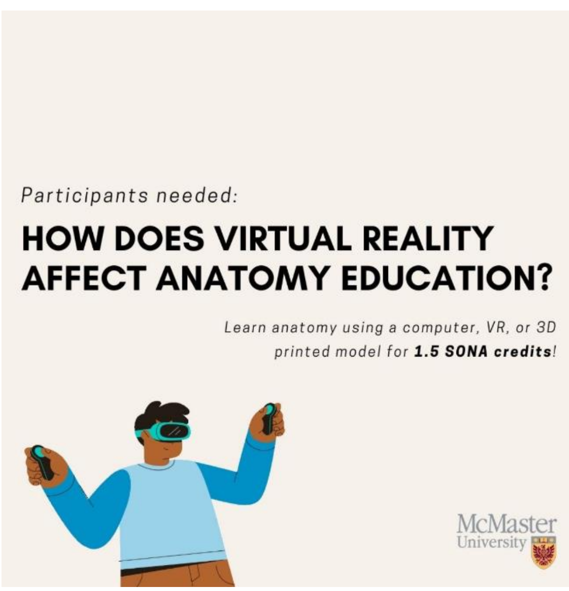
Appendices

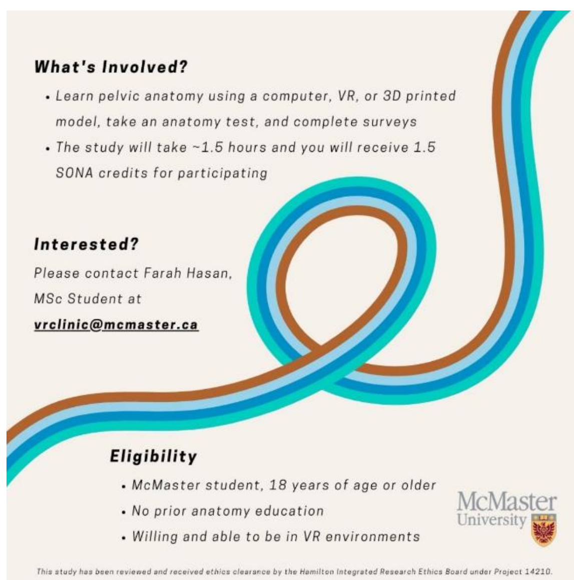
Appendix A: Online Recruitment Portal Advertisement

Credits	You will receive 1.5 credits for participating in the study.
Duration	1 session lasting approximately 1.5 hours
Abstract	The purpose of this study is to explore the effect virtual reality has on acquiring anatomical knowledge in the novice learner. You will receive 1.5 credits for participating in the study.
Description	Participants will learn anatomical structures using either a virtual reality headset, 3D printed models, or a computer. After learning, participants will be tested with a short, written test and will also complete a series of questionnaires. The study will take place in the Health Sciences Centre (HSC), room 1R1. Participation in the study will require approximately 1.5 hours of time. At the end of the study, participants will be compensated with 1.5 credits. After signing up, you'll receive a brief questionnaire via email. Please complete this ASAP – your responses will determine eligibility to participate. During the study, you may be asked to wear a Virtual Reality headset that has 2 straps that go over your head. If you have long hair, this is easiest done with hair worn down or in a low bun. If you anticipate having problems with the headset, please email the research team (vrclinic@mcmaster.ca) so we can work on accommodations that will allow you to participate in the study.
Inclusion/ exclusion criteria	Must not have prior classroom or laboratory experience in anatomy (ex. Mammalian Physiology, Anatomy & Physiology, Neuroanatomy, Embryology, Human Development, Histology). No known nausea/dizziness/anxiety when using VR equipment.
Researchers	Co-investigators: Dr. Ranil Sonnadara, PhD, Department of Surgery, McMaster University Dr. Bruce Wainman, PhD, Department of Pathology and Molecular Medicine, McMaster University Student investigator: Farah Hasan, MSc Student in the Health Science Education Program, McMaster University Undergraduate student investigators: Zhiyu (Zoe) Wu, Athena Li, Aida Esmaelbeigi Please contact Farah at (vrclinic@mcmaster.ca) if you have any questions.

Appendix B: Recruitment Poster and Social Media Advertisement







Appendix C: Participant Screening Questionnaire

Thank you for expressing interest in participating in the study entitled: “Efficacy of virtual reality in the acquisition of anatomical knowledge.” To confirm your candidacy for the study, please respond to all of the items in the following three sections of the participant recruitment form.

Background Information

Please provide your full name.	______________________________
Please provide your email so we can contact you regarding participation in this study.	_______________________________
Please indicate the age group which describes you.	○ 17 or younger ○ 18-20 ○ 21-23 ○ 24-26 ○ 27-29 ○ 30 or older
Please indicate your sex.	○ Female ○ Male ○ Intersex ○ Prefer not to disclose
Please indicate the degree you are currently working towards earning.	○ Bachelors ○ Masters ○ Doctorate ○ Other
Please identify why you selected ‘other’.	_____________________________
Please indicate the faculty or program you are currently enrolled in.	○ Faculty of Humanities ○ Faculty of Social Sciences ○ Faculty of Engineering ○ Faculty of Science ○ Faculty of Health Science ○ DeGroote School of Business ○ Arts and Science Program ○ Other
Please indicate your ‘other’ faculty or program.	____________________
Please indicate your current year of study.	○ 1 ○ 2 ○ 3 ○ 4 ○ 5

	○ 6 or more
Have you taken university anatomy and physiology courses?	○ Yes ○ No
If you responded "yes" for the previous question, please explain.	______________________________

Experience with Technology

Have you used a virtual reality device (e.g. HTC Vive, Samsung Odyssey, Oculus Rift) in any previous learning experiences?	○ Yes ○ No
Have you used a virtual reality device for experiences other than learning (e.g. gaming, movies, therapeutic applications, training)?	○ Yes ○ No

Please indicate your level of agreement with the statement below.

I am comfortable using technology (e.g. computer applications, websites, mobile devices, educational games, etc.) in learning.	○ Strongly disagree ○ Disagree ○ Agree ○ Strongly agree
Please include any other information you think may be helpful for the investigators to know.	____

Thank you for answering the items in this participant recruitment form. After your responses have been reviewed, you will be contacted by email regarding your candidacy for the study.

Appendix D: Participant Recruitment Email

Dear ____________________________,

Thank you for your interest in participating in the study entitled "Efficacy of virtual reality in the acquisition of anatomical knowledge". The investigators have reviewed the answers you submitted on the participant recruitment form and determined that you are a candidate for the study.

We are confirming that you have signed up for a time slot on (date) at (time). The approximate time needed to participate in the study is 1.5 hours. The study will take place in the Health Sciences Centre, in the Anatomy Lab – HSC 1R1. The lab can be difficult to find, so we have attached a map of the first floor of the McMaster Children's Hospital and the anatomy lab with common landmarks. We recommend you give yourself extra time to find the location. If you get lost or are not comfortable finding the lab, please email us and someone from our team will meet you at the William's in HSC (1st floor) to accompany you to the lab.

In preparation for the study, please review the following instructions carefully:

- During the study, you may be asked to wear a Virtual Reality headset that has 2 straps that go over your head (see attached photo). If you have long hair or you wear a religious head cover (e.g. hijab), this is easiest done with hair worn down or in a low bun. If you anticipate having problems with putting on or wearing the headset, please email us at your earliest convenience so that we can work on accommodations that will allow you to participate in the study.
- On the day of the study, ensure that you wear closed toed shoes and bring your student card to the lab.
- Lastly, if you wear glasses or contacts, please wear them for participation in the study.

The Letter of Information is attached to this email. Please review this document prior to your arrival. When you arrive for the study, a coordinator will review the study and content of the consent form with you one-on-one to answer questions you may have before beginning.

If you have any questions, please feel free to contact us at this email (vrclinic@mcmaster.ca). Thank you for choosing to participate in our study. Your contribution to the research is valued and appreciated.

Appendix E: Pre-test Assessment

1. Which of the following muscles is located within the superficial perineal pouch?
   a. Bulbospongiosus
   b. Coccygeus
   c. External urethral sphincter
   d. Internal anal sphincter
   e. Deep transverse perineus

2. Relative to the sacrum, is the location of the pelvic bone lateral, medial, or posterior?

   ____

3. What structure is deep to the deep transverse perineus?

   ____

4. A 29-year-old woman experiences tearing in the perineum after a long and difficult delivery. Which of the following muscles may result in urinary incontinence if affected by the tearing?
   a. Deep transverse perineus
   b. Iliococcygeus
   c. Ischiocavernosus
   d. Obturator internus
   e. Pubococcygeus

5. Identify the structure with the number 4 on it.

   ____

6. Which of the following connective tissue structures provides support to the pelvic floor?
   a. Obturator membrane
   b. Perineal body
   c. Perineal membrane
   d. Sacrospinous ligament
   e. Sacrotuberous ligament

7. Which of the following orifices/openings is located most posterior in the female perineum?
   a. Anus
   b. Urethra
   c. Vagina

8. Identify the structure with the number 20 on it.

   ____

9. Using ONE term from the anatomy orientation sheet, identify the location of the puborectalis relative to the iliococcygeus.

____

10. Identify the structure with the number 7 on it.

____

11. A 38-year-old female is provided with a series of exercises to strengthen the levator ani muscles of her pelvic floor. Which of the following muscles is part of the levator ani?
    a. Coccygeus
    b. External anal sphincter
    c. Iliococcygeus
    d. Obturator internus
    e. Superficial transverse perineus

12. Identify the structure with the number 6 on it.

____

13. A 68-year-old woman with a history of 5 vaginal births visits her doctor for flatal incontinence. She states that it has gotten worse over time. Weakening of which of the following muscles may be responsible for this patient's flatal incontinence?
    a. Bulbospongiosus
    b. Deep transverse perineus
    c. Sphincter urethrovaginalis
    d. Obturator internus
    e. Puborectalis

14. Identify the structure with the number 22 on it.

____

15. Using ONE term from the anatomy orientation sheet, identify the location of the opening of the vagina relative to the urethra.

____

Appendix F: Post-test Assessment

1. Which of the following muscles is located within the deep perineal pouch?
   a. Bulbospongiosus
   b. Coccygeus
   c. Deep transverse perineus
   d. Internal anal sphincter
   e. Purborectalis

2. Using ONE term from the anatomy orientation sheet, identify the location of the sacrum relative to the coccyx.

   ____

3. What structure is superficial to the deep transverse perineus?

   ____

4. A 29-year-old woman experiences tearing in the perineum after a long and difficult delivery. Which of the following muscles may result in urinary incontinence if affected by the tearing?
   a. Bulbospongiosus
   b. Coccygeus
   c. Compressor urethrae
   d. Puborectalis
   e. Piriformis

5. Identify the structure with the number 2 on it.

   ____

6. Which of the following connective tissue structures provides support to the pelvic floor and serves as an attachment point for several muscles in the perineum.
   a. Obturator membrane
   b. Perineal body
   c. Perineal membrane
   d. Sacrospinous ligament
   e. Sacrotuberous ligament

7. Which of the following orifices/openings is located most anterior in the female perineum?
   a. Anus
   b. Urethra
   c. Vagina

8. Identify the structure with the number 21 on it.

 ____

9. Using ONE term from the anatomy orientation sheet, identify the location of the pubococcygeus relative to the iliococcygeus.

 ____

10. Identify the structure with the number 16 on it.

 ____

11. A 38-year-old female is provided with a series of exercises to strengthen the levator ani muscles of her pelvic floor. Which of the following muscles is part of the levator ani?
    a. Coccygeus
    b. External anal sphincter
    c. Obturator internus
    d. Puborectalis
    e. Superficial transverse perineus

12. Identify the structure with the number 10 on it.

 ____

13. A 68-year-old woman with a history of 5 vaginal births visits her doctor for flatal incontinence. She states that it has gotten worse over time. Weakening of which of the following muscles may be responsible for this patient's flatal incontinence?
    a. Bulbospongiosus
    b. Deep transverse perineus
    c. Sphincter urethrovaginalis
    d. Obturator internus
    e. Puboccygeus

14. Identify the structure with the number 19 on it.

 ____

15. Using ONE term from the anatomy orientation sheet, identify the location of the greater vestibular gland relative to the opening of the vagina.

 ____

Appendix G: Anatomy Structure List

Anus
Bulbospongiosus
Coccygeus
Coccyx
Compressor urethrae
Deep transverse perineus
External anal sphincter
External urethral sphincter
Greater vestibular gland
Hip bones
Iliococcygeus
Internal anal sphincter
Ischiocavernosus
Obturator internus
Obturator membrane
Pelvic diaphragm
Perineal body
Perineal membrane
Piriformis
Pubococcygeus
Puborectalis
Sacrospinous ligament
Sacrotuberous ligament
Sacrum
Sphincter urethrovaginalis
Superficial transverse perineus
Urethra
Vagina

Appendix H: Anatomy Orientation Sheet

When indicated, use one of the anatomical orientation terms below when describing the location of structures on the test.

Anterior – at or near the front of the body (i.e. nose is anterior to ears)

Posterior – at or near the back of the body (i.e. ears are posterior to nose)

Medial – towards the midline of the body (i.e. nose is medial to eyes)

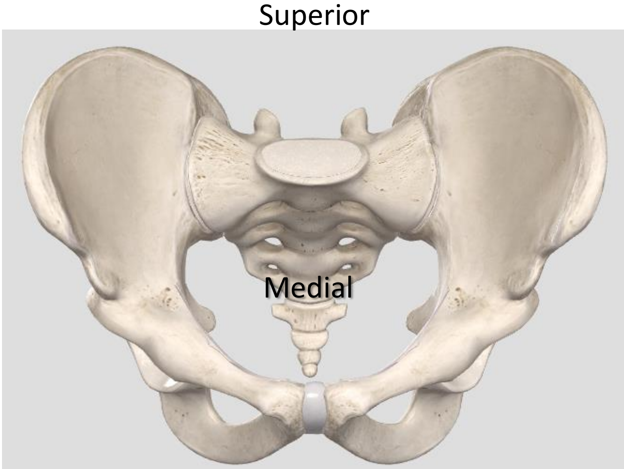
Lateral – away from the midline of the body (i.e. eyes are lateral to nose)

Superior – towards the upper part of the body (i.e. eyes are superior to nose)

Inferior – towards the lower part of the body (i.e. nose is inferior to eyes)

Superficial – towards the surface of the body (i.e. skin is superficial to bones)

Deep – away from the surface of the body (i.e. bones are deep to skin)



Appendix I: Simulator Sickness Questionnaire (SSQ)

While using the modality I felt (choose one for each row):

General discomfort	None	Slight	Moderate	Severe
Fatigue	None	Slight	Moderate	Severe
Headache	None	Slight	Moderate	Severe
Eye strain	None	Slight	Moderate	Severe
Difficulty focusing	None	Slight	Moderate	Severe
Increased salivation	None	Slight	Moderate	Severe
Sweating	None	Slight	Moderate	Severe
Nausea	None	Slight	Moderate	Severe
Difficulty concentrating	None	Slight	Moderate	Severe
Fullness of the Head	None	Slight	Moderate	Severe
Blurred vision	None	Slight	Moderate	Severe
Dizzy (eyes open)	None	Slight	Moderate	Severe
Dizzy (eyes closed)	None	Slight	Moderate	Severe
Vertigo	None	Slight	Moderate	Severe
Stomach awareness	None	Slight	Moderate	Severe
Burping	None	Slight	Moderate	Severe

Appendix J: NASA Task Load Index (NASA-TLX)

INTRODUCTION:

This survey has two sections which measure your perceived workload based on six factors (see descriptions on the right).

In the first part, you will use a scale to rate how each of the six factors contributed to your experience while you were learning with the pelvic anatomy module.

In the second part, you will select the factor which contributed most to your experience while you were learning with the pelvic anatomy module.

PART 1:

Please rate your experience during the learning task by putting an "X" on the six rating scales that the researcher provided you with.

When you are finished, please let the researcher know and they will provide you with part 2 of the survey.

Version 4 | 11/28/22

RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

INTRODUCTION:

This survey has two sections which measure your perceived workload based on six factors (see descriptions on the right).

In the first part, you will use a scale to rate how each of the six factors contributed to your experience while you were learning with the pelvic anatomy module.

In the second part, you will select the factor which contributed most to your experience while you were learning with the pelvic anatomy module.

PART 2:

You will be presented with a deck of cards and each card will have two factors to choose from. Using the dry erase marker, circle the factor that was the more important contributor to workload while you were learning with the pelvic anatomy module.

Please go through the cards one at a time.

When you are finished, please let the researcher know and they will collect the cards and test materials.

Version 4 | 11/28/22

RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Study ID: ____________

RATING SHEET

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

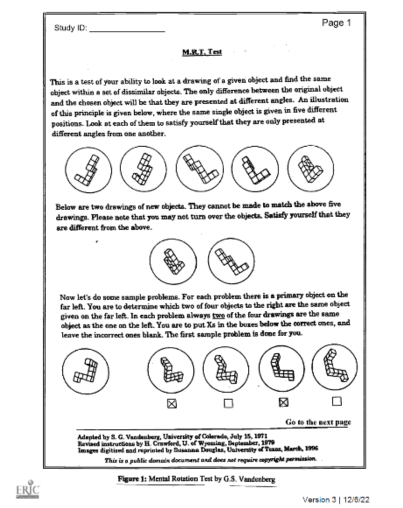
Low High

FRUSTRATION

Low High

Effort or Performance	Temporal Demand or Frustration	Frustration or Effort	Performance or Mental Demand
Temporal Demand or Effort	Physical Demand or Frustration	Performance or Temporal Demand	Mental Demand or Effort
Performance or Frustration	Physical Demand or Temporal Demand	Mental Demand or Temporal Demand	Effort or Physical Demand
Physical Demand or Performance	Temporal Demand or Mental Demand	Frustration or Mental Demand	

Appendix K: Mental Rotations Test (MRT)



Study ID: ______________ Page 1

M.R.T. Test

This is a test of your ability to look at a drawing of a given object and find the same object within a set of dissimilar objects. The only difference between the original object and the chosen object will be that they are presented at different angles. An illustration of this principle is given below, where the same single object is given in five different positions. Look at each of them to satisfy yourself that they are only presented at different angles from one another.

Below are two drawings of new objects. They cannot be made to match the above five drawings. Please note that you may not turn over the objects. Satisfy yourself that they are different from the above.

Now let's do some sample problems. For each problem there is a primary object on the far left. You are to determine which two of four objects to the right are the same object given on the far left. In each problem always two of the four drawings are the same object as the one on the left. You are to put Xs in the boxes below the correct ones, and leave the incorrect ones blank. The first sample problem is done for you.

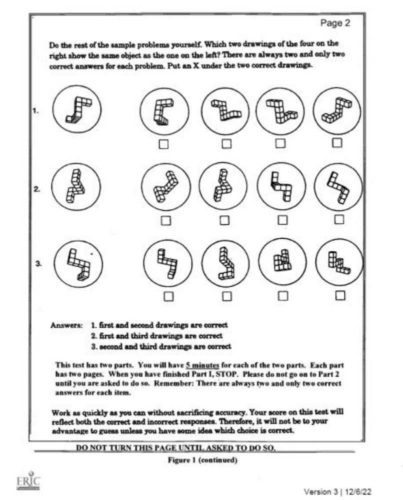
Go to the next page

Adapted by S. G. Vandenberg, University of Colorado, July 15, 1971
Revised instructions by H. Crawford, U. of Wyoming, September, 1979
Images digitized and reprinted by Susanna Douglas, University of Texas, March, 1996

This is a public domain document and does not require copyright permission.

Figure 1: Mental Rotation Test by G.S. Vandenberg

Version 3 | 12/6/22



Page 2

Do the rest of the sample problems yourself. Which two drawings of the four on the right show the same object as the one on the left? There are always two and only two correct answers for each problem. Put an X under the two correct drawings.

1.

2.

3.

Answers: 1. first and second drawings are correct
2. first and third drawings are correct
3. second and third drawings are correct

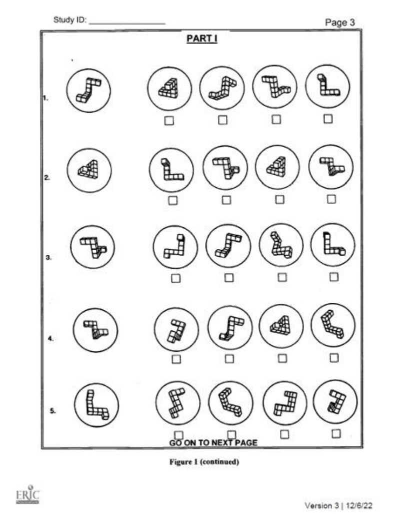
This test has two parts. You will have 5 minutes for each of the two parts. Each part has two pages. When you have finished Part I, STOP. Please do not go on to Part 2 until you are asked to do so. Remember: There are always two and only two correct answers for each item.

Work as quickly as you can without sacrificing accuracy. Your score on this test will reflect both the correct and incorrect responses. Therefore, it will not be to your advantage to guess unless you have some idea which choice is correct.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Figure 1 (continued)

Version 3 | 12/6/22



Study ID: ______________ Page 3

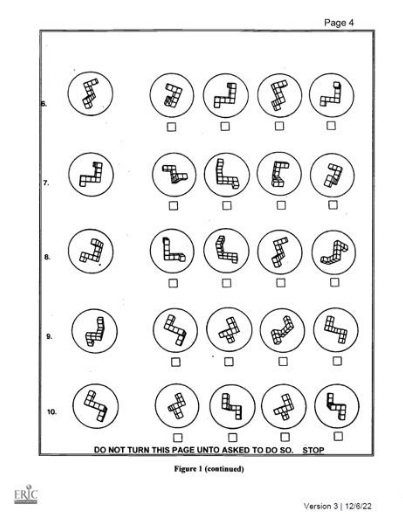
PART I

1.
2.
3.
4.
5.

GO ON TO NEXT PAGE

Figure 1 (continued)

Version 3 | 12/6/22



Page 4

6.
7.
8.
9.
10.

DO NOT TURN THIS PAGE UNTO ASKED TO DO SO. STOP

Figure 1 (continued)

Version 3 | 12/6/22

Study ID: ______________ Page 5

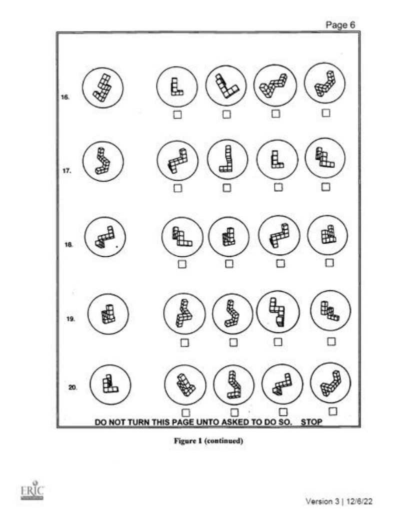
PART II

11.
12.
13.
14.
15.

GO TO NEXT PAGE

Figure 1 (continued)

Version 3 | 12/6/22



Page 6

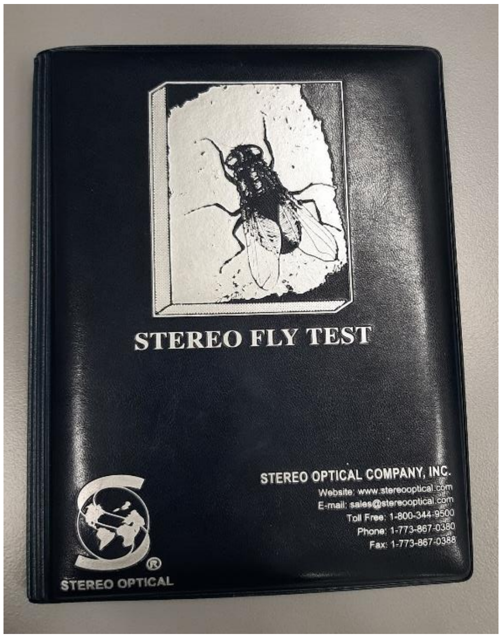
16.
17.
18.
19.
20.

DO NOT TURN THIS PAGE UNTO ASKED TO DO SO. STOP

Figure 1 (continued)

Version 3 | 12/6/22

Appendix L: Stereo Fly Test



Appendix M: User Engagement Scale (UES) and Pelvic Anatomy Learning Module Feedback Survey

The following statements ask you to reflect on your experience of engaging with the 10-minute pelvic anatomy learning module. For each statement, please use the following scale to indicate what is most true for you.

Question	Strongly Disagree 1	Disagree 2	Neither agree nor disagree 3	Agree 4	Strongly agree 5
I lost myself in this experience.					
I was so involved in this experience that I lost track of time.					
I blocked out things around me when I was using the pelvic anatomy learning module.					
When I was using the pelvic anatomy learning module, I lost track of the world around me.					
The time I spent using the pelvic anatomy learning module just slipped away.					
I was absorbed in this experience.					
During this experience I let myself go.					
I felt frustrated while using this pelvic anatomy learning module.					
I found this pelvic anatomy learning module confusing to use.					
I felt annoyed while using the pelvic anatomy learning module.					
I felt discouraged while using this pelvic anatomy learning module.					
Using this pelvic anatomy learning module was taxing.					
This experience was demanding.					
I felt in control while using this pelvic anatomy learning module.					

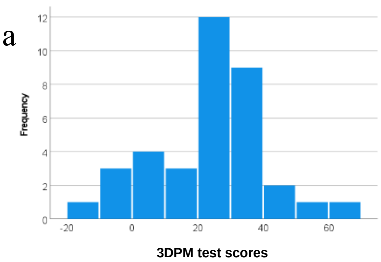
I could not do some of the things I needed to do while using the pelvic anatomy learning module.					
This pelvic anatomy learning module was attractive.					
This pelvic anatomy learning module was aesthetically appealing.					
I liked the graphics and images of the pelvic anatomy learning module.					
The pelvic anatomy learning module appealed to be visual senses.					
The screen layout of the pelvic anatomy learning module was visually pleasing.					
Using the pelvic anatomy learning module was worthwhile.					
I consider my experience a success.					
This experience did not work out the way I had planned.					
My experience was rewarding.					
I would recommend the pelvic anatomy learning module to my family and friends.					
I continued to use the pelvic anatomy learning module out of curiosity.					
The content of the pelvic anatomy learning module incited my curiosity.					
I was really drawn into this experience.					
I felt involved in this experience.					
This experience was fun.					

What facilitated or limited your ability to understand the educational content presented in the learning module?

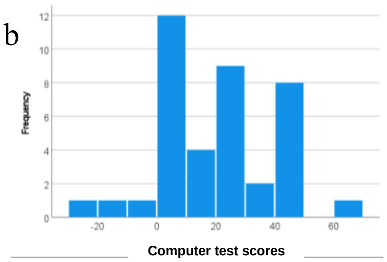
What recommendations do you have to improve the user experience for the learning module?

Additional comments for the investigator:

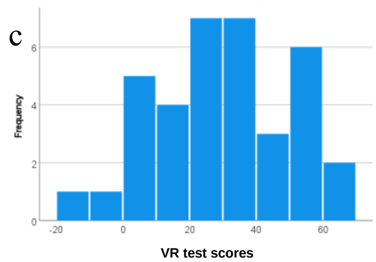
Appendix N: Statistical Pre-Analysis

a


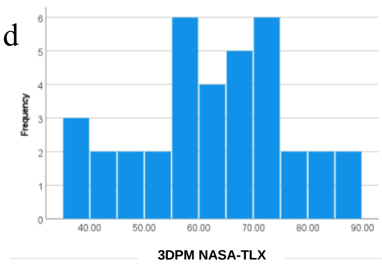
Mean (SD) = 22.22 (17.16), n = 36

b


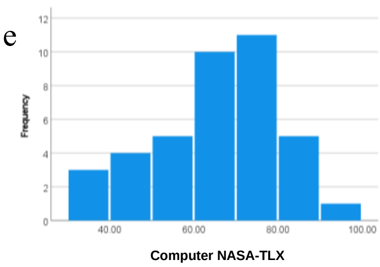
Mean (SD) = 19.15 (19.24), n = 39

c


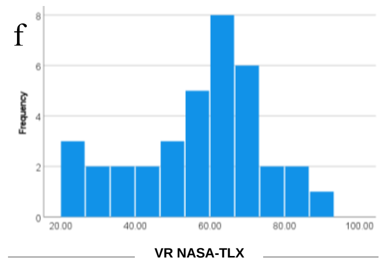
Mean (SD) = 28.33 (20.12), n = 36

d


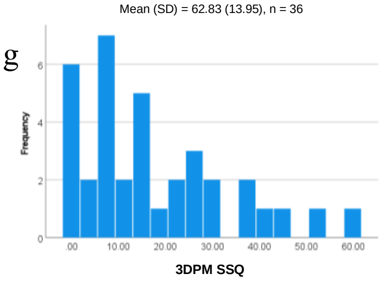
Mean (SD) = 62.83 (13.95), n = 36

e


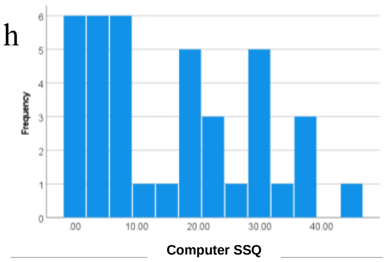
Mean (SD) = 64.97 (15.05), n = 39

f


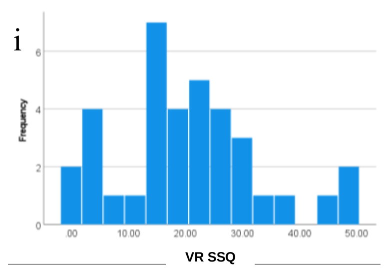
Mean (SD) = 57.43 (17.00), n = 36

g


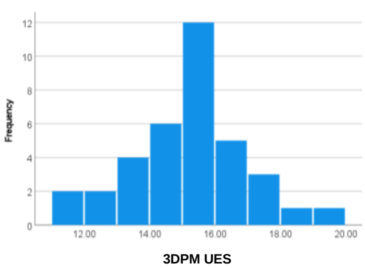
Mean (SD) = 17.56 (15.79), n = 36

h


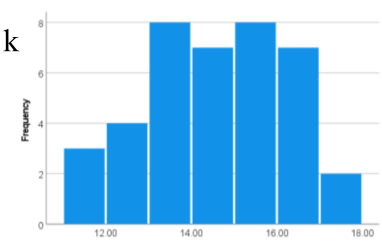
Mean (SD) = 15.92 (13.23), n = 39

i


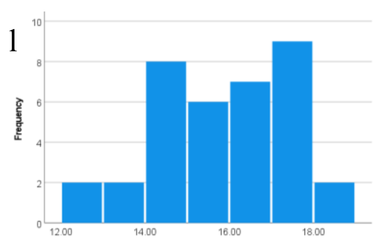
Mean (SD) = 20.36 (12.53), n = 36

j


Mean (SD) = 15.22 (1.80), n = 36

k


Mean (SD) = 14.57 (1.59), n = 39

l


Mean (SD) = 15.88 (1.73), n = 36

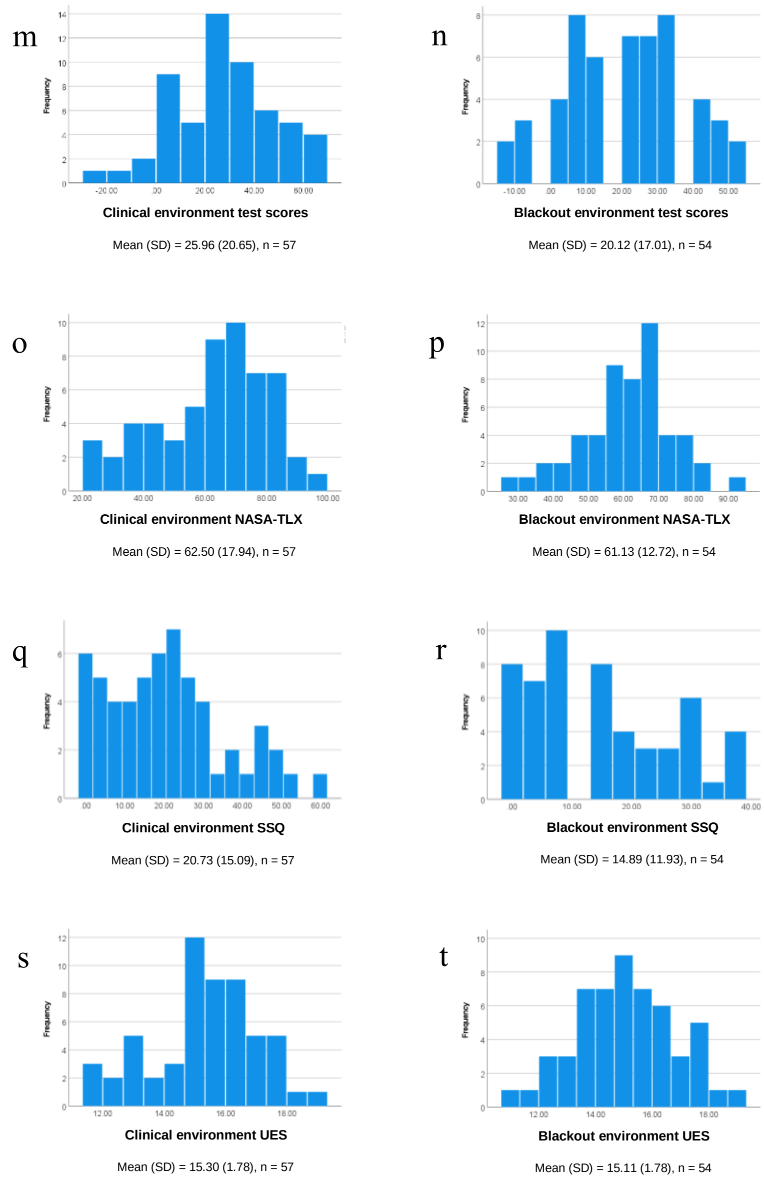


Figure 11. Histograms of data distributions by modality and environment for test scores (a-c, m-n), NASA-TLX (d-f, o-p), SSQ (g-I, q-r), and UES (j-l, s-t) (total n = 111).

Table 3. Levene's test statistics and significance for homoscedasticity (n = 111). P > 0.05 indicates equal variances.

Comparison	F statistic	Significance
Delta ANCOVA	1.466	0.207
NASA-TLX ANCOVA	1.673	0.148
SSQ ANCOVA	0.348	0.882
UES ANCOVA	1.323	0.260

Table 4. Shapiro-Wilk's test of normality statistics for test scores, NASA-TLX, SSQ, and UES (total n = 111). *p < 0.05 indicates non-normal distribution.

Variable	Learning Condition		Statistic	df	Significance	n
Test Scores	Modality	3D Printed Model	0.961	36	0.227	36
		Computer	0.975	39	0.523	39
		Virtual Reality	0.970	36	0.439	36
	Environment	Context	0.981	57	0.498	57
		No Context	0.974	54	0.275	54
NASA-TLX	Modality	3D Printed Model	0.968	36	0.382	36
		Computer	0.970	39	0.367	39
		Virtual Reality	0.954	36	0.139	36
	Environment	Context	0.955	57	0.032*	57
		No Context	0.983	54	0.638	54
SSQ	Modality	3D Printed Model	0.900	36	0.003*	36
		Computer	0.910	39	0.004*	39
		Virtual Reality	0.952	36	0.118	36
	Environment	Context	0.948	57	0.015*	57
		No Context	0.910	54	0.001*	54
UES	Modality	3D Printed Model	0.984	36	0.868	36
		Computer	0.973	39	0.456	39
		Virtual Reality	0.960	36	0.216	36
	Environment	Context	0.977	57	0.362	57
		No Context	0.988	54	0.876	54

Appendix O: Results

Table 9. F statistics, significance, and observed power for tests of between-subjects effects for test scores (n = 111). Observed power computed using $\alpha = .05$. $^{*}p < 0.05$.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	9052.305[a]	9	1005.812	3.265	0.002	0.975
Intercept	0.932	1	0.932	0.003	0.956	0.050
MRT	3714.937	1	3714.937	12.061	0.001*	0.931
NASA-TLX	240.025	1	240.025	0.779	0.379	0.141
SSQ	718.616	1	718.616	2.333	0.130	0.328
UES	84.140	1	84.140	0.273	0.602	0.081
Modality	1021.301	2	510.650	1.658	0.196	0.342
Environment	1407.641	1	1407.641	4.570	0.035*	0.563
Modality * Environment	2515.335	2	1257.668	4.083	0.020*	0.713
Error	31109.457	101	308.014			
Total	99511.111	111				
Corrected Total	40161.762	110				

Table 11. F statistics, significance, and observed power for tests of between-subjects effects for NASA-TLX (n = 111). Observed power computed using $\alpha = .05$. $^{*}p < 0.05$.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	4561.567	8	570.196	2.633	0.012	0.908
Intercept	5739.997	1	5739.997	26.508	0.000	0.999
MRT	716.705	1	716.705	3.310	0.072	0.437
SSQ	715.042	1	715.042	3.302	0.072	0.437
UES	199.445	1	199.445	0.921	0.339	0.158
Modality	633.268	2	316.634	1.462	0.237	0.306
Environment	0.045	1	0.045	0.000	0.989	0.050
Modality * Environment	1684.189	2	842.095	3.889	0.024*	0.691
Error	22087.071	102	216.540			
Total	451021.111	111				
Corrected Total	26648.639	110				

Table 13. F statistics, significance, and observed power for tests of between-subjects effects for SSQ (n = 111). Observed power computed using $\alpha = .05$. $^{*}p < 0.05$.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	4414.254	8	551.782	3.344	0.002	0.967
Intercept	1296.959	1	1296.959	7.860	0.006	0.793
MRT	115.372	1	115.372	0.699	0.405	0.132
UES	1422.479	1	1422.479	8.621	0.004*	0.829
NASA-TLX	544.859	1	544.859	3.302	0.072	0.437
Modality	1049.762	2	524.881	3.181	0.046*	0.597
Environment	1046.379	1	1046.379	6.342	0.013*	0.704
Modality * Environment	507.365	2	253.682	1.537	0.220	0.320
Error	16830.264	102	165.003			
Total	56775.668	111				
Corrected Total	21244.518	110				

Table 15. F statistics, significance, and observed power for tests of between-subjects effects for UES (n = 111). Observed power computed using $\alpha = .05$. $^{*}p < 0.05$.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	73.048	8	9.131	3.419	0.002	0.971
Intercept	1181.810	1	1181.810	442.485	0.000	1.000
MRT	3.809	1	3.809	1.426	0.235	0.219
NASA-TLX	2.460	1	2.460	0.921	0.339	0.158
SSQ	23.025	1	23.025	8.621	0.004*	0.829
Modality	35.627	2	17.813	6.670	0.002*	0.907
Environment	3.762	1	3.762	1.408	0.238	0.217
Modality * Environment	2.098	2	1.049	0.393	0.676	0.112
Error	272.426	102	2.671			
Total	26012.416	111				
Corrected Total	345.474	110				