12-2022

Understanding neuroanatomy in a virtual 3D environment: creation and use of a new survey tool to evaluate the effectiveness of 3D software in neuroanatomy education for understanding superficial and deep brain structures.

Akash Khare
University of Louisville

UNDERSTANDING NEUROANATOMY IN A VIRTUAL 3D ENVIRONMENT:

CREATION AND USE OF A NEW SURVEY TOOL TO EVALUATE THE EFFECTIVENESS OF 3D SOFTWARE IN NEUROANATOMY EDUCATION FOR UNDERSTANDING SUPERFICIAL AND DEEP BRAIN STRUCTURES

By

Akash Khare

M.B.B.S., King George's Medical University, 2010

A Thesis Submitted to the faculty of the School of Medicine of the University of Louisville in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in Anatomical Sciences and Neurobiology

Department of Anatomical Sciences and Neurobiology

University of Louisville

Louisville, Kentucky

December 2022

UNDERSTANDING NEUROANATOMY IN A VIRTUAL 3D ENVIRONMENT:

CREATION AND USE OF A NEW SURVEY TOOL TO EVALUATE THE EFFECTIVENESS OF 3D SOFTWARE IN NEUROANATOMY EDUCATION FOR UNDERSTANDING SUPERFICIAL AND DEEP BRAIN STRUCTURES

By

Akash Khare

A Thesis Approved on

November 30, 2022

by the following Thesis Committee:

Dr. Brian Davis

Dr. Chad Samuelsen

Dr. John Pani

ACKNOWLEDGMENTS

First, I would like to thank my mentor and thesis advisor, Dr. Brian Davis for believing in me, and my ideas for this project and immensely supporting me throughout the process. He has always been there to give professional support and personal guidance during difficult times that I am sure every graduate student goes through worldwide. Words cannot express how grateful I am for your support and encouragement.

This thesis would only have existed with the contributions and support of faculty. I was lucky to have Dr. John Pani, as my thesis co-advisor as he is a legend and widely cited author in the use of 3D technologies for neuroanatomy education and the role of spatial ability. Secondly, could only have done it with the support of Dr. Erin Davis, who shined a light on various approaches we could take and helped all along the way. I want to thank my other co-advisor Dr. Chad Samuelsen, who has incredible analytic skills and is a fantastic person. I want to thank Dr. Robert F. Lundy for allowing and supporting this project in his course.

The department of Anatomical sciences and Neurobiology (ASNB) at UofL would not be as excellent as it is, if not for Dr. William Guido, who is a great neuroscientist, outstanding teacher, and firm administrator, and I thank him for allowing me to study and work in his department. I want to thank Dr. J. Patrick Moore for being there during and since I joined this school and his great energy. I would also like to thank Dr. Robin Krimm for her dedicated support to graduate students. Also, Charles H. Hubscher had supported us during admissions and the first year, for which I greatly thank him. I would also like to thank Dr. Guillermo Rougier for being a wonderful person and a great thinker and for his support. I thank Dr. Aaron W. McGee, for being one of the most analytical and scientific thinkers I have met and for his teaching and support in the Anatomy seminars course.

It was not a random decision for me to come to this school. Gross anatomy at UofL is of very high standards, and all the students, including me, are lucky to have Dr. Jennifer Brueckner-Collins and Dr. Nicole R. Herring taking care of such a mammoth task in such a brilliant way through their fabulous courses. I want to thank the late Dr. Robert D. Acland for being the spiritual guide to this accomplishment and for his contributions to the field of gross anatomy. I would also like to thank other faculty members and graduate students in the department for being part of the incredible family we call ASNB at UofL.

I would also like to take this opportunity to thank my past teachers, my family and friends, without who I could not succeed. Thank you for giving me continuous encouragement throughout my academic career.

ABSTRACT

UNDERSTANDING NEUROANATOMY IN A VIRTUAL 3D ENVIRONMENT:

CREATION AND USE OF A NEW SURVEY TOOL TO EVALUATE THE EFFECTIVENESS OF 3D SOFTWARE IN NEUROANATOMY EDUCATION FOR UNDERSTANDING SUPERFICIAL AND DEEP BRAIN STRUCTURES

Akash Khare

November 30, 2022

Visualization is an important factor determining success in science and technological fields. Neuroanatomy education can be tricky as it contains many deep and complex structures that can be difficult to visualize and understand. Students often face difficulties regarding their position, structure, and spatial relationships, leading to a fear of neuroanatomy. Studying cross-sections is a critical approach to learning and testing knowledge in neuroanatomy and is used clinically in radiological examinations. Cadaver-based learning is essential, but its use has been decreasing over time, and the role of 3D technologies have been gradually increasing in medical education, especially after the COVID-19 pandemic, which caused the prolonged closure of schools and colleges across the world. Regular 3D software on 2D screens is the most used technology (as virtual reality is still in its nascent stage) and has been shown to enhance learning regardless of initial spatial abilities. Many other studies show mixed reviews and often lack an appropriate instrument to measure the effectiveness; hence a new survey was created.

The prospective study was conducted during school shutdowns as an immediate response to the COVID-19 pandemic in a quasi-experimental one-group pre-post interventional design in an online setting by creating, implementing, and evaluating the effectiveness of a virtual lab in neuroanatomy for all neuroscience students enrolled in the Fundamentals of Neuroscience course in our department at the University of Louisville (ASNB 502/602). The study population consisted of a total of 35 students, which included 25 undergraduate and ten graduate students. There were more female students overall (57%) and in the undergraduate population (64%), but graduate students consisted of a male majority (60%). Study modules were created using the 2D resources used in previous years and 3D web applications of Visible Body and AnatomyLearning.com software. A newly developed 13-item Reaction-Relevance-Result survey measured the effectiveness of these resources, along with Confidence in topics survey instruments and tests, and involved both quantitative and qualitative analysis. The surveys were based on pre-validated and widely used Kirkpatrick and ARCS models and other surveys. They were tested for validity by factor analysis, Cronbach's coefficient alpha, and other correlations. The use of surveys along with tests gives a more balanced picture and student side of the story.

Results of the study confirmed the advantages of using 3D software for neuroanatomy education in increasing student engagement, learning, confidence, and performance, with mostly large effect sizes for the pre-post effects on the RRR and Confidence in topics surveys and test results, while other significant group differences observed in this study mainly were with medium effect sizes. The study sheds some light on social need and justice regarding the utility of 3D intervention to bring equitable learning among all genders and academic levels without any effects of earlier performances. Female students, while facing little difficulties while studying the 2D resources (due to differences in innate spatial abilities) and may encounter some usage-related barriers for 3D software, found these 3D resources to be as helpful as the male students did and performed equally as well. While the outcome for female students might be equal to male students, 3D software provides benefits over 2D resources more to female students than male students, bringing in some social justice. This is important in neuroscience education which has shown increased participation by female students and may one day be dominated by the female gender due to changing demographics in neuroscience and other fields of science and technology. The study also uncovered some bias in student perception of the advantages of 3D software for students with any previous neuroanatomy experience, suggesting that novice students (but not necessarily young) are more comfortable using these 3D resources. However,

they performed the same as experienced students. Qualitative analysis showed that the 3D software brings in more accessibility, manipulation of views, and dissection features, leading to a better overview and spatial understanding of the brain, especially for deeper structures. However, the software needs to add more clinical material and evolve to provide more realistic renderings and smoother runtime.

The student insight regarding their knowledge of topics was better only after the 3D intervention; hence the use of surveys can be inadequate for less effective interventions without considering test results. 3D software increased understanding of superficial and deep structures but was more beneficial for deeper structures, thus bridging the difficulty gap between superficial and deep structures. When combined with the effect of gender, this influence may support our finding that the equality brought upon by 3D software may not be accurate as the overall perception and scores may seem equal but the gender differences seep deep down cognitively to complex and deeper levels where female students benefit more from the superficial structures, and male students benefit more from the deep structures, leading to male students being more successful in narrowing the difficulty gap between the superficial and deep structures. Further training on spatial abilities and prolonged use of 3D software may eventually lead to these differences fading away. The goal of the study was not to prove that 3D technologies should replace 2D resources or cadaveric dissection but to suggest that a multi-dimensional approach may be a better method for providing neuroscience education instead of a single (2D) or bi-dimensional approach (2D + cadavers) that is mainly used in schools.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE PAGE

INTRODUCTION

Anatomy education and various modes of teaching

Science, technology, engineering, mathematics, and medicine (STEMM) disciplines are heavily loaded with text material that needs visualization for a clear understanding; medical sciences, and especially its most critical foundational branch, the anatomical sciences, contains a lot of structures and their relationships that can only be comprehended with proper visual resources (Naaz, 2012; Gillbert, 2007). Anatomy education builds the foundation for medical and surgical sciences. However, studying anatomy requires high levels of spatial visualization as the body is a complex system of various structures arranged in complicated relations. Also, it keeps getting more and more complicated over time as technological advances and new research leads to the addition of more information to learn and retain (e.g., the growing importance of imaging studies).

Neuroanatomy in particular can be challenging, especially when studying deep and complex structures and their interrelationships (Schon et al., 2002; Flanagan et al., 2007). It becomes especially troublesome when much material has to be studied in a short amount of time with fewer resources; if these resources are ineffective, students are left with much discomfort and bitterness; 'Neurophobia' is frequently used to describe the fear of neuroscience (Jozefowicz, 1994; Zinchuk et al., 2010; Flanagan et al., 2007). Medical students and physicians often feel that neuroanatomy is their Achilles' heel as their knowledge of it is inadequate compared to other subjects and disciplines, leading to difficulties in diagnosing and treating neurological disorders and a decrease in preference for entering the field of neurology, while neurological conditions constitute a significant part of acute hospitalizations and inpatient care of an aging population more susceptible to these disorders (Allen, 2018; Sotgiu et al., 2020). Moreover, it is not easily possible to dissect and study complicated structures like the brain's deep cortical nuclei and white matter tracts; hence, courses do not teach their organization, function, and clinical correlates with great detail, leading to incomplete neuroanatomy education (Latini & Ryttlefors, 2019).

Cadaver-based learning in human anatomy is considered the gold standard by many academics. While it may be interesting to encounter variations in anatomical structures seen in actual donors compared to some abstract description in a textbook, it necessitates allocating a lot of time and money in the form of donor numbers, infrastructure, and other resources; also, it faces challenges like public perception, scarcity of donors, expensive facilities and labs, proper handling of cadavers, and associated health risks (Kuyatt, 2012; Wang et al., 2020). PowerPoint presentations, plastic and clay modeling, X-ray and other radiological imaging, drawing and learning surface anatomy on volunteers, peer-teaching, clinical integration, and problem-based learning (PBL) are a few of the resources and learning environments that were developed to enhance learning in anatomy to supplement cadaver-based learning.

The role of technology has been gradually increasing in education, also popularly known as E-learning, especially recently due to growth in the tech industry. This led to increased online platforms, 3D software/apps on mobile devices, computers, Virtual reality headsets, and 3D printing applications. These new technologies bring adaptability, accessibility, ease, safety, and cost-efficiency to an educational system that often lacks resources. It brought the student-friendly concept of asynchronous e-learning, where the students get to study virtually at their own time and pace, and the 'flipped classroom' approach, where students can use their official school hours more efficiently, hence the increased demand by students for recorded lectures. Also, the rendering ability and quality of images and animation have been greatly improved and are visually more realistic than static images reproduced in many traditional 2D resources.

Evolving crisis in anatomy education

Anatomical sciences education can often be challenging due to the complex and voluminous study material that sometimes has to be studied with limited time and resources. This is not a new problem; Eldred and Eldred (1961) pointed out that the deficit in teaching faculty has always been present, and the anatomical teaching hours in medical schools gradually decreased from over 800 hours at the start of 20th century to around 330 hours on an average around the 1950s. Presently this number lies somewhere in the 80-200 hours range for most schools but may further disappear in many places due to plans for a totally clinically-oriented medical education (Allen, 2018; Arantes et al., 2018).

Decreased funding is an essential issue in post-modern times of financial crisis, and courses often get restructured so that students graduate faster and cheaper (Brenton, 2011). When medical schools face much crisis of time and resources, it is not surprising that education of nursing and allied health professionals face even more severe budget cuts (Kuyatt, 2012). They are equally involved in patient care and are very interested in but systematically excluded from advanced courses in neuroanatomy or gross anatomy leading to resentment and a lack of knowledge and skills (Latini & Ryttlefors, 2019). Anatomy education at high school and undergraduate levels face similar, if not more, crises.

A significant decline in teaching hours dedicated to anatomy has been observed in many countries (Brenton, 2011; Wang et al., 2020; Craig et al., 2010), and a shortage of cadavers and other resources has been present since historic times; it was essential to keep developing more effective and efficient learning methods for teaching anatomy. Moreover, COVID-19 greatly impacted higher education and research in the health industry as many medical, dental, nursing, and other professional schools were forced to suspend or reduce in-person lectures and labs for almost a year (Attardi et al., 2020). As schools suspended classes and implemented stay-at-home policies, they were forced to adopt a distance learning/remote model for didactic teaching, where all classes were held online with the help of web-based learning (Deery, 2020; Kogan et al., 2020; Desai, 2020; Ahmed et al., 2020; Chick et al., 2020; Sahi et al., 2020; Newman &

Lattouf, 2020). A challenge in this transition was to ensure that the students have a similar learning experience and are assessed as previously to ensure their competency as graduates (Deery, 2020).

Evolution of newer technological solutions

Students frequently face problems understanding and memorizing neuroanatomy, which leads to a constant need for additional resources. Availability of cadavers can be limited due to decreased time and resource allocation; hence computer-based learning environments in the form of various online and offline platforms and software have been commonly used in neuroanatomy education by many medical schools (Naaz, 2012). These technologies keep evolving and can give beautiful and realistic human brain renderings. They are developed using actual clinical data from MRI/CAT scans and fresh/cadaveric sections, sometimes involving multi-institutional collaboration and grand projects like the Visible Human Project and Human Connectome Project. They are less expensive than a dedicated cadaver for each student. They can be easily accessed much more quickly (even remotely/online on a mobile device) and more frequently, leading to flexible learning. They also allow better interaction and manipulation of the brain through tools like reversible virtual dissection, slicing, rotation, and zooming, while retaining the tag/labels of these structures to promote and enhance exploration.

Various research has been conducted on the use and applicability of web resources and computer-assisted learning for learning anatomy, where these tools can be used to visualize and understand 3D structures and the spatial relationships between them (Wang et al., 2020). Various technologies have been tested and put in use for anatomy education, like visualization of 3D structures on a 2D screen via 3D software, use of 3D glasses (polarized and anaglyph types used in 3D movie theatres), and screens (with stereo display and digital hologram technologies), and virtual reality and mixed/augmented reality (VR/AR, where a head-mounted device is used to project 3D images to the eyes). VR/AR is a preferred way of learning by students and greatly benefits motivation, learning, and long-term retention of nominal and spatial information through reduced cognitive load on the mind of the subject, but it suffers from drawbacks such as novelty issues requiring training to mitigate it, and distractions, dizziness and discomfort issues requiring adaptation (Wang et al., 2020). While all the 3D technologies can be very effective, their use and applicability are limited by their availability and ease of access. That is why one of the most tested and used technology is 3D software for computer/mobile on a 2D screen, though virtual/mixed reality is expected to take over in the future.

The replacement of cadavers with 3D software is undesirable and frowned upon by many in the medical education community (Papa et al., 2022). Nevertheless, studies have shown that adding a 3D model increases the effectiveness of lecture-based and web-based courses (Allen, 2018; Wang et al., 2020). Even in studies where increased effectiveness is not proven, students strongly suggest a multidimensional approach and often demand lectures be supplemented with 3D resources (Hu et al., 2010; Wang et al., 2020). Hence it may be beneficial to teach lectures on interactive 3D software along with traditional 2D PowerPoint presentations or use these 3D resources as a primer or supplement, if not a replacement for cadaveric labs.

Technology-related solutions supported teaching and learning during school closures during COVID-19. Faculty leaned on various pre-developed programs to adjust quickly to this new distance learning/remote model (Deery, 2020). As the pandemic continued for over a year, faculty felt the need to continue to use and integrate new methods of online learning, and schools of higher education were forced to temporarily re-evaluate the curriculum until faculty and students were able to return to in-person classes (Deery, 2020; Kogan et al., 2020; Desai, 2020).

While an exclusively online curriculum is neither desired nor necessary post-pandemic, it is likely that some form of it will be retained as a permanent change and will mark the beginning of a new virtual era in education and research (Deery, 2020; Kogan et al., 2020; Desai, 2020). As we move forward with these web-based distance learning experiences, it is essential to evaluate the applicability and effectiveness of these programs and their role in the future classroom (Sahi et al., 2020).

Sectional Neuroanatomy and use of 3D software

Sectional neuroanatomy has been used in various studies to measure the effectiveness of visualization tools, as it is an essential aspect of anatomy and has applications in other fields like pathology, radiology, and surgical sciences (Chariker et al., 2011, 2012; Pani et al., 2013). It is based on two-dimensional representations of the (primarily three-dimensional) structures from gross specimens or radiological (MRI or CAT) scans in various planar views as traditionally printed on paper or radiology films. Due to its widespread use in clinical and experimental sciences, it has become an essential part of anatomy education. It is also a great way to learn and test whether students understand the anatomical structures in a 3D space.

However, due to limited spatial abilities and the absence of appropriate resources, students may learn neuroanatomy in a simple single-plane textbook-style view or a few of the various canonical views used in radiological imaging, enabling them to learn, memorize and identify structures only in those key views rather than having global knowledge of those structures and their spatial interrelationships, which may have worked traditionally in some fields but demand innovation due to changing situations and the emerging crisis in higher medical education (Chariker, 2009).

A complete and relevant understanding of neuroanatomy requires a more profound and higher level of learning compared to other regions of gross anatomy. Various structures can appear in totally different configurations in the planar views, as many structures are very complex in shape. Curved and C-shaped morphology of some structures (like the caudate nucleus, lateral ventricles, and fornix) can cause changes in their location, shape, and size throughout different planes which may lead to confusion and misidentification. Even structures with a fixed and simple location, easy shape, and large size (like the thalamus) can be challenging to identify if similar structures surround them. Thus, sectional neuroanatomy can be much more challenging than other body parts' sectional anatomy, requiring additional resources for 3D visualization.

Traditionally it may seem that letting the learners integrate and map the 2D information into 3D knowledge on their own, no matter how difficult it is for them to do so, may lead to better learning due to the increased effort by the learner; hence the difficulties may be desirable; however, some situations are complicated enough to block these mapping due to high mental resistance, so the best way to learn spatial transformations in such challenging situations is to see them explicitly (Naaz, 2012; Pani et al., 2005). Also, the decreasing attention span and changing psychologies of future students may not let them give the same effort compared to historical standards, and things will have to change.

Studies have shown that 3D visualization of neuroanatomical structures done before or preferably concurrently/alternating while studying sectional neuroanatomy induces better understanding by providing a better framework for the organization of spatial relationships leading to better learning, memory, and generalization of knowledge for future learning (on different modes/technologies of planar views like cadaveric sections, MRI/CAT scans or any other 2D/3D representations that they may encounter later on in the future) (Naaz, 2012; Pani et al., 2005, 2013; Chariker et al., 2011, 2012; Allen, 2018).

Again, the complexity of involved structures plays a major role in the success of these interventions, the more complex structures requiring a better approach to learning that provides the students with a better way to recall and integrate 2D and 3D information in their minds (Chariker, 2009; Naaz, 2012). For simple structures, 3D visualization may or may not offer any additional benefits over traditional methods, but it may be indispensable for complex structures.

Theories in cognitive sciences dictate that this spatial transformation of 3D objects into 2D representations depends on simplicity and ease of the condition, as it depends on concepts of 3D and 2D mapping; and spatial knowledge is global/schematic and has a tendency of spatial alignment which may project a different idea in the mind rather than the actual situation (e.g., even after having seen the globe earlier in life people may think North and South America are vertically aligned when in fact they are not, and similarly, they may not be able to figure out how shadows may project differently on the wall even for simple structures/hand gestures) (Naaz, 2012; Pani et al., 1996, 1997, 2005). Hence the use of 3D software, which not only improves spatial visualization but also shows clearly the spatial transformations of the 3D structures in a 2D plane, becomes essential for successful learning in sectional anatomy, especially for the more complex areas like neuroanatomy, which need a more accurate and organized mental model.

Enhanced engagement and deliberate practice are cognitive principles that support the success of 3D software (Kuyatt, 2012). The presence of colorful graphics and labels or tags in the 3D software are features that may not be easily or reliably possible on cadavers. Specific visual cues help the students to learn faster, even without intermittent feedback or testing, as it helps them eliminate other structures and vast possibilities that they may not have studied or encountered (Naaz, 2012). It helps them focus on a particular structure and be more confident that they understand it.

Visualization also depends on the cognitive abilities of the students. Spatial ability has been linked to the efficiency of skill and knowledge acquisition in anatomical and other STEMM fields; young adults (20-40 years) and males have been shown to have higher spatial ability, with studies linking prenatal androgen levels to developmental disorders critical in spatial abilities during adulthood, and its decline to old age (Allen, 2018). However, it is highly malleable and can be durably improved by performing complex tasks and spatial training, with studies reporting no gender difference in the final outcome; moreover, while young students (aged less than 30 years) are generally more tech-savvy, it is the students with more academic preparation and older students (more than 30 years) that do better in virtual learning using 3D software (Kuyatt, 2012). The benefits of 3D software for visualization are initially stronger in students with high spatial ability; while they may offer little to no benefit to students with low spatial ability on an initial stage or with brief exposure, they can also offer an equal advantage if given enough time for repetition or mastery of the subject (Naaz, 2012; Allen, 2018). These students may face initial mental resistance but adapt quickly and soon experience learning at the same rate as high-spatial ability students; hence spatial ability is only an entry barrier and does not affect final outcomes if given enough time (Naaz, 2012).

Spatial ability has a reciprocal relationship with spatially complex tasks where the success in a task depends on spatial ability, but these tasks themselves lead to an increase in spatial ability, leading to greater benefits over time to populations with low spatial ability (Allen, 2018). Hence the use of 3D software offering great ease and unlimited access may be more beneficial for such populations with low spatial ability, that have less cognitive ability for complicated visualizations based on 2D resources and are also impeded by limited access to cadavers and accurate models due to the lack of these resources as discussed earlier.

Reviews of the effectiveness of 3D technologies

Most software and tools used for visualization in science education are built based on the developers' intuitions and personal opinions and not on a cognitive theory-based approach; very few of them are scientifically evaluated for their effectiveness in actual learning (Naaz, 2012). It has been reported that students consider 3D resources a more effective and preferable method than 2D resources, associating it with increased motivation, excitement, satisfaction, and confidence (Arantes et al., 2018).

Various studies have shown that computer-based programs and 3D software tools are more likely to enhance anatomy learning as compared to traditional methods (lectures and textbook-style 2D resources), while other studies have given mixed results, and few even reported these new technologies to be inferior to the traditional approach which may be due to unrealistically short duration of study design not corresponding to actual course durations, use of a software which itself is too basic/simple as compared to textbook diagrams, or use in topics that are spatially too simple and do not require complex visualizations in the first place (Naaz, 2012; Arantes et al., 2018; Wang et al., 2020; Welch et al., 2020; Drapkin et al., 2015). The development and evolution of newer software and hardware bring forth new capabilities in visualization that were not possible before. Adequate and flexible learning environments and program designs that follow principles of cognitive theory can further ensure their effectiveness (Naaz, 2012).

Measure of effectiveness

Several studies have qualitatively investigated student feedback and perception of motivation in learning, but usually these studies lack a central theory of learning or evaluation and use various attributes of student feedback on a (seemingly) arbitrary basis. Moreover, only a few studies analyze the quantitative effects on performance.

Various questionnaires and survey tools have been independently developed and validated to suit specialized needs. A few commonly used among these are the IMI (Intrinsic Motivation Inventory) scale, CAP (Cognitive, Affective, and Psychomotor learning) scale, and VARK (Visual, Aural, Read/write, and Kinesthetic learning) model and inventory (Edward et al., 1989; Rovai et al., 2009; Leite et al., 2010).

It is necessary to have a proper strategy for learning as well as evaluation, to identify its strengths and weaknesses, and to understand the value and effectiveness of the educational program (Maddineshat et al., 2018). The Kirkpatrick evaluation model has been one of the most acclaimed and widely used methods of evaluating health sciences educational programs for over 50 years (Dorri et al., 2019). It assesses an educational program at four levels: (1) reaction (of the learner); (2) learning (outcomes and increase in knowledge or skills); (3) behavioral (changes and successful transfer of knowledge into practice in the workplace); and (4) results (of training, organizational performance) (Dorri et al., 2019; Frye & Hemmer, 2012; La Duke, 2017; Bates, 2004). Many course evaluations miss one or more of these levels, focus on the first two, and often neglect the last two levels (Clunie et al., 2018; Bewley & O'Neil, 2013). Using the model at all levels is the most effective educational program evaluation.

The Attention, Relevance, Confidence, and Satisfaction (ARCS) model of motivation has been another popular tool for assessing medical and health science education programs for over 40 years (Daugherty, 2019; Chang et al., 2019). It is similar to the Kirkpatrick model and shares some overlapping features with its four levels. Attention can be grouped with the learner's reaction, while the program's relevance determines learning outcomes. Also, confidence can be grouped with behavioral changes, while satisfaction can be regarded as a result of the training.

The Instructional Materials Motivation Survey (IMMS) and its Reduced IMMS form (RIMMS) are pre-validated tools to measure motivation in self-directed learning based on the ARCS model (Stephan et al., 2017; Loorbach et al., 2015).

Aspects of all these models can be combined or modified to get more comprehensive and detailed feedback on the efficacy of an educational program. Using a suitable learning/evaluation model for generating questionnaires will help effectively measure student feedback. Our study created a questionnaire based on the Kirkpatrick and ARCS models to assess the impact of 3D software on student confidence and learning in neuroanatomy education. In tailoring assessment specifically for neuroanatomy education, it is also important to elucidate which attributes of brain structures (superficial or deep-seated) are better addressed by 3D resources compared to 2D resources. Another questionnaire for confidence in topics and tests were designed to measure its efficacy with regard to particular characteristics of brain anatomy. Qualitative feedback was also collected using open-ended questions.

Objectives

Due to COVID-19, the University of Louisville restructured cadaver-based neuroanatomy labs for graduate and undergraduate students. Affected students were supposed to complete neuroanatomy labs remotely without hands-on access to cadaveric brains, spinal cords, or resin-embedded slices. To address this shift to remote instruction, an online learning module focused on identifying brain structures in 3D space and the interrelationships of these structures was created using a combination of licensed software and freely available web resources. This study aimed to evaluate the success of this learning module as a tool for students learning neuroanatomy, using the newly created surveys and test scores.

Research Questions

1. a) Was the 3D neuroanatomy learning module an effective learning tool for student learners?

   b) Was the neuroanatomy 3D module more effective when compared to traditional 2D learning resources?
2. Did the neuroanatomy 3D module differentially improve student understanding of superficial (simple) structures versus deep (complex) structures?

Research Hypotheses

1. a) The 3D neuroanatomy module was an effective learning tool for graduate and undergraduate students in a neuroscience course.

   b) The 3D module was more effective than traditional 2D learning resources for student learning. (Effectiveness was measured by higher scores on the newly developed Reaction-Relevance-Result survey, Confidence in topics survey and test results, as well as qualitative feedback.)

2. The 3D module differentially improved student understanding of deep structures to a greater extent than superficial structures.

METHODS

A prospective study was conducted in the online setting as a virtual program to replace a pre-existing in-person neuroscience lab course while retaining all of its difficulty level and complexity; hence, the results can be generalized to actual instruction and provides ecological validity (Andrade, 2018). It was designed as a one-group/within-population pre-post design for surveys and tests. As used in this study, the quasi-experimental nature of program evaluation research is more acceptable by educational and ethical standards and provides a balance between an artificially controlled experimental condition and a complex real-world situation.

Research site. The location of this project was the department of Anatomical Sciences & Neurobiology (ASNB) at the University of Louisville (UofL) School of Medicine (on all levels: data collection, data analysis, etc.). The online labs were conducted on the Visible Body (Argosy Publishing) and Anatomylearning.com web applications. While the Visible Body software was licensed through the Kornhauser Library at UofL and available for student use, Anatomylearning.com was directly available for public use on their website. Data collection was done online on the Blackboard platform.

Study design and population. The study involved implementing and evaluating an educational intervention that used a newly developed neuroanatomy learning module and surveys and tests. The study population included all students enrolled in the Fundamentals of Neuroscience (ASNB 502/602) course in the department, and no sampling was done as the expected sample size was small (n<50). The project was done with permission and under the supervision of Dr. Robert Lundy (course director) and Dr. Chad Samuelsen (coordinator of the neuroanatomy lab module in the course and co-investigator in the project) at ASNB, UofL.

Although graduate and undergraduate students have differential course requirements and difficulty levels, they attended the lectures, labs, and exams together as they had been for the previous years. All the participants were adult students (18 years and older) of various gender and ethnicity, including minorities, and there was no involvement of any particular groups, including children, prisoners, pregnant women, or participants with cognitive impairments.

Inclusion/Exclusion Criteria. The only inclusion criterion for the participants was that they must be enrolled in the ASNB 502/602 Fundamentals of Neuroscience course at ASNB, UofL. There were no exclusion criteria.

The study population consisted of 10 graduate and 25 undergraduate students, totaling 35 participants. One student dropped out mid-study. This study was not funded by any sponsor. The study qualified for a partial waiver to the requirement of document informed consent process (waiver of signed consent and use of an unsigned consent form, a preamble) according to 45 CFR 46.117(c) federal regulations for the protection of human subjects and was ruled exempt by the school's Institutional Review Board (IRB) according to 45 CFR 46.101(b) under Category 2. The outcome letter(s) are presented in Appendix A. An unsigned consent form (a preamble, see Appendix B) was used for the study. Signed consent forms were not obtained from the enrolled participants because of the following reasons:

1. The research presented minimal to no risk of harm to participants and involved no procedures for which written consent is typically required. The study involved using regular educational procedures/interventions like computer-assisted learning, collecting student survey data, and analyzing their performance on exams before and after the intervention. All procedures performed in this project were typical and usually done in traditional educational settings.
2. Participant rights and confidentiality were also protected while using official data (performance on exams) by avoiding identifiable data or sensitive/private information. All data were treated as confidential and only available to the investigators. For data analysis, the exam scores were obtained by Dr. Chad Samuelsen and deidentified after interlinking each participant's pre- and post-surveys and test scores.
3. It did not involve any special participant populations.

4. The study uses additional educational interventions to supplement the regular coursework and does not involve grouping or randomization, which did not affect the participant's rights to education and did not interfere with their student experience.
5. The post-intervention test was an integral part of the regular exams in the course as in previous years. The surveys and pre-intervention tests were voluntary and anonymous; there was no opportunity for the participant to sign an informed consent or identify themselves in any way on the survey. Implied consent was obtained by completing the pre-survey. If a prospective participant did not complete the survey, it indicated that he/she did not consent to the study.

Consent process

The IRB-approved preamble contained details of the study (investigators' details, study title, purpose, risks and benefits, data storage, and safety measures) and consent (its voluntary nature, overview by the IRB and the Human Subjects Protection Program Office). It was sent electronically to the students by the PI (principal investigator) and co-investigator, along with the introduction, recruitment, and orientation email one week before starting the study module. The participants were expected to take about 5 minutes to read the preamble and had a week to review it and participate when the study module started. They were informed about the survey and that their responses would remain anonymous and have no impact on future course evaluations. The students were sent another reminder just before the start of the neuroanatomy module to maximize the response rate. If any questions arose, the PI and co-investigator were available via phone and email. Anonymous completion of the pre-survey indicated consent. The students emailed their surveys back to Dr. Chad Samuelsen so that he could organize them and immediately remove any identifiers. The students who did not wish to participate were free to withdraw from the study through lack of response to the emails and by not logging into the Blackboard folder containing the surveys. Students could also stop a survey or pre-test in progress without penalty. The one student who dropped out of the course mid-study had only completed the pre-survey but was not given the pre-test. Another student did not submit the post-survey but gave the post-test. A few other students skipped or missed questions on the surveys and tests. All these instances were treated as missing data in the data analysis.

Data Handling and Privacy. All study personnel who distributed survey instruments or assisted in analyzing data were CITI (Collaborative Institution Training Initiative) certified. New study personnel were added to this research via an approved IRB amendment before being allowed to interact with participants. The surveys and educational tests were administered via email and in an online module on the University of Louisville Blackboard site. The de-identified data was stored on password-protected University of Louisville encrypted computers, to which only approved study personnel had access. Should the data from this study be published, individual participant's identities would always remain protected as the data did not use identifiers.

Risk to Human Participants. This was a no-risk/low-risk study. Risks to human participants included discomfort, irritation, and fatigue. However, to alleviate these risks, the surveys and educational tests were administered online and could each be completed within 15-30 minutes whenever the participant had time during the day. All students were invited to complete the survey and test, but they were free to decline to fill out all or part of them. The study used data on performance on the exam but without identifiers, so there were no specific academic, physical, psychological, social, economic, or legal risks of participating in this study. A statement regarding this was also disclosed in the preamble.

Benefits to Human Participants. The participants in this study had an opportunity to experience additional educational interventions and a better learning environment for studying the 3D anatomy of the brain. The benefit was also to medical educators who hoped to improve the Neuroscience curriculum by identifying areas in need of emphasis, resulting in better understanding and performance of students in exams and indirectly helping the students. Participants were not paid for their time to participate in this study. Instead, it was presented as an opportunity to earn five extra credit points on the practical exam by participating in the study, and it was mentioned clearly in the emails and was also implied by posting the surveys and preamble document in the Blackboard folder titled 'Extra credit opportunity.' The extra credit was awarded based on participation and not performance on the surveys or tests to prevent cheating on the pre-test, which was conducted in a non-proctored online quiz format.

Adverse Event Reporting and Data Monitoring. The investigators in the study closely monitored data collection (which did not use identifiable private information) and assured anonymity and confidentiality throughout this study. Research personnel distributing surveys also monitored for stress or frustration on the part of participants. The IRB was to be notified immediately of any adverse events, whether expected or unexpected, but none occurred.

Learning modules: The 2D learning resource was based on the lab manuals provided for the course during previous years to learn the brain's internal structures. The lab manuals contained well-labeled images of the brain, focusing on its cross-sections (sample page in Fig. 1). It consisted of two PDF documents containing sectional anatomy (horizontal and coronal sections) of the brain as it was taught in previous years (before COVID pandemic) and was made available through the Blackboard platform for the ASNB 502/602 course. They were uploaded one week prior to the start date of the 3D module/online labs. The students were advised to read it carefully as they needed to answer five questions based on its content later. They had a week to study and review these resources on their own before taking a voluntary, online pre-interventional survey and pre-test containing questions about the PDF document, which was sent the following weekend and was to be completed by Sunday evening before midnight.

The following week started with a 3D educational module serving as an online neuroanatomy lab on sectional anatomy of the brain thru online software. The school already licensed the Visible Body software, and the investigators in this study found it very useful, especially for learning an overview of the brain because of its ease of dissection and various good tools and features. Most body regions were extremely well-detailed; hence, it was used in medical school and other courses at UofL. However, we felt it needed to be better for studying cross-sectional anatomy and lacked the necessary details in the brain region; thus, we searched for another software and found Anatomylearning.com to satisfy the rest of our needs. The rendering quality and details in the brain region were much better than most other software, and their slicing tool was quite impressive, hence very useful in learning sectional anatomy. Also, it was among the few software we found that had free public use (at that time on the WebGL/browser version) and written copyright instructions for its usage. We had to go for these two software as any single one did not satisfy all needs, although a few students later complained in the qualitative feedback that they wished they had to use just one software due to their complexity and novelty issues; AnatomyLearning.com would have been a better choice for us in that case.

Lab manuals were provided to the students via the Blackboard platform, containing details of basic software functions/features and instructions to use it to get to specific views/scenes in the software. A checklist of anatomical structures was provided in the manual for each scene/view so students knew what to look for. The online labs, a mandatory component of the ASNB 502/602 course, consisted of a brief introduction to the software and neuroanatomy by Dr. Chad Samuelsen. The students were then divided into groups and used the 3D module and online lab manual to learn neuroanatomy independently under the supervision of the teaching assistants. They logged into Visible Body and AnatomyLearning.com and explored various predetermined views/scenes to observe the 3D brain structures and cross sections from different angles, zoom in/out, and click on various tags/labels, to learn about various neuroanatomical features. All the instructions were provided in the online lab manual (sample page in Fig.2). The online labs acted more as an orientation/training to use the software, and the students still had to study independently after the lab session.

They received a voluntary, online post-interventional survey the same evening, which they had to finish within two days to complete participation in the study. The first practical exam for ASNB 502/602 course was held at the end of the same week. This exam supplemented the post-interventional assessment for our study. The practical exam consisted of many texts and image-based questions on neuroanatomy, including internal features of the brain. Five pre-decided image-based questions were used as the post-interventional test to measure content mastery following the 3D module on cross-sectional anatomy.

Fig. 1: A sample page from the 2D learning resource. (ASNB 502/602 Fundamentals of Neuroscience, University of Louisville)

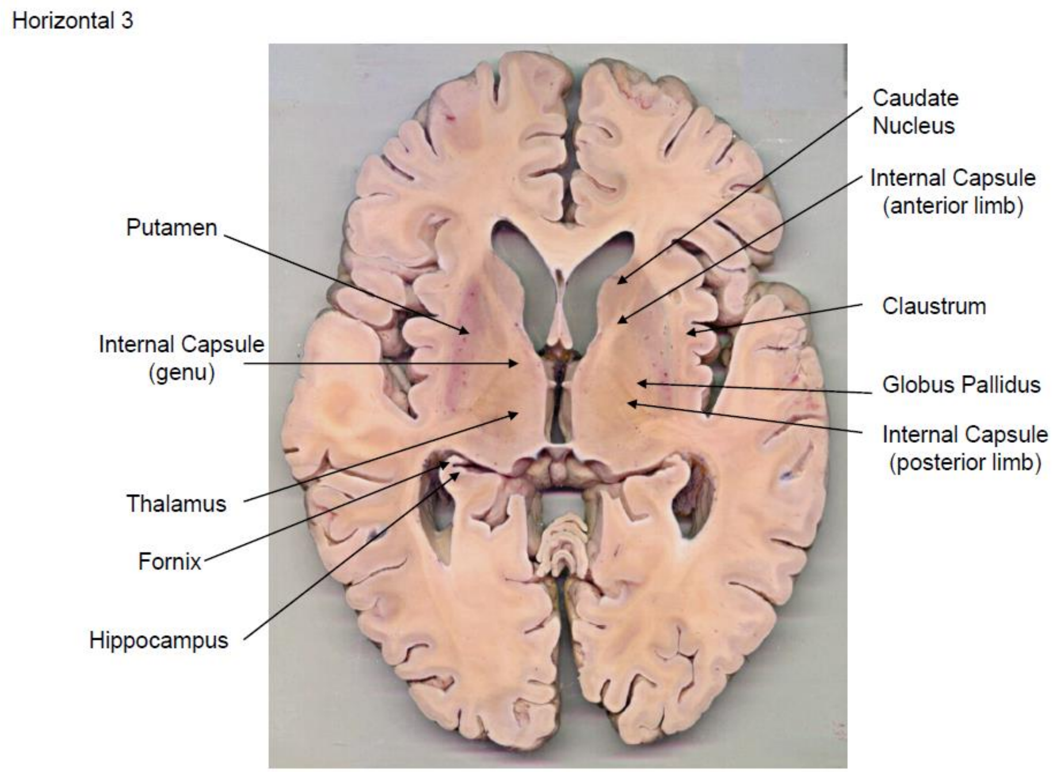
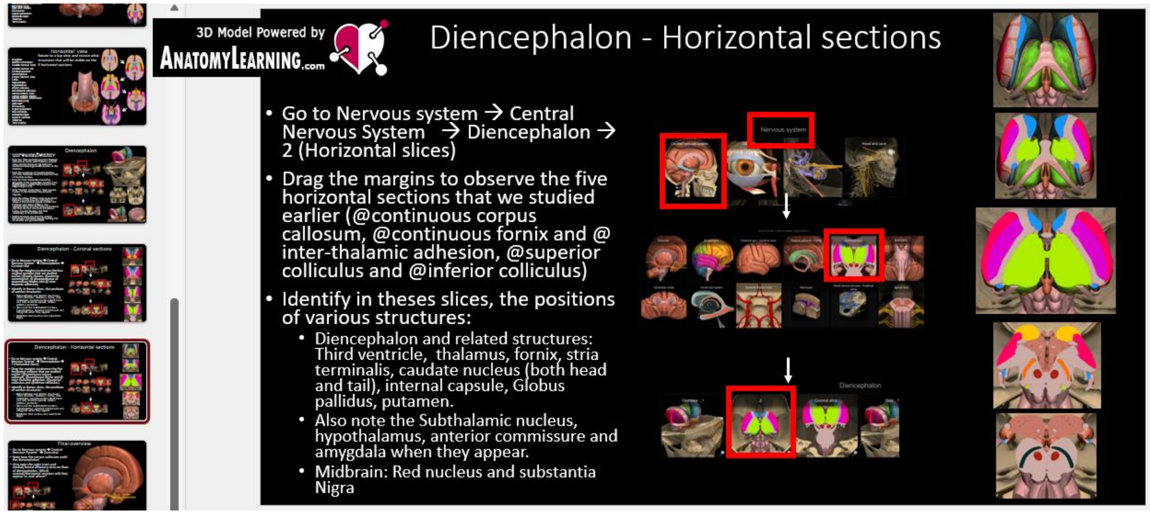


Fig. 2: A sample page from the online lab manual. 3D software available at www.AnatomyLearning.com



Measurements

Two online surveys and one online test were conducted specifically for this study. The surveys had 23 and 26 items, the online pre-intervention test had five questions, and an additional five questions were extracted from the online practical exam for post-interventional analysis as the post-test. Answering those surveys required no more than 25 minutes of the participant's time. The pre-and post-intervention test also took around 5 minutes to complete. The time commitment required per participant was about 60 minutes (25 mins each for pre- and post-survey + 5 mins each for pre- and post-test).

Survey design: Two surveys (pre- and post-interventional) were developed for this study. The surveys (detailed in Appendix C) contained questions about basic demographic information (but not personal identifiers), academic background and prior neuroanatomy experience, software usage and experience, questions regarding confidence in particular topics in neuroanatomy, open-ended questions regarding what they liked/disliked and improvements about the module, and questions modified from pre-validated questionnaires and models of learning evaluation.

Studies have defined the 'effectiveness' of computer-based tools in teaching neuroanatomy based on student performance, attitudes, and perceptions (Arantes et al., 2018). Student performance in our study was evaluated by pre- and post-interventional tests. The pre-and post-interventional surveys asked questions with themes to identify student reactions, learning experience, confidence, and satisfaction. Positive responses on these themes, measured on a 5-point Likert scale, indicated the effectiveness of the 3D module if the average score was greater than 3. Further, comparing these mean Likert scores for the 2D module and 3D module and comparing performance on the tests supported the idea of whether the 3D module is more effective than the 2D module. A similar comparison for questions regarding specific neuroanatomy structures concluded that the 3D module differentially improves student understanding of superficial or deep brain structures. Effects of gender, academic level, or previous experience in neuroanatomy were studied by observing group differential motivation, confidence, and performance following the 2D or 3D module.

A pool of questions (Table 1) was created by adapting from pre-validated models and surveys like the Kirkpatrick evaluation model, ARCS model of motivation, IMI questions, CAP scale, IMMS, and RIIMS surveys, among many others, which was then carefully vetted by the investigators and reduced to about half its size. These questions possessed face and content validity as we ensured that each level of Kirkpatrick's or ARCS model was well represented in our survey. The pre-survey, consisting of 23 questions, was developed to assess the 2D learning resources. It contained three questions on demographics and academic background, three on ‘Reaction’ (and attention), six on ‘Relevance’ (and learning experience), two on confidence and change in attitude, and two on satisfaction and overall impact. The last four questions were grouped as 'Result,' giving rise to the 13-question ‘Reaction-Relevance-Result’ (RRR) survey. The pre-survey contained five additional questions specific to confidence regarding various neuroanatomical structures and two regarding prior experience using 3D software. The second survey (post-survey), consisting of 26 questions, was similar to the first and contained the same RRR questions but focused on the 3D module (no demographic questions or prior software usage), along with questions on time spent on the 3D software (2 questions), ease and efficacy of the 3D software (3 questions), and three open-ended questions in the end regarding what they liked/disliked about the 3D module and what could be improved. While designing all these survey questions, they needed clarity and simplicity, were not double-barreled, were not loaded, and no negative wording was used (Jones et al., 2013).

Table 1: Pool of questions created for the survey development

Demographic survey
Age:
Gender:
Current Level of study: Undergraduate/Graduate student
Did you study neuroanatomy previously?
Program usage statistics:
How easy is it for you to learn to use any new software?
Have you previously used Visible Body / AnatomyLearning.com? If yes, how familiar are you with the program?
Did you have any prior experience in using other software to study neuroanatomy? If yes, then which particular one is your favorite?
How easy was it for you to use this particular software?
Regarding time spent on software during labs and before exams:
The average number of times you opened the software (0,1, 2-5, 6-10, >10):
Average time spent on each use (in minutes):
Evaluation of Reaction
How much did you enjoy studying with the software?
How much attention did you give to the software?
How much did it stimulate your curiosity and interest?
How engaged and immersed were you while using the software?
How focused or goal-oriented were you using the software (exploring on your own versus focusing on only checklist structures)?
How much did it motivate you to learn neuroanatomy, keep working on it, and become active?
How important did you feel to study it and use it on your own, or did you feel forced and did it because it was mandatory, and you had to do it?
How much effort and energy did you put into using the software?
How ready were you to use the software, or did you need more training and guidance?
Did you feel relaxed, or were you nervous, anxious, tense, or pressured while using it?
Did you feel comfortable, or did you feel dizzy, headache, nausea, or any other unpleasant feeling while using it for a longer time?
Evaluation of Learning
Overall, how useful or beneficial was the learning experience using this software?
How detailed was the text/information presented in the software?
How detailed was the 3D/visual content presented in the software?
How much did it help you acquire knowledge and identification skills and improve your understanding of the brain?
How much did it improve the visualization of complex structures?
How much did it improve your spatial understanding of structures in the brain and the interrelationships between these structures?
How much did it help to get a holistic view of the topic that you study on it?
How much did it help to relate gross anatomy to real-life clinical or laboratory situations?
How much does it match your study objectives, needs, and motives and fulfill your learning requirements?
How relatable was the content presented understandably and related to your experience and preferences?
How pleasant was the learning experience for you?
How well organized was the software, so it was easy to use and navigate?
How successfully could you follow the instructions to use the software successfully?
How easy was it to virtually dissect the brain?
How often did you struggle with handling structures while performing virtual dissections?
How useful was it for understanding 3D relationships?

How much did you feel like you had control over the success of your learning experience, or did you feel lost?
How much did it provide challenges and meaningful opportunities for learning anatomy?
How memorable was studying it to help in long-term retention and recall?
How much would you recommend for future use in other/similar courses for career development?
How much would you recommend for future use for other students in this course or other similar courses?
Evaluation of Confidence
How much did studying the software change your attitude and confidence toward neuroanatomy?
How much did its use stimulate your critical thinking or enhance reasoning skills in neuroanatomy?
How difficult do you now think neuroanatomy is actually?
How self-confident/prepared do you personally feel about neuroanatomy now?
How difficult do you think it will be to identify structures on the exam now?
Evaluation of Overall Impact
How satisfied do you feel after learning about the software?
How much more self-reliant do you feel after using the software?
How much did it help you to prepare better for exams and increase your performance on the test?
How much did it help you to improve your efficiency in the future laboratory or clinical work?
Sub-topic-specific inquiry
How easy or difficult do you think it is to learn about superficial cortical structures like cortical gyri and sulci?
How easy or difficult do you think it is to learn about deep cortical structures like subcortical nuclei or internal capsule?
How easy or difficult do you think it is to learn about C-shaped structures like the Caudate nucleus, Lateral ventricle, or Fornix?
How easy or difficult do you think it is to learn about ovoid and simple shaped structures like the thalamus or lentiform nucleus?
How easy or difficult do you think it is to learn about small sized structures like the Red Nucleus?
How easy or difficult do you think it is to learn about large sized structures like the thalamus?
How much attention and effort did you give to learning about superficial cortical structures like cortical gyri and sulci?
How much attention and effort did you give to learning about deep cortical structures like subcortical nuclei or internal capsule?
How much attention and effort did you give to learning about C-shaped structures like the Caudate nucleus, Lateral ventricle, or Fornix?
How much attention and effort did you give to learning about ovoid and simple shaped structures like the thalamus or lentiform nucleus?
How much attention and effort did you give to learning about small sized structures like the Red Nucleus?
How much attention and effort did you give to learning about large sized structures like the thalamus?
How confident are you about your knowledge and identification skills concerning superficial cortical structures like cortical gyri and sulci?
How confident are you about your knowledge and identification skills concerning deep cortical structures like subcortical nuclei or internal capsule?
How confident are you about your knowledge and identification skills concerning C-shaped structures like the Caudate nucleus, Lateral ventricle, or Fornix?

How confident are you about your knowledge and identification skills concerning ovoid and simple shaped structures like the thalamus or lentiform nucleus?
How confident are you about your knowledge and identification skills concerning small sized structures like the Red Nucleus?
How confident are you about your knowledge and identification skills concerning large sized structures like the thalamus?

Test design

The pre-and post-test consisted of questions that tested the spatial understanding of neuroanatomical structures and their inter-relationships via identifying them on tagged cross-sectional images. Such questions testing information with visible cues is more helpful as they eliminate the possibility of distraction by structures that have not been studied (Naaz, 2012). Five questions were carefully selected and used in a multiple-choice question format after approval by the study investigators and course instructors of the Fundamentals of Neuroscience course, ensuring that these structures were essential in fulfilling the learning objectives. These questions were selected based on face and content validity. They contained two questions from the superficial aspect (Lateral sulcus and Cingulate gyrus), one from the intermediate zone (Claustrum), and two from the deep aspect of the brain (Thalamus and head of caudate nucleus). The same five questions were also used to design the confidence in topics survey. The pre-test included images from the 2D learning resource (Fig. 3, derived from the same PDF files) containing tagged images of cadaveric brain sections. The post-test contained tagged images of the same five structures in a different order to ensure testing equivalence but from screenshots of the 3D software used in the study so that the performance on these images was more suggestive of the effects of the 3D intervention and possessed minimal residual effect of 2D learning. Also, the screen-captured images were used in greyscale to test generalization after 3D intervention on an image that was not as colorful as the 3D software they studied (Fig. 4).

Fig. 3: Sample pre-test image.

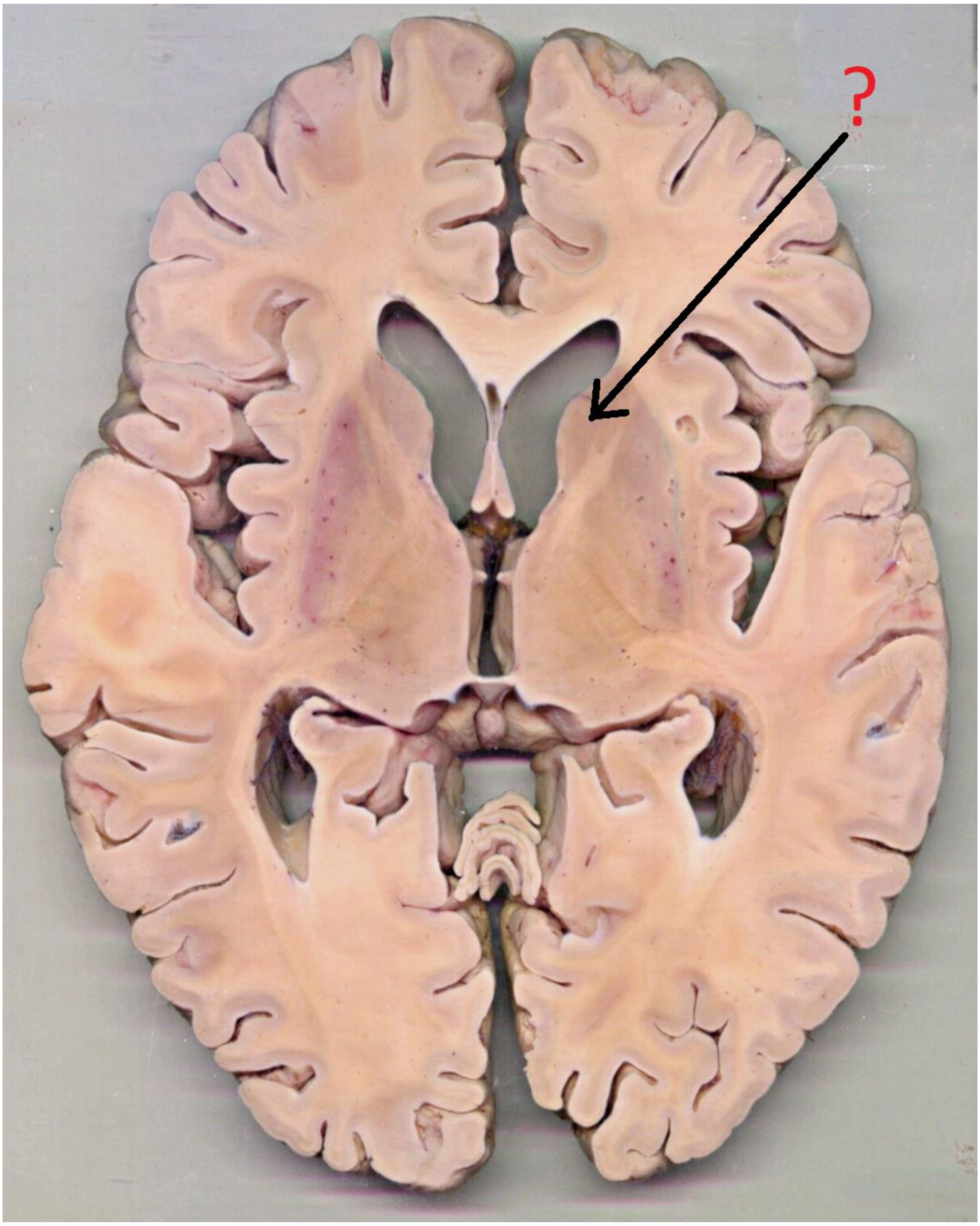
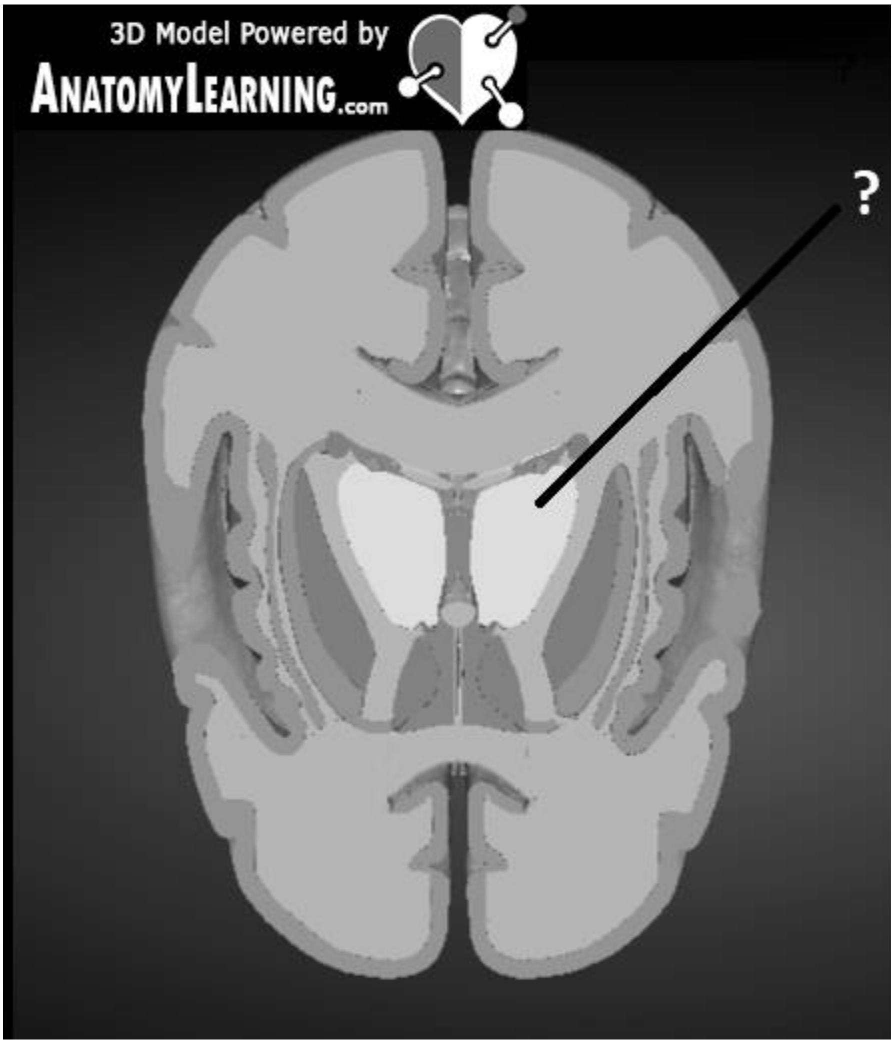


Fig. 4: Sample post-test image. Images captured from www.AnatomyLearning.com



Analysis Plan

RRR survey validity: As the RRR tool was adapted and inspired from a combination of several other pre-validated surveys to evaluate the effectiveness of the 3D module, we validated the internal structure of the post-survey using Principal Component Analysis after confirming its applicability by checking Kaiser–Meyer–Olkin (KMO) value of sampling adequacy and Bartlett's test of sphericity, and setting maximum iteration count 1000, eigenvalues greater than 1, and acceptable factor loadings greater than 0.4. As the survey tool was complex and composite, analysis without fixing the number of factors and allowing Oblimin rotation for inter-component correlation gave uninterpretable factors and inconclusive results. However, the three sub-components of the RRR survey were pre-determined; factor analysis of each of these components individually (without rotation) confirmed them to be unifactorial. Hence, we averaged the value of individual items under the respective component. Furthermore, the factor analysis of the RRR survey using only these three components confirmed it to be unifactorial, so these three components were averaged under RRR to get an overall RRR score. Factor analysis of the Confidence in topics post-survey with rotation extracted only one component; hence all five individual items could be averaged under Confidence in topics. The correlation of the averaged value of these two surveys with the final score in the practical exams provided convergent validity. The internal reliability of surveys and tests was analyzed using Cronbach's alpha coefficient of internal consistency. Missing data in this part of the analysis were handled by listwise deletion.

Quantitative analysis

The demographic data were compared by two-sided Fisher's exact tests (FET) as the comparisons involved nominal data. Data collected through the questionnaires were described as means and standard deviation (SD). Norman (2010) argued that this could be done if the distribution is reasonable enough, along with other mathematical manipulations (like change in scores), especially if a Likert scale is calculated by sums of different Likert items, just like the final score in multiple-choice-question kind of test becomes interval or ratio variable. Still, the data in this study was considered nonparametric as it was initially ordinal, and the normality of the data was questionable as it was asked on a Likert scale for a relatively small sample size (n<50). For nonparametric analysis, a Mann-Whitney U test was used to compare the responses of independent samples. A Wilcoxon signed-rank test was used to compare the differences in means of the matched pairs/dependent samples (pre- vs. post-intervention scores on the survey and test results). The pre-post increase in scores was calculated to analyze the group differences in them. The superficial-deep difference in scores was also calculated for analysis concerning the difficulty of structures.

All the results were shown along with calculated effect size measures, which play an essential part in reporting the actual strength of effects of the observed significant differences, which is required in comparison of different studies in the review of the literature and other educational research, especially if it is medium size or larger as it becomes visible in observations; the effect sizes, which are rarely reported for nonparametric tests, were described in this study as the correlation coefficient using the Rosenthal formula ["$r_{\text{effect size}}$ = r_{contrast} in two groups design" =Z/(square root of n)] where n is the total number of observations, and this effect size is interpreted same as the Pearson's r correlation coefficient (Rosenthal, 1991; Rosnow, 2003; Tomczak & Tomczak, 2014). The correlations of surveys and tests among themselves and with the final scores were explored by calculating Kendall's tau-b (τ_b) due to the presence of ordinal data.

Statistical analysis of quantitative data was done using IBM SPSS Version 24.0 software (IBM Corp., Armonk, NY). Statistical significance was established at $p < .05$. Missing data in this part of the analysis was handled by excluding cases on a test-by-test basis. Roughly half of the educational and psychological research studies contain missing data, usually 15-20%, but have been reported up to 60% (Dong & Peng, 2013). In our study, it was within 3% for tests and mostly 9% for surveys. In only a few instances, it reached up to 15% (for both software usage questions, once in RRR post-survey and twice in confidence post-survey topics). Hence, the test analysis was most reliable in case of conflicting results, followed by the RRR surveys and the confidence in topics survey (pre-survey being more reliable than post-survey).

Qualitative analysis

A thematic analysis was applied to the three open-ended answers from the post-survey via identifying important repeated themes on close examination of student feedback with an inductive and semantic approach. Firstly, the authors familiarized themselves with the data by collecting and arranging all the statements in respective groups (of questions) to give a better overview while reading it. Then in a systematic and iterative approach, key ideas were identified, marked in bold, and given codes. We kept repeating until almost all the text was marked and divided into meaningful codes. Then, these codes were analyzed, grouped, and rearranged under appropriate themes or concise codes by following the same iterative approach, leading to a significant reduction in the number of themes. In the end, we could group everything under five concise codes. Finally, the original data was reread and checked for a complete representation of all the statements into the final and concise codes, if the structure of codes looked logical and fair to the data, and if they needed any changes. Then these codes were described under the three questions we asked, citing important examples as evidence and providing validity to our conclusions.

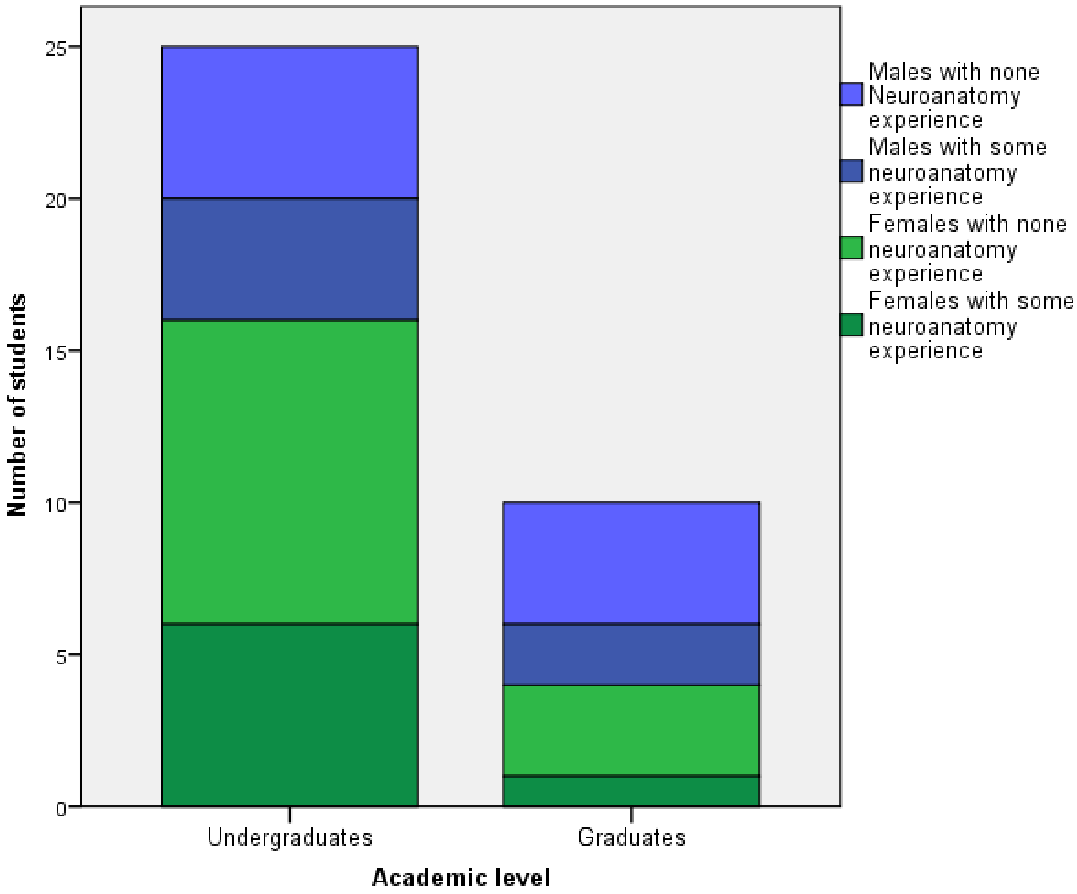
RESULTS

MAIN RESULTS

DEMOGRAPHICS

The study population, 35 students, (data graphed in Fig. 5) consisted of 15 males (42.9%) and 20 females (57.1%); this increased female composition was a result of undergraduate student composition (64% females), as compared to the opposite gender ratio in graduate students (40% females).

Fig. 5: Student demographics

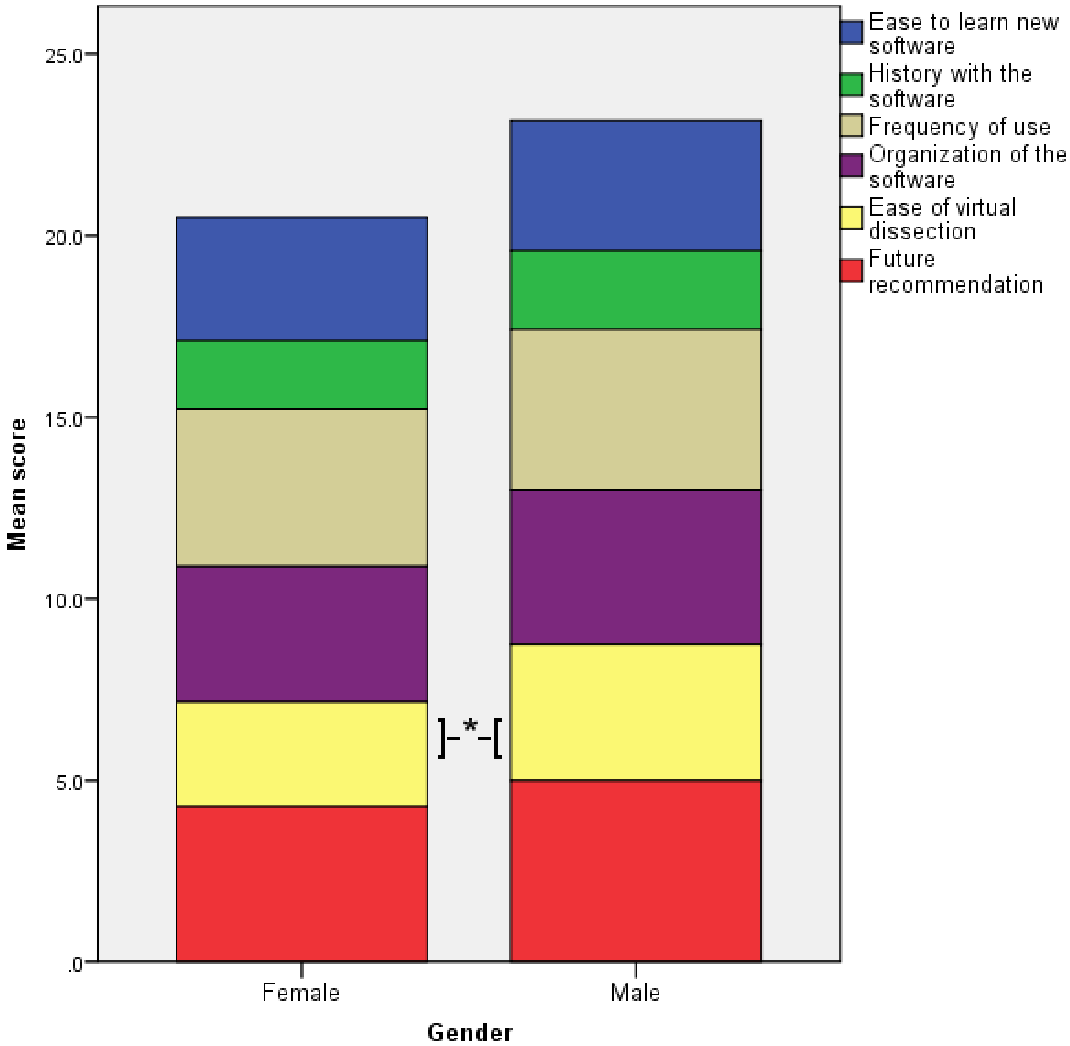


COMFORTABILITY WITH SOFTWARE USAGE

One item on the survey (Average time spent outside of class) seemed to need clarification, as it was misinterpreted by some students who reported total time over the course instead of per session. After rigorously removing the outliers, the modified mean was greatly reduced to around 51 minutes, so it can be assumed that students roughly spent an hour or less each instance they opened the software. The item ‘frequency of use’ was initially asked as an open-ended question, but was later converted to a 5-point Likert item having a mean of 4.39, which suggested that students, on average, opened the software between 5-10 times. The item 'history with the software' had a mean of 2.11 on the Likert scale, which meant that students, on average, had seen the software and maybe used it a little (but were not familiar with it much). Students reported little agreement for ease in learning any new software, and if the software used was well organized. However, they still reported high agreement in future recommendations. For all the corresponding items, the mean scores of males were generally higher than females (Fig. 6). However, statistical significance was observed only for the item 'ease of virtual dissection' in which male students expressed slight ease, and in contrast, female students expressed slight unease (M_1=3.86, M_2=2.89, U=69.5, p=.025, Z=-2.247, N=32, $r_{\text{effect size}}$=0.40) with effect size corresponding to medium (0.3< $r_{\text{effect size}}$ <0.5).

Fig. 6: Gender differences in comfortability with software usage

* Males scored higher than females in all the items and overall, but a statistically significant difference was observed only in the item ‘ease of virtual dissection’ (in yellow)

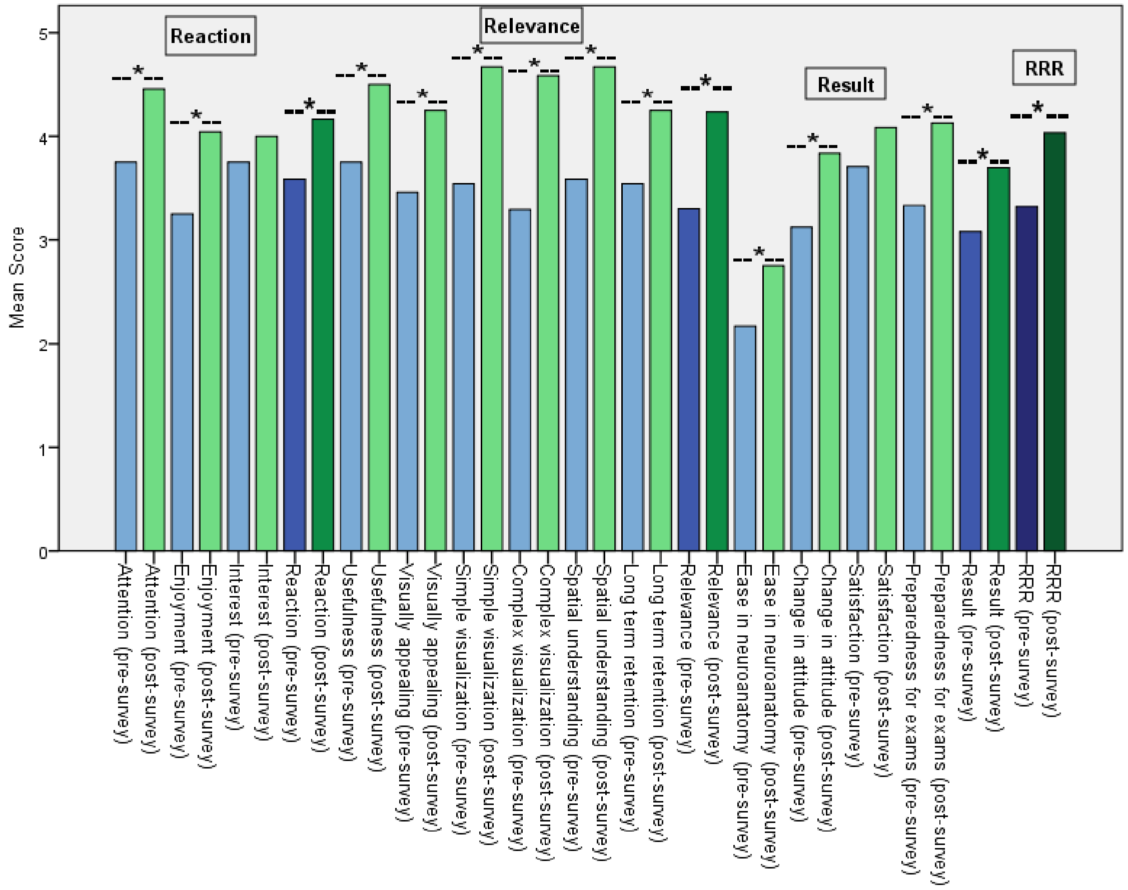


THE REACTION-RELEVANCE-RESULT (RRR) SURVEY

The mean scores of the post-survey were higher than the pre-survey for the RRR survey overall, as well as in all individual components and items (shown in Fig. 7). Almost all these differences were statistically significant (except the ‘interest’ and ‘satisfaction’ items), with effect sizes varying from large (for most individual items, all dimensions and RRR survey overall) to medium (for the 'enjoyment,' 'interest,' 'visually appealing,' 'long term retention' and 'satisfaction' items), as detailed in Tables 4-6.

Fig. 7: Outcome of the Reaction-Relevance-Result survey

* Statistically significant pre-post differences were observed in the RRR survey overall, as well as in all individual components and most items (except the ‘interest’ and ‘satisfaction’ items).

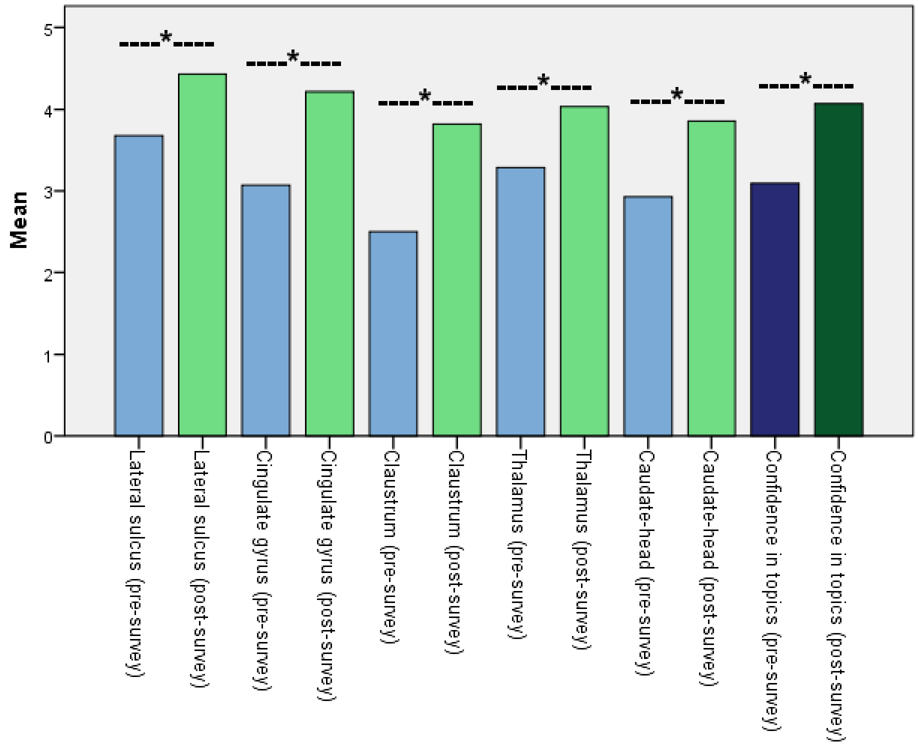


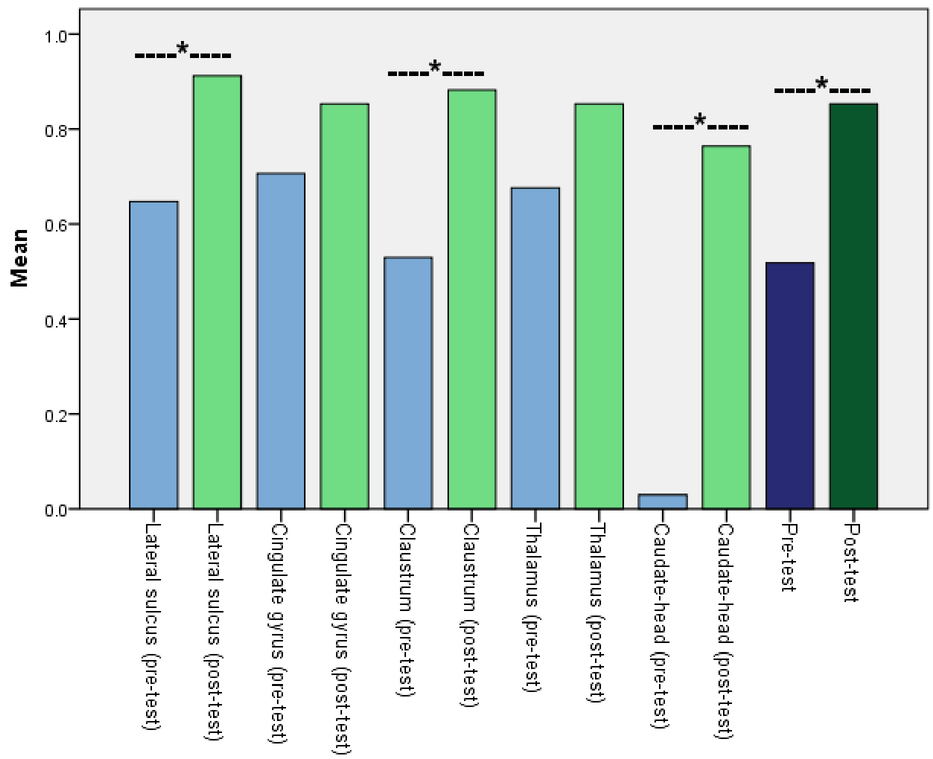
THE CONFIDENCE IN TOPICS SURVEY AND TESTS

The mean scores on the post-survey and post-test were higher than the pre-survey and pre-test, respectively, in overall as well as in all individual components and items (Fig. 8). Most of these differences were statistically significant (except for Cingulate gyrus and Thalamus in tests), with effect sizes varying from large (for most items and overall, in surveys and tests, details in Tables 8-9) to medium (for Lateral sulcus in tests).

Fig. 8: Outcome of the Confidence in topics surveys and tests

* Statistically significant pre-post differences were observed overall and in most individual items.



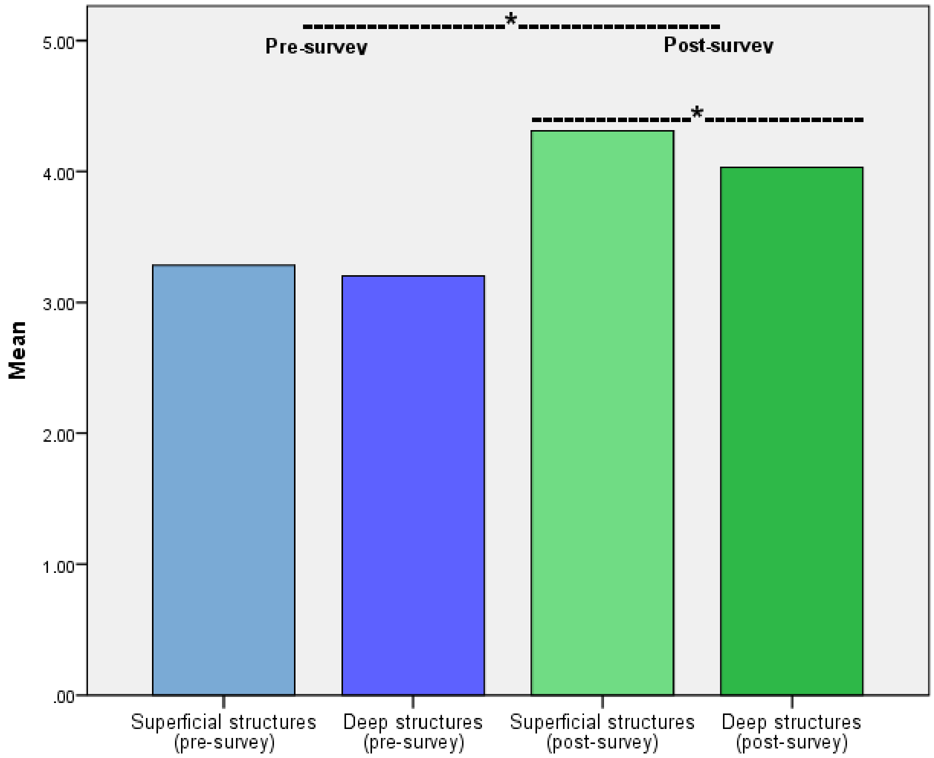


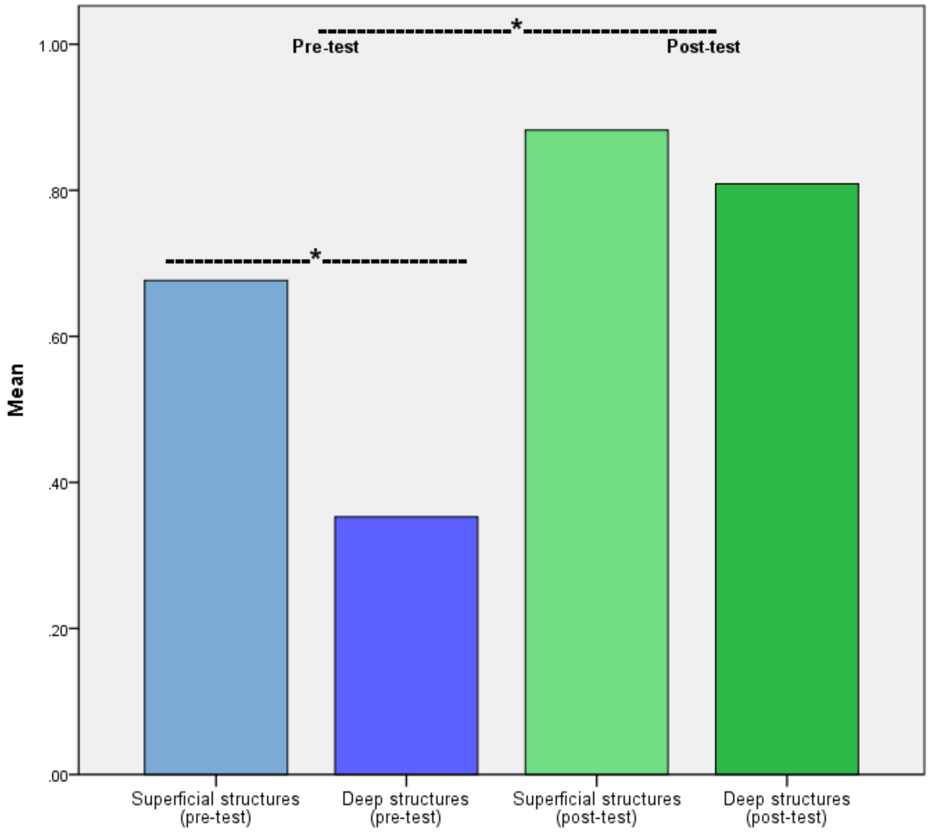
SUPERFICIAL VS DEEP STRUCTURES

As discussed earlier, the mean scores of the Confidence in topics post-survey and post-test were higher than in pre-survey and test overall. Likewise, the post-intervention scores were also greater than pre-intervention scores for the subgroups (superficial and deep structures, Fig. 9). All these differences were statistically significant, with effect sizes varying from large (for most sub-groups) to medium (for Test-superficial structures), as detailed in Table 11.

Fig. 9: Effect of the superficiality of structures on outcomes of surveys and tests.

* Statistically significant pre-post differences were observed overall and in both sub-groups (superficial and deep structures) for surveys and tests. Significance was also observed for superficial-deep differences within the post-survey and pre-test.





QUALITATIVE ANALYSIS

Regarding technology, students appreciated the rendering and accessibility they had; however, many students experienced glitches and trouble with the heavy and intensive software on their computers. They suggested that there should be one unified application that allows them to comfortably use the program as much as they want to without glitches.

Students also discussed the ease of navigation and helpful features such as zooming and rotating the view to look at different angles and delineation by color-coding and contrast. They also liked the organization of structures into different perspectives and how they could search to add or remove structures and could change their transparency, isolation, and slicing in cross-sections. Still, some problematic areas included difficulty in searching structures, isolating and handling them, and navigating through cross-sessions. They suggested improvements like better search, isolation, navigation, dissection, assembly and rotation locks and control, and the ability to add tags/labels. They also expressed the need for a greyscale view, like in real-life brain and exams, and more training in the software and its features.

Another area students discussed was content, stating that the complex structure and fine details of the brain structure were helpful. The holistic view and relations significantly helped them in spatial awareness and mental mapping, leading to genuine learning and retention. However, they needed some help with the tags' consistency and the software's absence of descriptive texts. They would appreciate it if the tags were compatible with the checklist provided and suggested having more text, clinical images, and offline supplementary material would be helpful.

Regarding realism, the students appreciated how the program was more helpful than a typical 2D resource; however, they expressed that some areas were too simple and cartoonish and lacked the feel of a real hands-on experience. They suggested having realistic colors and shape-accurate structures throughout.

Lastly, some students felt that the course duration was too short and that having more time would be incredibly helpful.

RESULTS – ADDITONAL OBSERVATIONS

DEMOGRAPHICS

According to the academic level, the population consisted of 25 undergraduate (71.4 %) and ten graduate (28.6 %) students. For previous neuroanatomy experience (with an average of 0.5 semesters per student), four students reported that they had taken two semesters, nine students reported one semester, and 22 students reported no previous experience. Furthermore, the previous experience groups were consolidated into 13 students (37.1 %) with none and 22 students (62.9 %) with some prior experience.

Two-sided Fisher's exact tests (FET) were performed, and preferred, over Chi-square tests of independence because of the small sample sizes (Kim, 2017). The FET did not reveal any statistically significant relationships between gender and previous experience (p= 1.000, FET), level and previous experience (p= .709, FET), and gender and level (p= .266, FET). These results were crosschecked with Chi-square tests of independence, which also did not reveal any statistically significant relationships.

COMFORTABILITY WITH SOFTWARE USAGE

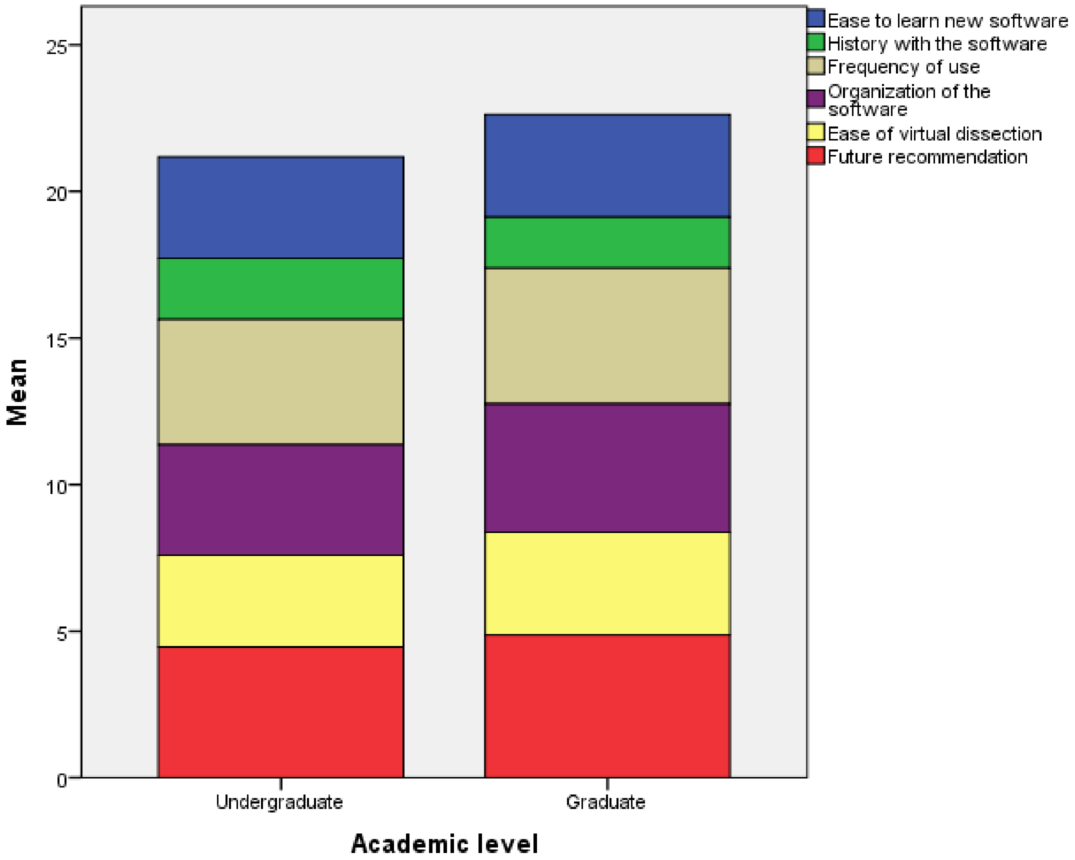
The item (Average time spent outside of class) seemed confusing and misinterpreted by some students who reported total time instead of average time spent per session. The mean of 376 minutes (6 hours 16 minutes) was exaggerated due to outliers like 1200 minutes and 1600 minutes. After removing only a few outliers (values equal to or more than 5 hours because it is less likely to study on the software for that long), the modified mean was around 67 minutes. After removing outliers more rigorously (values equal to or more than 2 hours), the modified mean was around 51 minutes. Unnecessarily removing more outliers (values more than an hour) did not change the mean much (45 minutes). The item 'frequency of use' was also changed. Initially, it was asked as an open-ended question but was later converted to a 5-point Likert item (based on the five options that were suggested to the students and used by them).

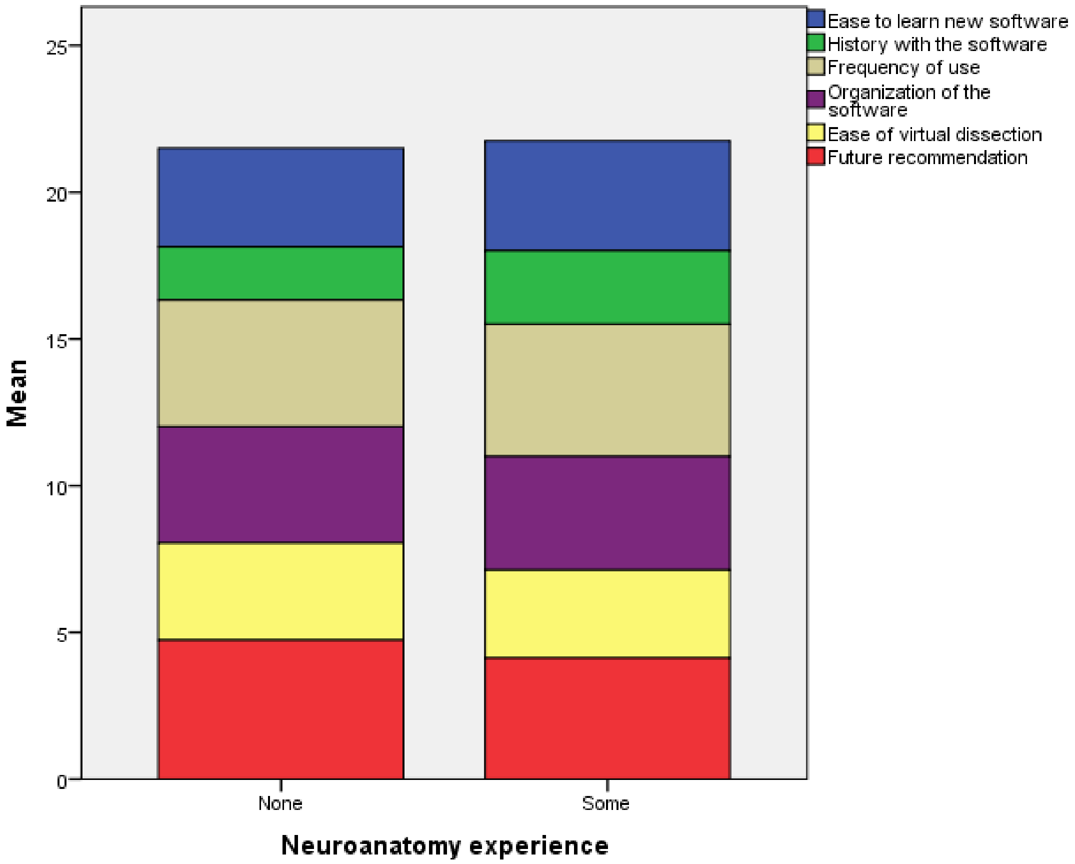
The mean scores and statistically significant group differences regarding software usage are described in Table 2. For most of the Likert scale items used in this project (except the item 'history with the software'), a mean of 3 suggests a neutral point, while 2 or 4 suggests a minor disagreement or agreement, and 1 or 4 suggests extreme disagreement or agreement with the context; higher scores meaning more favorable outcomes. No significant group difference was observed for an academic level or previous experience in neuroanatomy, and scores were similar for these groups (Fig. 10).

Table 2: Comfortability with software usage summary

Items	Mean (SD)	Significant Group differences
Ease to learn new software	3.46 (1.34)	
History with the software	2.11 (1.49)	
Frequency of use	4.39 (0.84)	
Ease of virtual dissection	3.31 (1.23)	Male>Female (M_1=3.86, M_2=2.89, U=69.5, p=.025, Z=-2.247, N=32, $r_{\text{effect size}}$=0.40)
Organization of the software	3.94 (1.01)	
Future recommendation	4.53 (0.84)	

Fig. 10: Academic level and experience differences in software comfortability





REACTION-RELEVANCE-RESULT (RRR) SURVEY VALIDITY

Principal Component Analysis (PCA) examination of the factor structure of the post-survey indicated that all the corresponding items loaded significantly under Reaction, Relevance, and Result dimensions (Table 3). While most items had good loadings (more than 0.7), four had factor loadings at satisfactory levels (0.4 - 0.7). Hence, all the items suitably represented their respective dimensions and were averaged under the respective dimension in future analysis.

All the dimensions were confirmed suitable by KMO for sample adequacy (KMO well above 0.5 for all). Bartlett's test of sphericity (*p*-values far less than .05). Cronbach's coefficient alpha for internal validity indicated high levels (α>0.8, very strong) of internal consistency for RRR survey overall, Reaction, Relevance, and confidence in topics dimensions; but only acceptable ($\alpha > 0.6$, moderate) for the Result dimension. All the dimensions were also tested for validity by correlating with the final score on the neuroanatomy exam. As expected, all the dimensions showed a strong correlation (τ_b >.32) with the final score.

Table 3. PCA results and internal validity estimates of RRR post-survey.

Dimension	Items	KMO, Bartlett's test	Factor Loadings	Cronbach coefficient alpha	Correlation with Final score
Reaction		KMO=0.72, *p*< .001	0.89	0.84	N=33, r_b=.36, *p*=.007
	Attention		0.85		
	Enjoyment		0.88		
	Interest		0.9		
Relevance		KMO=0.66, *p*< .001	0.95	0.84	N=33, r_b=.50, *p*<.001
	Usefulness		0.92		
	Visually appealing		0.69		
	Simple visualization		0.8		
	Complex visualization		0.58		
	Spatial understanding		0.82		
	Long term retention		0.79		
Result		KMO=0.71, *p*= .006	0.89	0.65	N=32, r_b=.66, *p<.001*
	Ease in neuroanatomy		0.67		
	Change in attitude		0.82		
	Satisfaction		0.64		
	Preparedness for exams		0.74		
RRR Survey		KMO=0.70, *p*< .001		0.88	N=33, r_b=.54, *p*<.001

The mean scores and statistically significant group differences for the Reaction dimension described in Table 4. This dimension and the items under it (except the 'Interest' item) had statistically significantly greater scores on the post-survey as compared to the pre-survey, with effect sizes varying from large ($r_{\text{effect size}}$>0.5, for the 'Attention' item and Reaction dimension overall) to medium (0.3< $r_{\text{effect size}}$ <0.5, for the 'Enjoyment' and 'Interest' items).

In the pre-survey, male students had statistically significantly greater scores than female students in 'Attention,' 'Interest' items, and the Reaction dimension overall, with medium effect sizes.

While in the post-survey, students with no previous experience of neuroanatomy had statistically significant greater scores in 'Attention,' 'Enjoyment' items, and Reaction dimensions overall than those with some previous experience, with medium effect sizes.

We observed a significant pre-post increase in the item 'Attention', which was greater for students with no previous experience than students with some previous experience, with a medium effect size.

Table 4: Outcomes of the Reaction dimension.

	Pre-survey		Post-survey		Pre-Post Comparison	
	Mean (SD)	Significant Group differences	Mean (SD)	Significant Group differences	Pre-Post differences	Significant Group differences
Attention	3.63 (0.94)	Male>Female (M_1=4.00, M_2=3.35, U=96.5, p=.043, Z=-2.03, N=35, $r_{\text{effect size}}$=0.34)	4.45 (0.71)	None>Some previous experience (M_1=4.68, M_2=4.00, U=64.5, p=.014, Z=-2.45, N=33, $r_{\text{effect size}}$=0.43)	Pre<Post (N=33, Z=-3.58, p<.001, $r_{\text{effect size}}$=0.62)	$\Delta_{\text{None}}>\Delta_{\text{Some previous experience}}$ (ΔM_1=1.0, ΔM_2=0.27, U=72, p=.043, Z=-2.021, N=33, $r_{\text{effect size}}$=0.35)
Enjoyment	3.23 (1.16)		4.03 (0.86)	None>Some previous experience (M_1=4.33, M_2=3.45, U=52, p=.007, Z=-2.71, N=32, $r_{\text{effect size}}$=0.48)	Pre<Post (N=32, Z=-2.41, p=.016, $r_{\text{effect size}}$=0.43)	
Interest	3.59 (1.08)	Male>Female (M_1=4.00, M_2=3.3, U=80.5, p=.031, Z=-2.16, N=34, $r_{\text{effect size}}$=0.37)	4.09 (1.04)		Pre<Post (N=32, Z=-1.82, p=.070, $r_{\text{effect size}}$=0.32)	
Reaction Overall	3.48 (0.90)	Male>Female (M_1=3.87, M_2=3.2, U=82, p=.022, Z=-2.28, N=35, $r_{\text{effect size}}$=0.39)	4.20 (0.77)	None>Some previous experience (M_1=4.45, M_2=3.70, U=54, p=.009, Z=-2.61, N=33, $r_{\text{effect size}}$=0.45)	Pre<Post (N=33, Z=-3.06, p=.002, $r_{\text{effect size}}$=0.53)	

The mean scores and statistically significant group differences for the Relevance dimension are described in Table 5. This dimension and all the items under it had statistically significant greater scores on the post-survey as compared to the pre-survey, with effect sizes varying from large (for most items and Relevance dimension overall) to medium (for the 'Visually appealing' and 'Long term retention' items).

In the pre-survey, male students had statistically significantly greater scores than female students in 'Simple' and 'Complex visualization' items and Relevance dimensions, with medium effect sizes.

While in the post-survey, students with no previous experience of neuroanatomy had statistically significant greater scores only in the 'Simple visualization' item than those with some previous experience, with medium effect size.

Table 5: Outcomes of the Relevance dimension.

	Pre-survey		Post-survey		Pre-Post Comparison
	Mean (SD)	Significant Group differences	Mean (SD)	Significant Group differences	Pre-Post differences
Usefulness	3.71 (0.82)		4.55 (0.71)		Pre<Post (N=33, Z=-3.38, p=.001, $r_{\text{effect size}}$=0.59)
Visually appealing	3.49 (1.24)		4.31 (0.86)		Pre<Post (N=32, Z=-2.62, p=.009, $r_{\text{effect size}}$=0.46)
Simple visualization	3.63 (1.26)	Male>Female (M_1=4.13, M_2=3.25, U=91.5, p=.042, Z=-2.04, N=35, $r_{\text{effect size}}$=0.34)	4.70 (0.47)	None>Some previous experience (M_1=4.82, M_2=4.45, U=77, p=.035, Z=-2.11, N=33, $r_{\text{effect size}}$=0.37)	Pre<Post (N=33, Z=-3.67, p<.001, $r_{\text{effect size}}$=0.64)
Complex visualization	3.29 (1.38)	Male>Female (M_1=3.87, M_2=2.84, U=76, p=.018, Z=-2.37, N=34, $r_{\text{effect size}}$=0.41)	4.59 (0.62)		Pre<Post (N=31, Z=-3.28, p=.001, $r_{\text{effect size}}$=0.59)
Spatial understanding	3.69 (1.13)		4.63 (0.61)		Pre<Post (N=32, Z=-3.39, p=.001, $r_{\text{effect size}}$=0.60)
Long term retention	3.54 (1.04		4.28 (0.96)		Pre<Post (N=32, Z=-2.67, p=.008, $r_{\text{effect size}}$=0.47)
Relevance Overall	3.36 (0.86)	Male>Female (M_1=3.66, M_2=3.14, U=89.5, p=.043, Z=-2.02, N=35, $r_{\text{effect size}}$=0.34)	4.30 (0.52)		Pre<Post (N=33, Z=-4.04, p<.001, $r_{\text{effect size}}$=0.70)

The mean scores and statistically significant group differences for the Result dimension and RRR survey are described in Table 6. The RRR survey and Result dimension (and all the items under it except 'Satisfaction') had statistically significant greater scores on the post-survey as compared to the pre-survey, with effect sizes ranging from large (for most items, Result dimension, and RRR survey overall) to medium (for the 'Satisfaction' item).

In the pre-survey, male students had statistically significant greater scores than female students in the 'Change in attitude' item, Result dimension, and RRR survey overall, with medium effect sizes.

While in the post-survey, students with no previous experience of neuroanatomy also had statistically significant greater scores in the 'Change in attitude' item, Result dimension, and RRR survey overall than those students with previous experience, with medium effect sizes.

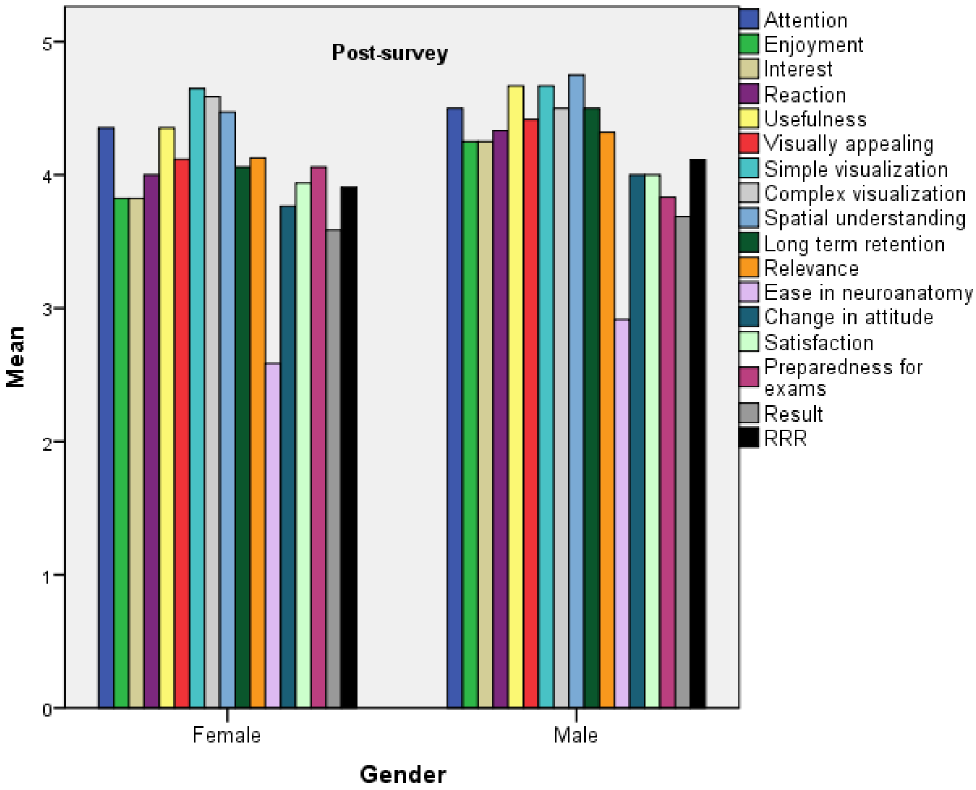
Table 6: Outcomes of the Result dimension and RRR survey overall.

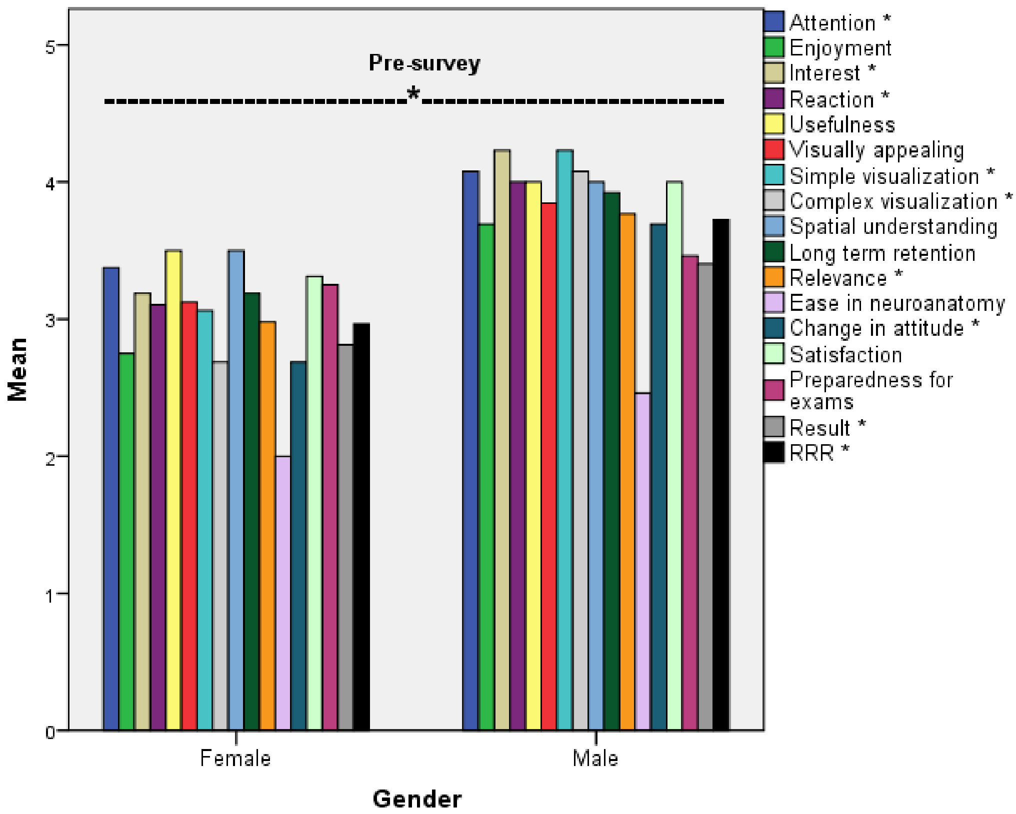
	Pre-survey		Post-survey		Pre-Post Comparison
	Mean (SD)	Significant Group differences	Mean (SD)	Significant Group differences	Pre-Post differences
Ease in neuroanatomy	2.25 (0.92)		2.84 (1.13)		Pre<Post (N=28, Z=-3.21, p=.001, $r_{\text{effect size}}$=0.61)
Change in attitude	3.12 (0.84)	Male>Female (M_1=3.53, M_2=2.79, U=74, p=.011, Z=-2.55, N=34, $r_{\text{effect size}}$=0.44)	3.91 (0.64)	None>Some previous experience (M_1=4.09, M_2=3.50, U=63, p=.020, Z=-2.33, N=32, $r_{\text{effect size}}$=0.41)	Pre<Post (N=32, Z=-3.93, p<.001, $r_{\text{effect size}}$=0.69)
Satisfaction	3.60 (0.91)		4.06 (0.95)		Pre<Post (N=32, Z=-1.92, p=.055, $r_{\text{effect size}}$=0.34)
Preparedness for exams	3.20 (1.16)		4.00 (0.92)		Pre<Post (N=32, Z=-3.57, p<.001, $r_{\text{effect size}}$=0.63)
Result Overall	3.05 (0.55)	Male>Female (M_1=3.29, M_2=2.88, U=89, p=.040, Z=-2.05, N=35, $r_{\text{effect size}}$=0.35)	3.71 (0.65)	None>Some previous experience (M_1=3.88, M_2=3.35, U=59, p=.036, Z=-2.09, N=32, $r_{\text{effect size}}$=0.37)	Pre<Post (N=32, Z=-4.00, p<.001, $r_{\text{effect size}}$=0.71)
RRR Overall	3.30 (0.66)	Male>Female (M_1=3.60, M_2=3.07, U=69.5, p=.007, Z=-2.69, N=35, $r_{\text{effect size}}$=0.45)	4.08 (0.59)	None>Some previous experience (M_1=4.25, M_2=3.74, U=65, p=.032, Z=-2.14, N=33, $r_{\text{effect size}}$=0.37)	Pre<Post (N=33, Z=-4.05, p<.001, $r_{\text{effect size}}$=0.70)

The mean scores of males were generally higher than females for all items in the pre-survey and most items in the post-survey. These gender differences were relatively larger (Fig. 11) and statistically significant (as shown in Tables 4-6) for a few individual items ('Attention,' 'Interest,' 'Simple,' and 'Complex visualization,' and 'Change in attitude'). Each RRR dimension and RRR survey overall, only in the pre-survey compared to the post-survey where differences were relatively smaller and statistically insignificant.

Fig. 11: Gender differences in outcome of RRR survey items.

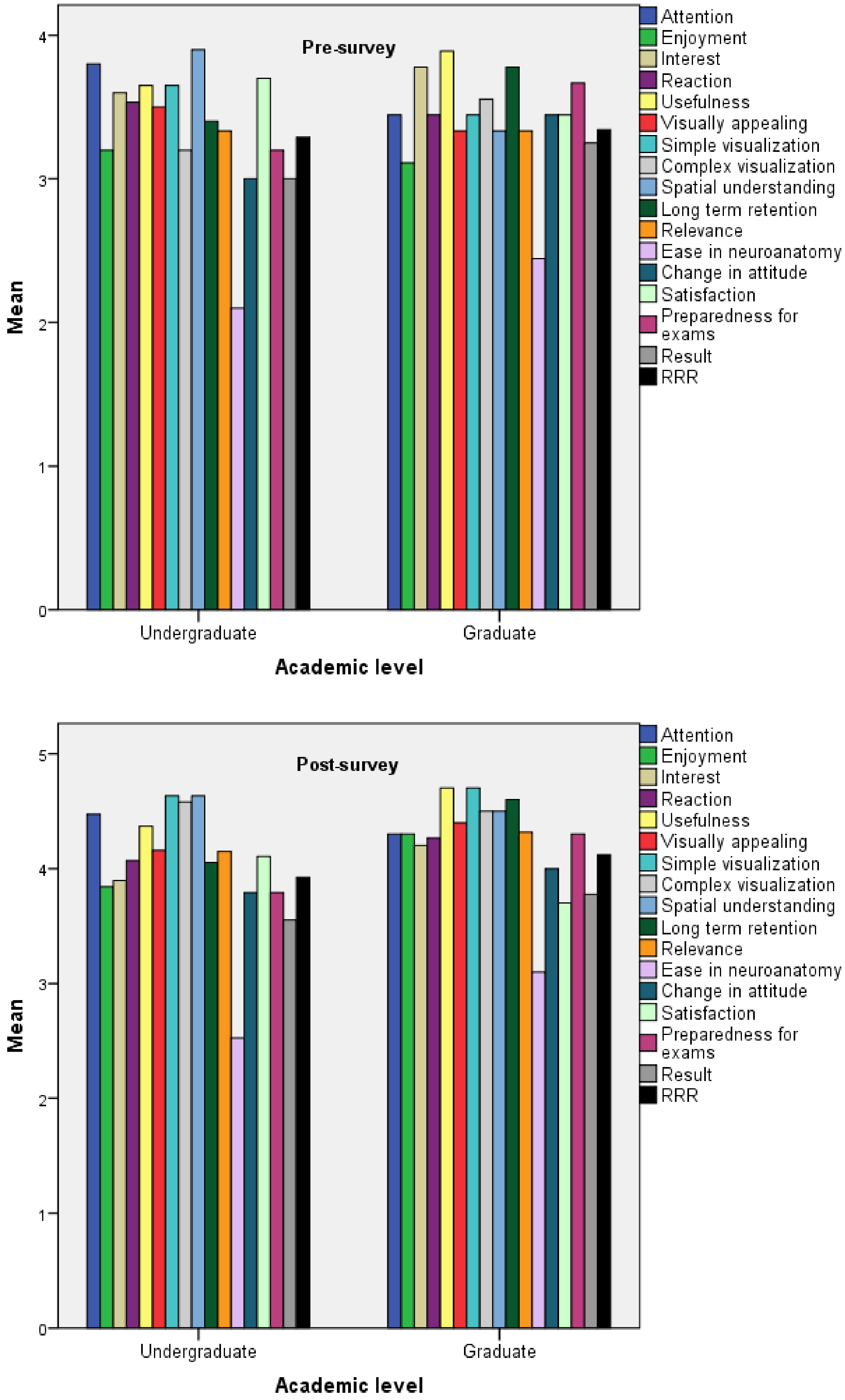
* Items, dimensions, and survey for which statistically significant gender differences were observed





The mean scores of Graduate and Undergraduate students were generally similar (Fig. 12), and no statistically significant differences were observed (Tables 4-6).

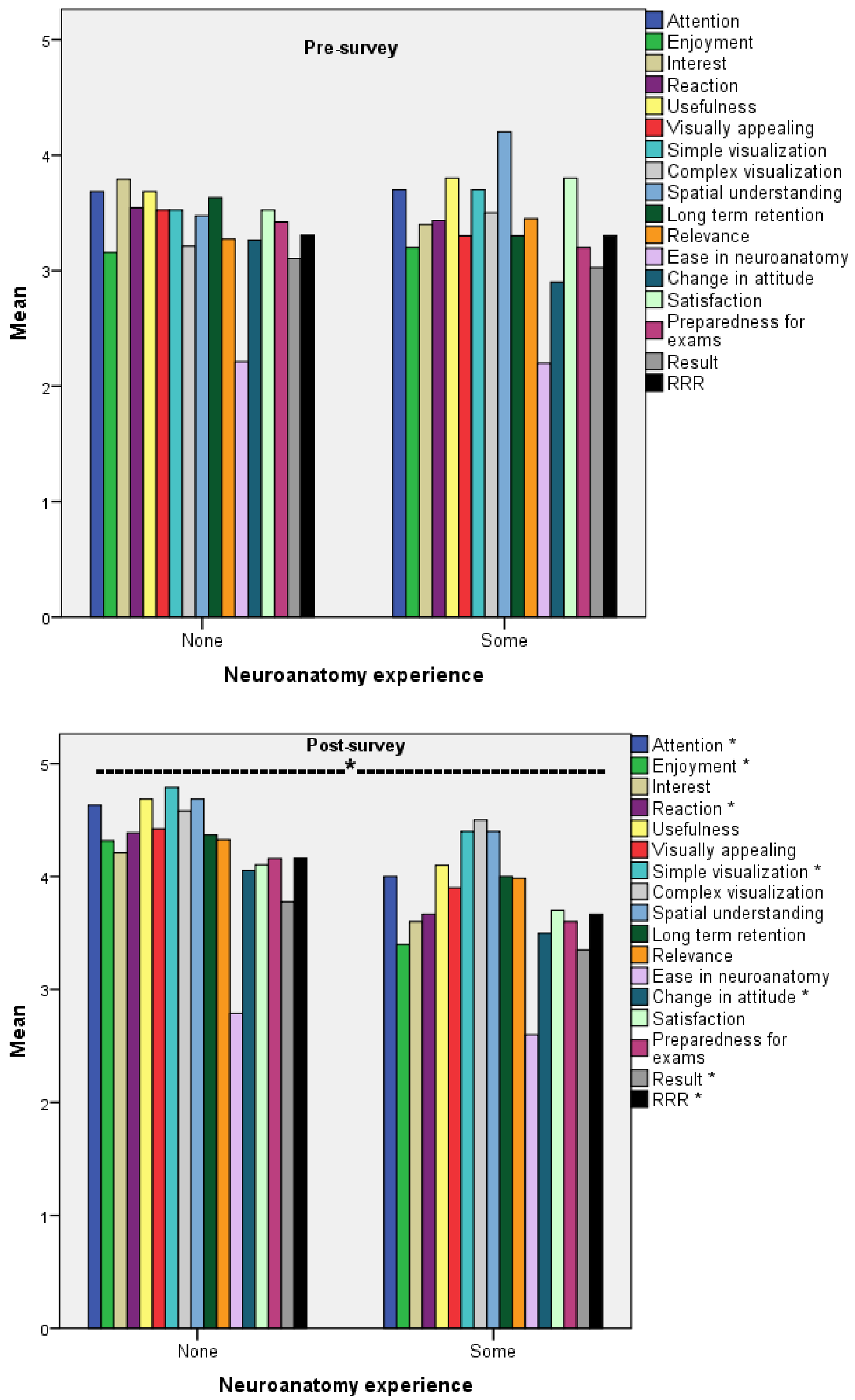
Fig. 12: Academic level differences in outcome of RRR survey items.



For the pre-survey, the mean scores of students with no previous experience of neuroanatomy were generally similar to students with some prior experience (Fig. 13), and no statistically significant differences were observed between them (Tables 4-6). However, for the post-survey, the scores of students with no previous experience were all greater than students with some previous experience. These differences were statistically significant for a few individual items ('Attention,' 'Enjoyment,' 'Simple visualization,' and 'Change in attitude'), Reaction and Result dimensions, and the RRR survey overall.

Fig. 13: Differences in outcome of RRR survey items based on previous experience of neuroanatomy.

* Items, dimensions, and survey for which statistically significant differences were observed between none and some previous experience in neuroanatomy.



THE CONFIDENCE IN TOPICS SURVEY AND TESTS

PCA examination of the factor structure of the post-survey indicated that all the corresponding items loaded significantly for the confidence in topics survey under a single factor with good loadings (for all except 'Lateral sulcus,' for which levels were satisfactory, Table 7).

It was also confirmed to be suitable by KMO (for sample adequacy) and Bartlett's test of sphericity. Cronbach's coefficient alpha for internal validity indicated high levels of internal consistency. The confidence in topics survey was tested for external validity by correlating with the RRR post-survey overall and final score on the neuroanatomy exam. As expected, it showed a strong correlation with both.

Table 7: PCA results and internal validity estimates of confidence in topics survey.

	KMO, Bartlett's test	Factor Loadings	Cronbach coefficient alpha	Correlation with RRR survey	Correlation with Final score
Confidence in Topics Survey	KMO=0.71, $p < .001$		0.82	N=32, $\tau_b=.57$, $p<.001$	N=32, $\tau_b=.57$, $p<.001$
Lateral sulcus		0.6			
Cingulate gyrus		0.85			
Claustrum		0.76			
Thalamus		0.79			
Caudate-head		0.83			

The mean scores and statistically significant group differences for the Confidence in Topics Survey are described in Table 8. All the items had statistically significant greater scores on the post-survey with large effect sizes, which was also true for overall confidence in topics.

While no significant group differences were observed in the pre-survey, on post-survey, students with no previous experience of neuroanatomy had statistically significant greater scores in the 'Thalamus' item and overall than those students with some previous experience, with medium effect sizes.

Table 8: Outcomes of the Confidence in Topics Survey.

	Pre-survey		Post-survey		Pre-Post Comparison
	Mean (SD)	Significant Group differences	Mean (SD)	Significant Group differences	Pre-Post differences
Lateral sulcus	3.57 (1.31)		4.44 (0.72)		Pre<Post (N=32, Z=-3.12, p=.002, $r_{\text{effect size}}$=0.55)
Cingulate gyrus	3.00 (1.11)		4.19 (0.96)		Pre<Post (N=32, Z=-3.91, p<.001, $r_{\text{effect size}}$=0.69)
Claustrum	2.60 (1.17)		3.91 (1.25)		Pre<Post (N=32, Z=-3.98, p<.001, $r_{\text{effect size}}$=0.70)
Thalamus	3.35 (1.20)		4.07 (0.87)	None>Some previous experience (M_1=4.35, M_2=3.50, U=50, p=.018, Z=-2.37, N=30, $r_{\text{effect size}}$=0.43)	Pre<Post (N=29, Z=-3.01, p=.003, $r_{\text{effect size}}$=0.56)
Caudate-head	3.11 (1.23)		3.94 (0.89)		Pre<Post (N=31, Z=-2.99, p=.003, $r_{\text{effect size}}$=0.54)
Confidence in Topics Overall	3.13 (0.86)		4.11 (0.72)	None>Some previous experience (M_1=4.30, M_2=3.7, U=58.5, p=.035, Z=-2.10, N=32, $r_{\text{effect size}}$=0.37)	Pre<Post (N=32, Z=-4.44, p<.001, $r_{\text{effect size}}$=0.79)

The mean scores and statistically significant group differences for the Test items are described in Table 9. Most of the items (except the 'Cingulate gyrus' and 'Thalamus' items) had statistically significant greater scores on the post-test, with effect sizes varying from large (for the 'Claustrum' and 'Caudate-head' items) to medium (for the 'Lateral sulcus' item), which was also true for the Tests overall (significant difference and having a large effect size). Gender differences were observed for the 'Cingulate gyrus' item, with female students having a greater pre-post increase than male students, with a medium effect size. While no significant group differences were observed in the pre-test, on the post-test, male students had statistically significant greater scores only in one item ('Thalamus') than female students, with a medium effect size.

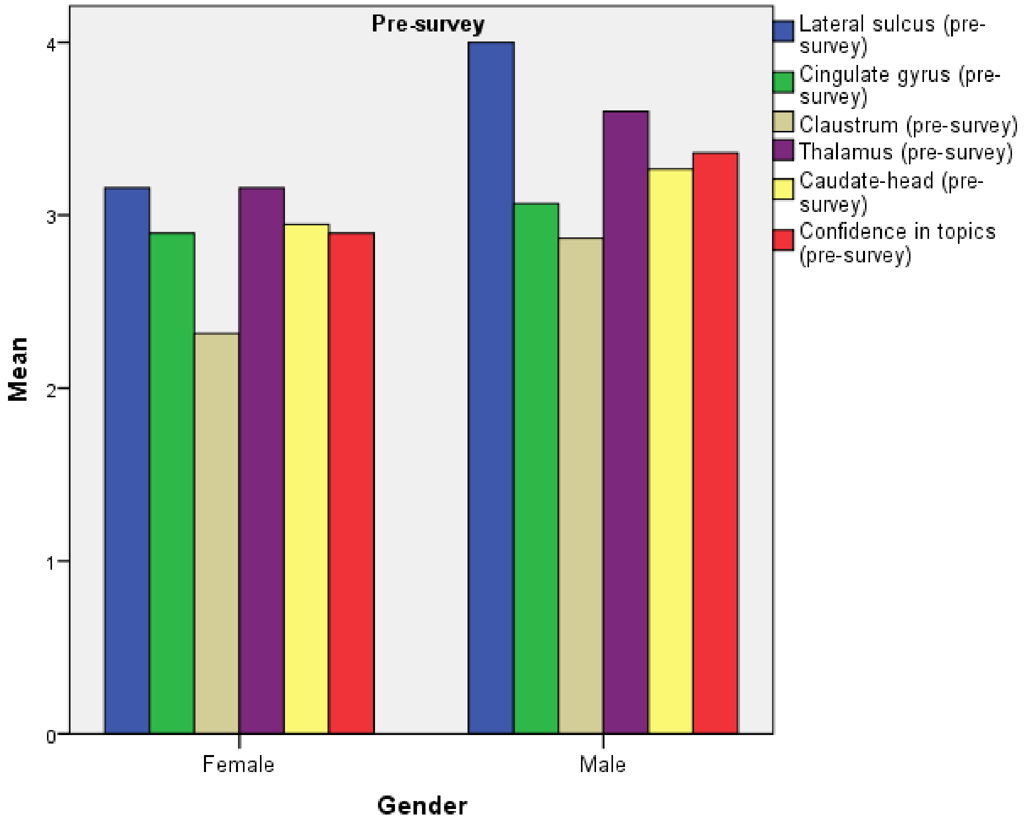
Table 9: Outcomes of the Test items.

	Pre-test		Post-test		Pre-Post Comparison	
	Mean (SD)	Significant Group differences	Mean (SD)	Significant Group differences	Pre-Post differences	Significant Group differences
Lateral sulcus	0.65 (0.48)		0.91 (0.29)		Pre<Post (N=34, Z=-2.50, p=.013, $r_{\text{effect size}}$=0.43)	
Cingulate gyrus	0.71 (0.46)		0.85 (0.36)		Pre<Post (N=34, Z=-1.51, p=.132, $r_{\text{effect size}}$=0.26)	$\Delta_{\text{Female}}>\Delta_{\text{Male}}$ (ΔM_1=0.32, ΔM_2=-0.07, U=94.5, p=.043, Z=-2.02, N=34, $r_{\text{effect size}}$=0.35)
Claustrum	0.53 (0.51)		0.88 (0.33)		Pre<Post (N=34, Z=-3.21, p=.001, $r_{\text{effect size}}$=0.55)	
Thalamus	0.68 (0.48)		0.85 (0.36)	Male>Female (M_1=1.0, M_2=0.74, U=105, p=.034, Z=-2.12, N=34, $r_{\text{effect size}}$=0.36)	Pre<Post (N=34, Z=-1.90, p=.058, $r_{\text{effect size}}$=0.33)	
Caudate-head	0.03 (0.17)		0.76 (0.43)		Pre<Post (N=34, Z=-5.00, p<.001, $r_{\text{effect size}}$=0.86)	
Test Overall	0.52 (0.30)		0.85 (0.26)		Pre<Post (N=34, Z=-4.10, p<.001, $r_{\text{effect size}}$=0.70)	

The mean scores of males were generally higher than females for most of the items and overall in the Confidence in topics pre-survey, pre-test, and post-test (except the 'Cingulate gyrus' item in post-test, Fig. 14-15), with statistical significance, observed only in a post-test item 'Thalamus' (Table 9).

However, for the Confidence in topics post-survey, the mean scores of males were very similar to females, again with no statistically significant difference observed.

Fig. 14: Gender differences in the outcome of Confidence in topic surveys.



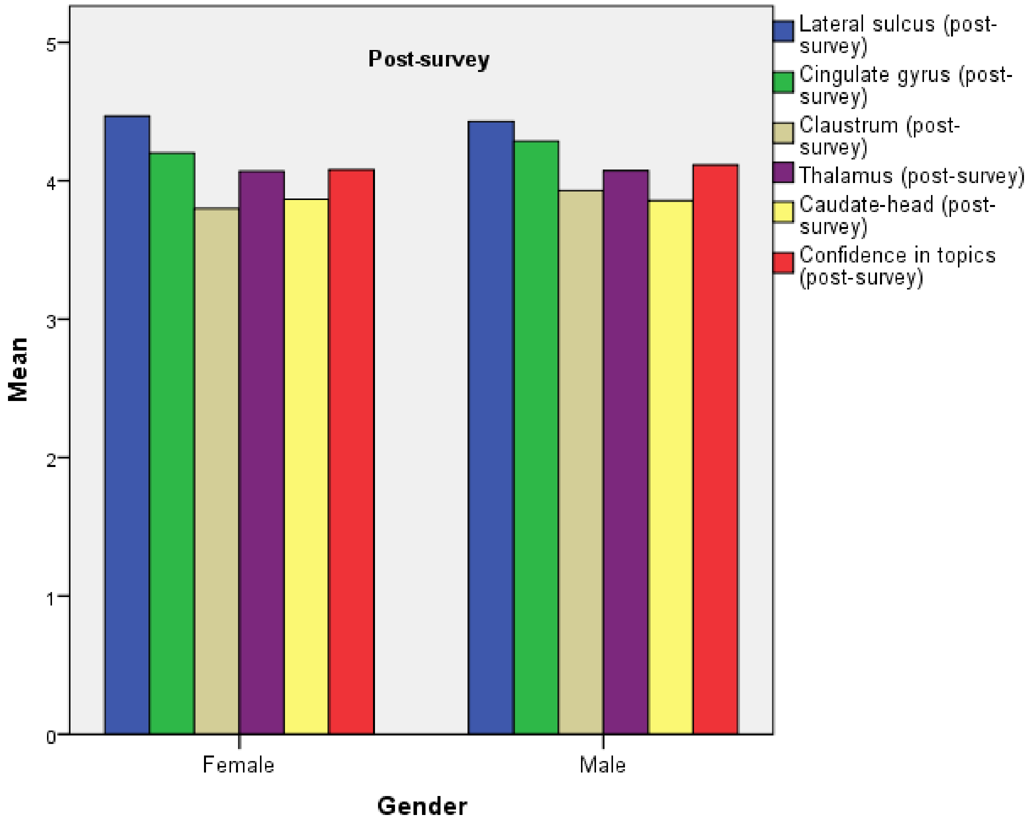
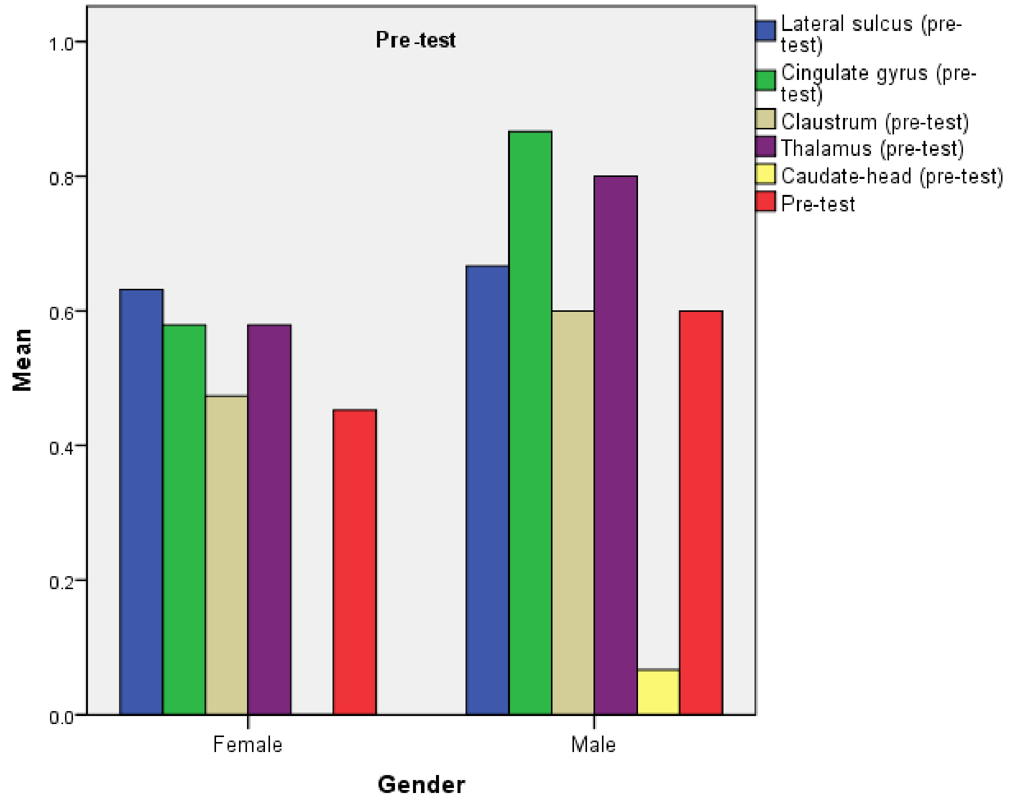
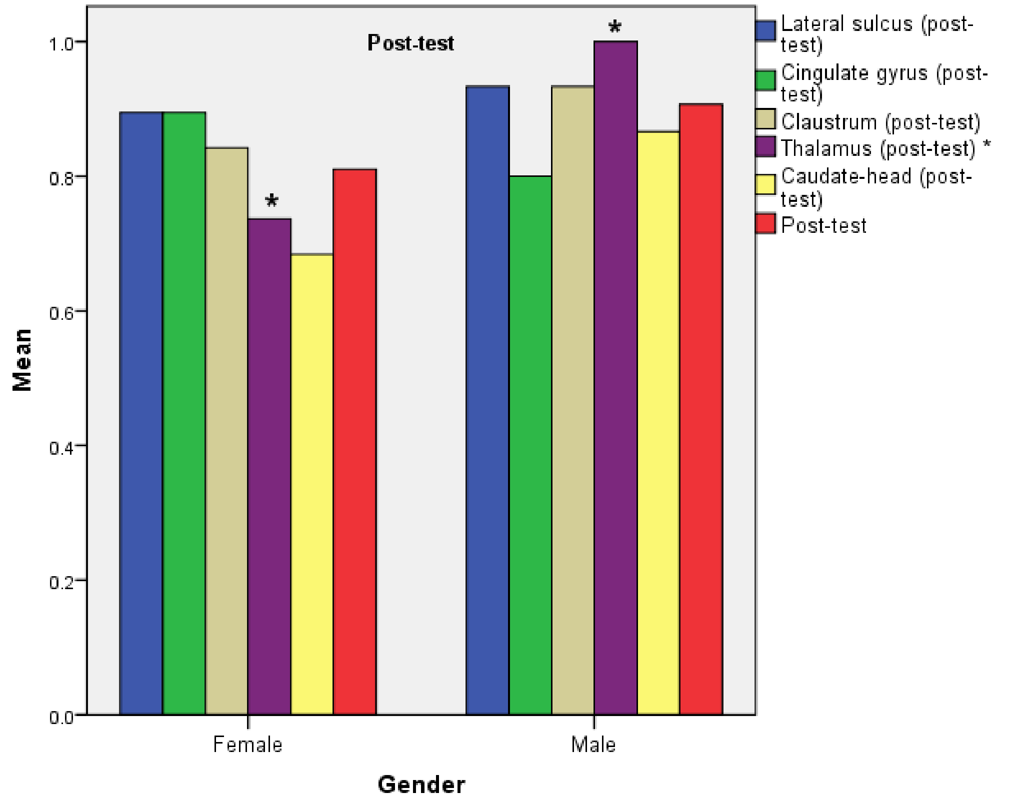


Fig. 15: Gender differences in the outcome of tests.

* Item ('Thalamus' post-test) where a statistically significant gender difference was observed



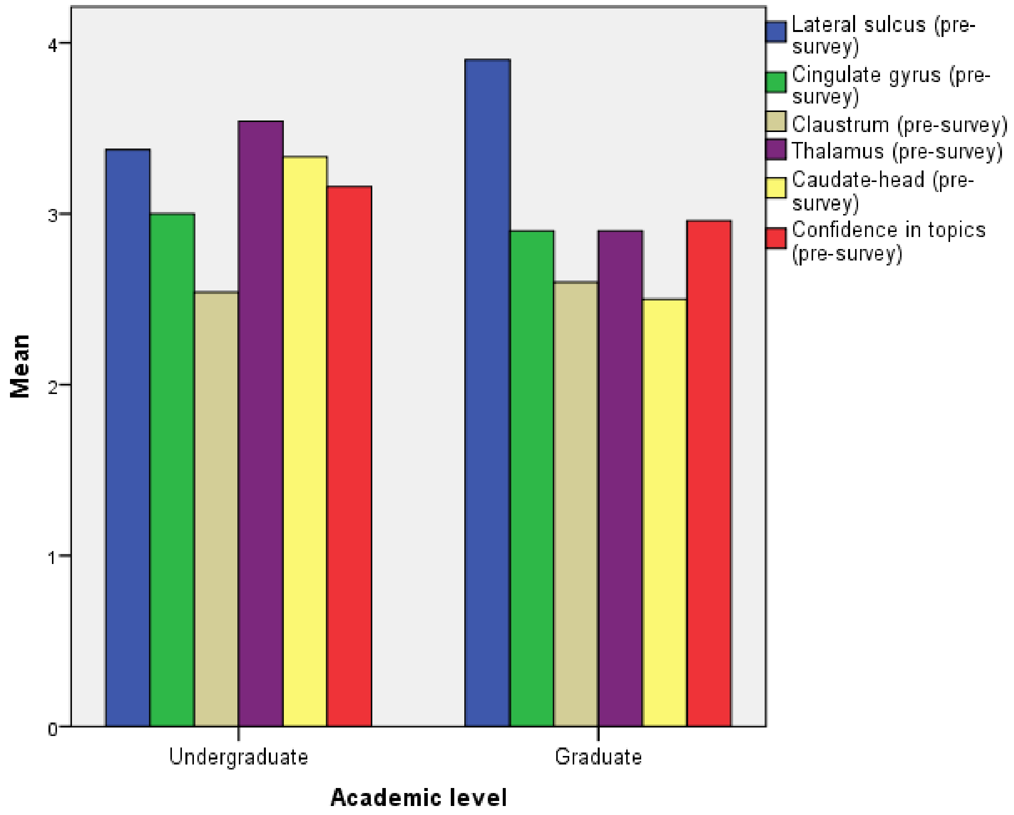


The mean scores of undergraduate students were generally higher compared to that of graduate students for most items and, overall, in the Confidence in topics pre-survey, post-survey, pre-test, and post-test (Fig. 16-17).

These differences were relatively smaller, and scores were more similar post-survey and post-test.

However, no statistically significant academic level difference was observed in any of the Confidence in topics surveys or tests (Tables 8-9).

Fig. 16: Academic level differences in the outcome of Confidence in topic surveys.



.

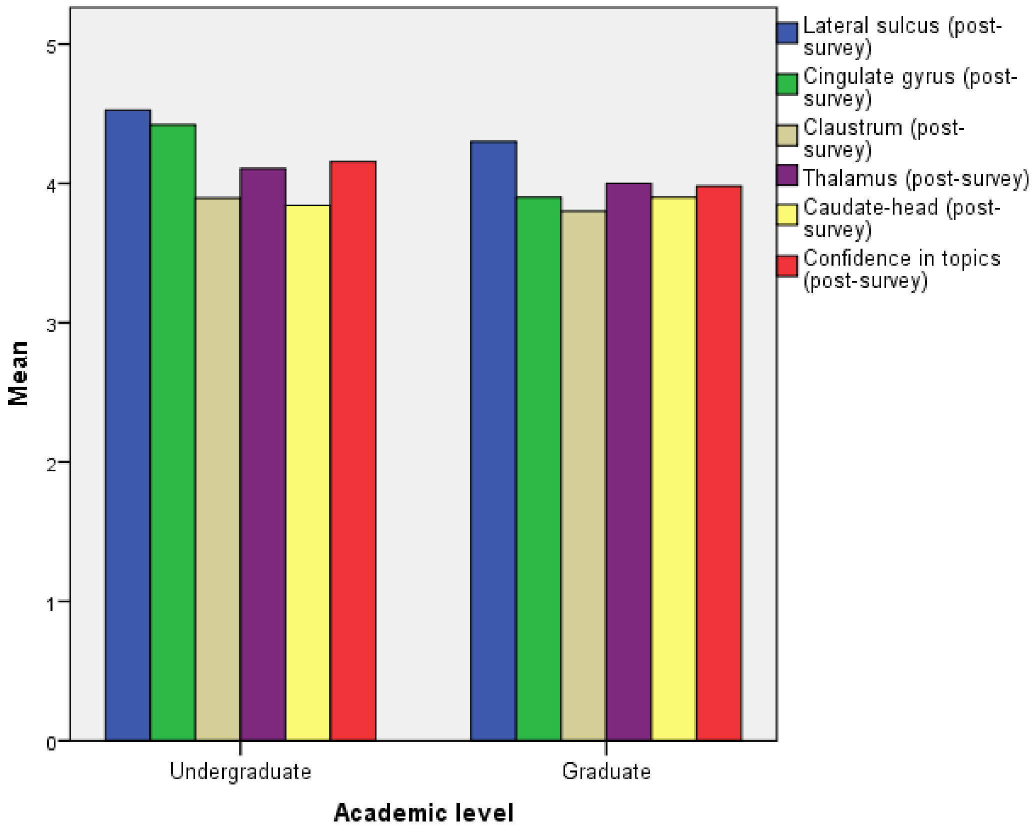
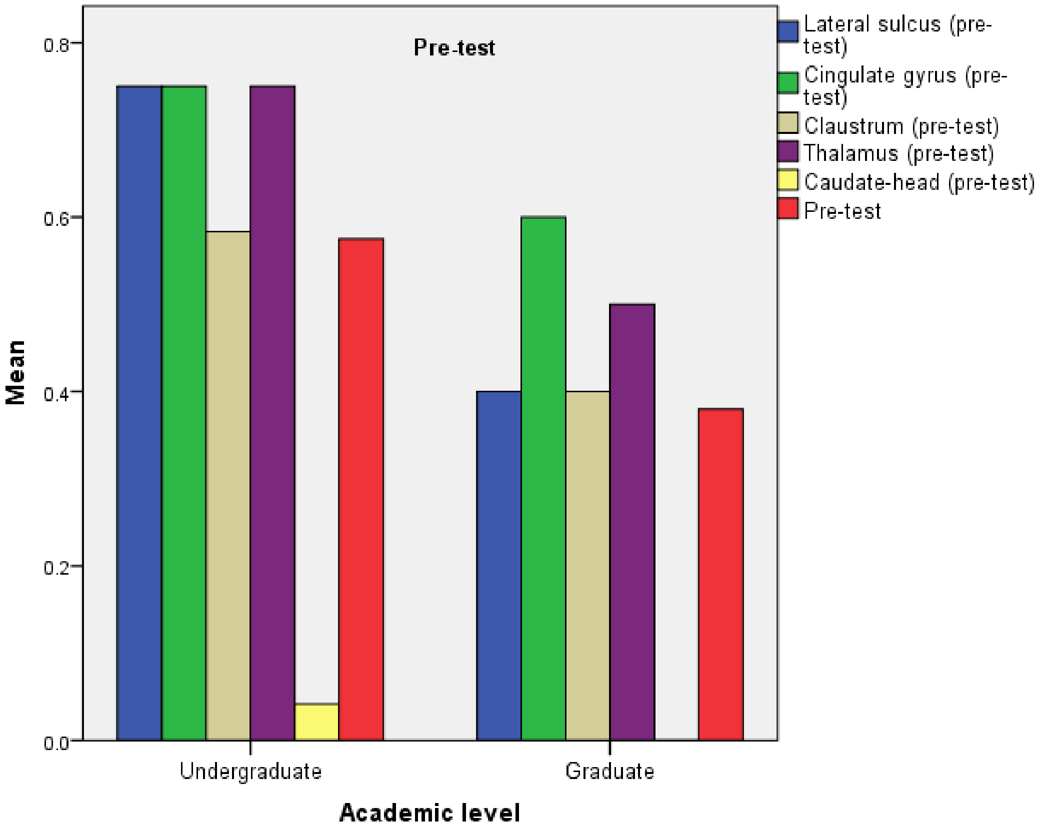
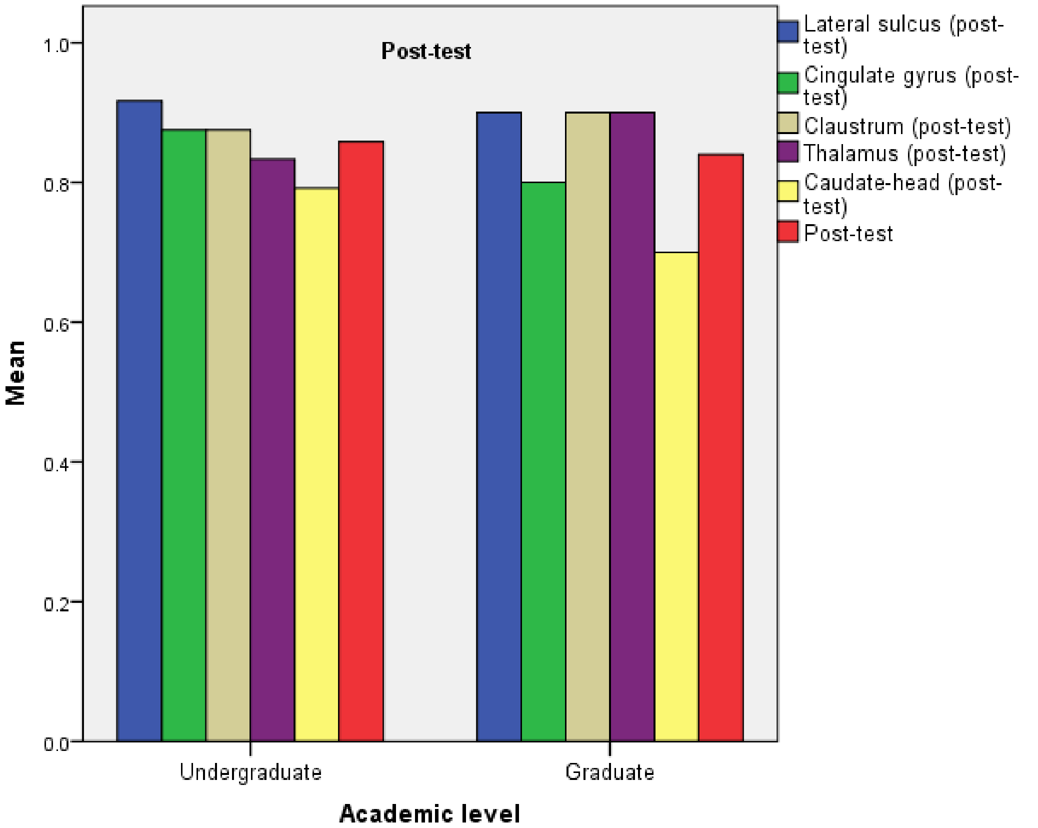


Fig. 17: Academic level differences in the outcome of tests.



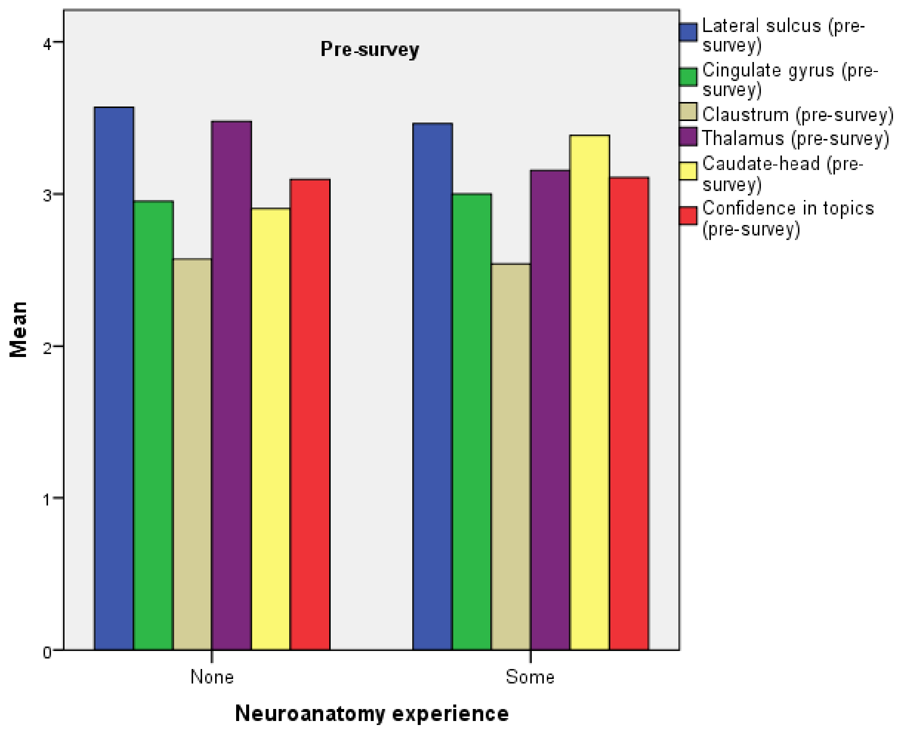


In the post-survey, students with no previous experience of neuroanatomy scored higher than students with some prior experience in all items and overall (Fig. 18). Statistical significance was observed only in the item 'Thalamus' and overall, with medium effect sizes (Table 8).

No statistically significant difference due to previous experience with neuroanatomy was observed in the pre-survey, pre-test, and post-test (Tables 8-9, Fig. 18-19).

Fig. 18: Differences in the outcome of Confidence in topic surveys, based on previous experience of neuroanatomy

* Item and survey for which statistically significant differences were observed between students with none or some prior neuroanatomy experience.



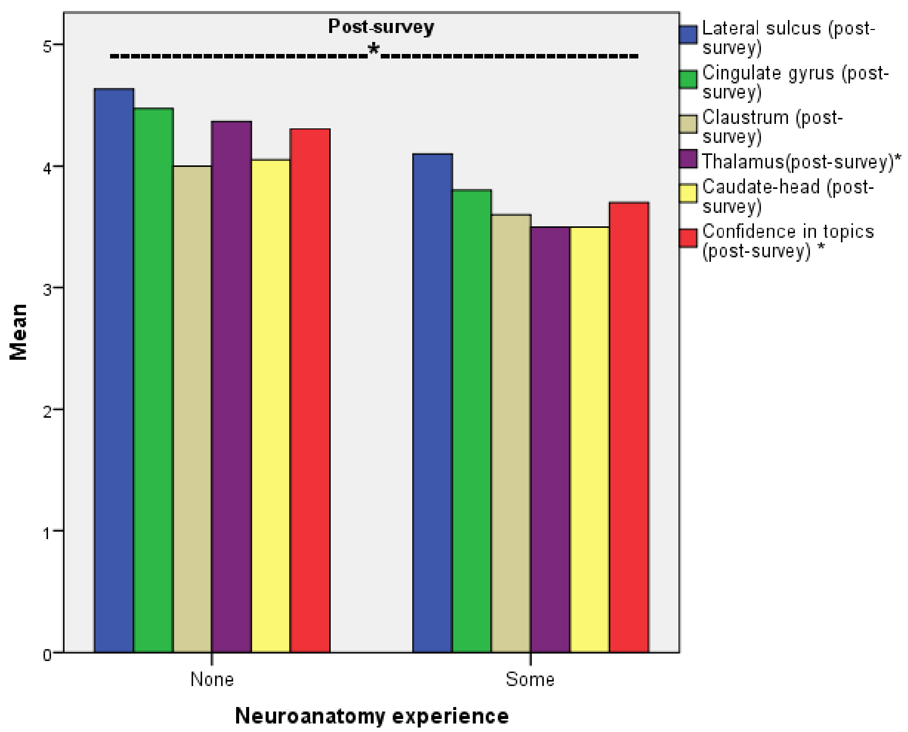
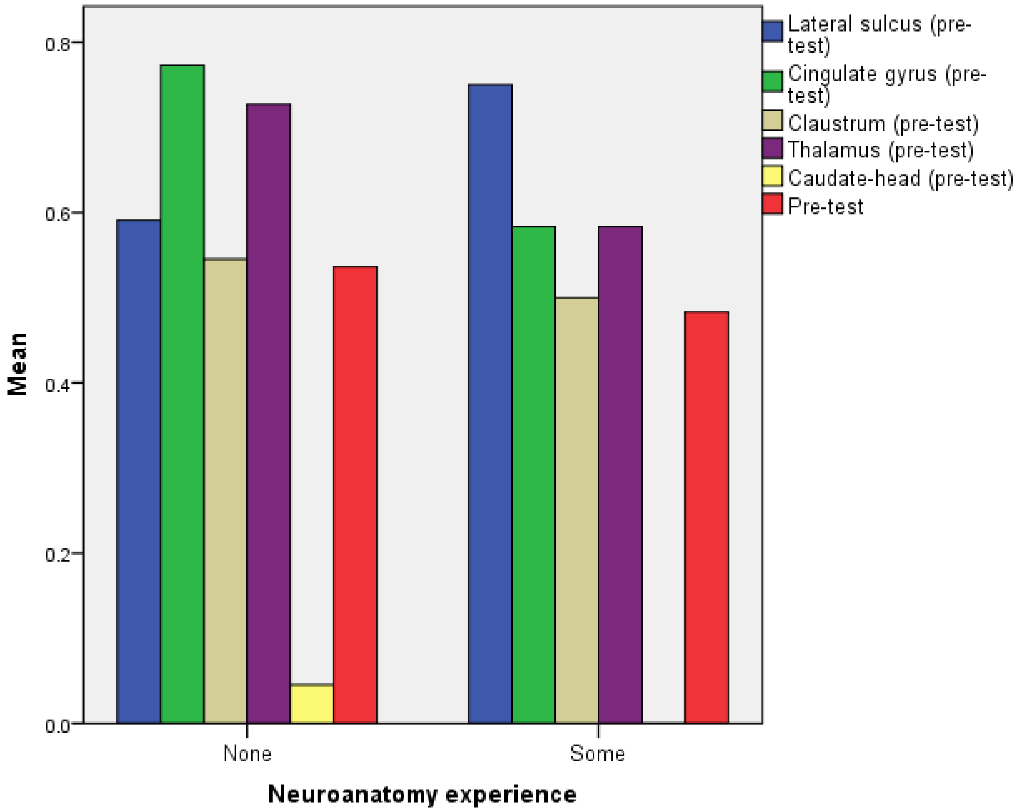
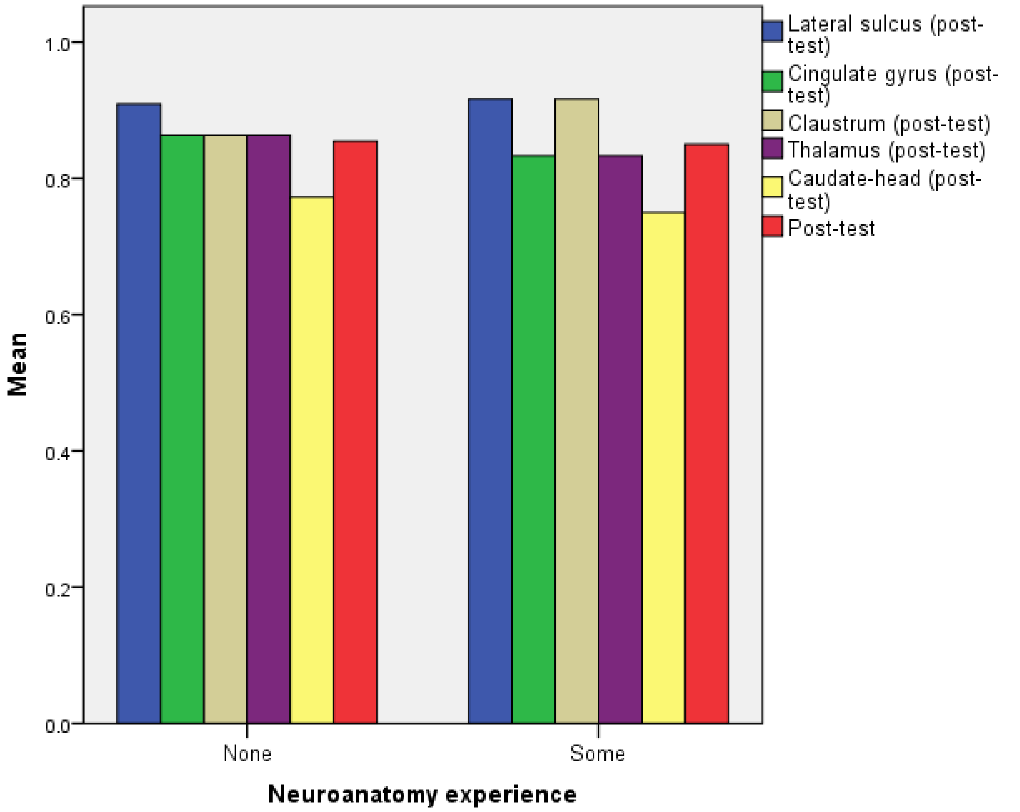


Fig. 19: Differences in the outcome of tests, based on previous experience of neuroanatomy.





On Correlational analysis, the confidence in topics dimensions post-survey and post-test showed strong and significant (τ_b >.32, p<.01) correlations with each other (Table 10). However, the Confidence in topics pre-survey and pre-test showed weak and insignificant correlations with each other and the post-intervention counterparts.

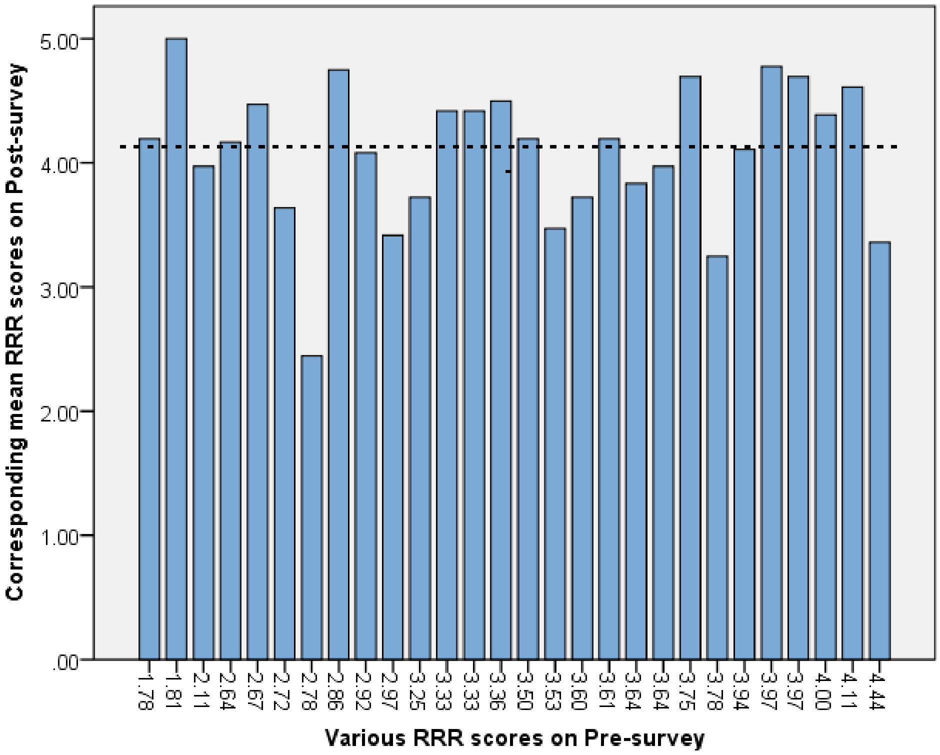
The lack of correlation between the confidence in topics pre- and post-survey, or the pre-and post-test, was further visualized by plotting the average post-intervention scores for each value of the pre-intervention score (Fig. 20), in which the values of post-intervention scores were generally high for all values of the pre-intervention scores (not just high ones).

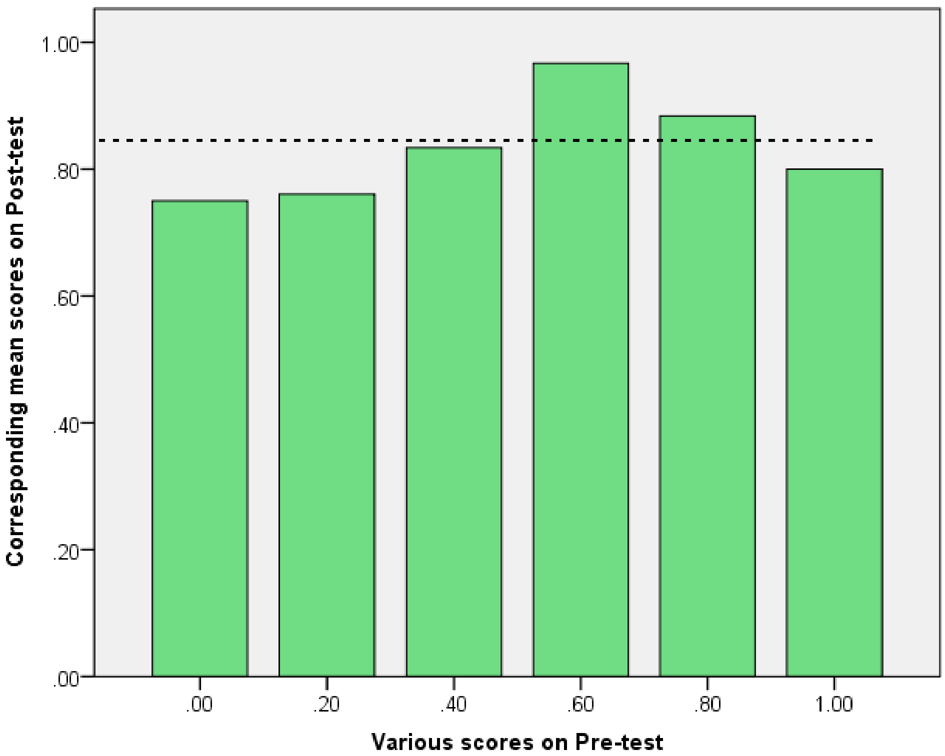
Table 10: Correlational analysis of the Confidence in topics surveys and tests

	Confidence in topics-pre	Pre-test	Confidence in topics-post	Post-test
Confidence in topics-pre	.	N=34, τ_b=.20, p=.140	N=32, τ_b=.19, p=.145	
Pre-test	N=34, τ_b=.20, p=.140	.		N=34, τ_b=.08, p=.581
Confidence in topics-post	N=32, τ_b=.19, p=.145	.	.	N=32, τ_b=.42, p=.005*
Post-test		N=34, τ_b=.08, p=.581	N=32, τ_b=.42, p=.005*	.

* Correlation is significant at the 0.01 level (2-tailed)

Fig. 20: Pre-Post comparison of scores in Confidence in topics surveys and tests. Dotted line represents the average post-intervention score.





SUPERFICIAL VS DEEP STRUCTURES:

The superficial and deep groups in the Confidence in topics surveys and tests were also analyzed individually to find differences concerning gender, academic level, and previous experience in neuroanatomy. Statistical significance was observed only in the post-survey for deep structures, where students with no previous experience in neuroanatomy scored higher than those with some previous experience, with a medium effect size (Fig. 23, Table 11).

Comparing the (pre to post-intervention) increases in the score of superficial structures Vs. deep structures, with depth as the repeated measure, statistical significance was observed only in the tests, where the increase in scores was greater for deeper structures, with medium effect size and no group interactions.

On comparing the superficial Vs. deep structures within the same condition/time, with depth as the repeated measure (within the same survey or test), the mean scores of the superficial structures were greater than the deep structures. However, overall statistical significance was only observed in the post-survey and pre-test, with effect sizes varying from large (for pre-test) to medium (for post-survey), and no group interactions (Fig. 9, Table 11). But in the post-test (although no significance overall), females did have significantly higher (and positive) increase in scores concerning the superficiality of structures, which was counterbalanced by males having a lower and negative increase in scores for superficiality (therefore greater scores for deeper topics, as described in the previous section), with medium effect size (Fig. 21, Table 11).

Finally, on analyzing the increased difficulty for deep structures (the difference in scores of superficial and deep structures within the same condition/time) and comparing their increase in pre- to post-intervention scores, statistical significance was observed only in the tests in which the increased difficulty for deeper structures was more in the pre-test as compared to the post-test, with medium effect size and no group interactions.

Table 11: Outcomes of various comparisons of the mean scores of Superficial Vs. Deep structures

	Pre-intervention		Post-intervention		Comparisons	
	Mean (SD)	Significant Group differences	Mean (SD)	Significant Group differences	Pre-post differences	$\Delta_{Superficial}$-Δ_{Deep} comparisons
Confidence in topics - superficial structures	3.28 (1.07)		4.31 (0.73)		Pre<Post (Z=-4.16, p<.001, N=32, $r_{effect\ size}$=0.74)	$\Delta_{superficial}>\Delta_{Deep}$ (M_1=1.03, M_2=0.83, Z=-0.81, p=.419, N=32, $r_{effect\ size}$=0.14)
Confidence in topics - deep structures	3.24 (1.09)		4.03 (0.78)	None>Some previous experience (M_1=4.27, M_2=3.50, U=50, p=.013, Z=-2.49, N=32, $r_{effect\ size}$=0.44)	Pre<Post (Z=-3.49, p<.001, N=32, $r_{effect\ size}$=0.62)	
Test - superficial structures	0.68 (0.39)		0.88 (0.28)		Pre<Post (Z=-2.35, p=.019, N=34, $r_{effect\ size}$=0.40)	$\Delta_{Superficial}<\Delta_{Deep}$ (M_1=0.20, M_2=0.46, Z=-2.40, p=.017, N=34, $r_{effect\ size}$=0.41)
Test - deep structures	0.35 (0.26)		0.81 (0.33)		Pre<Post (Z=-4.31, p<.001, N=34, $r_{effect\ size}$=0.74)	

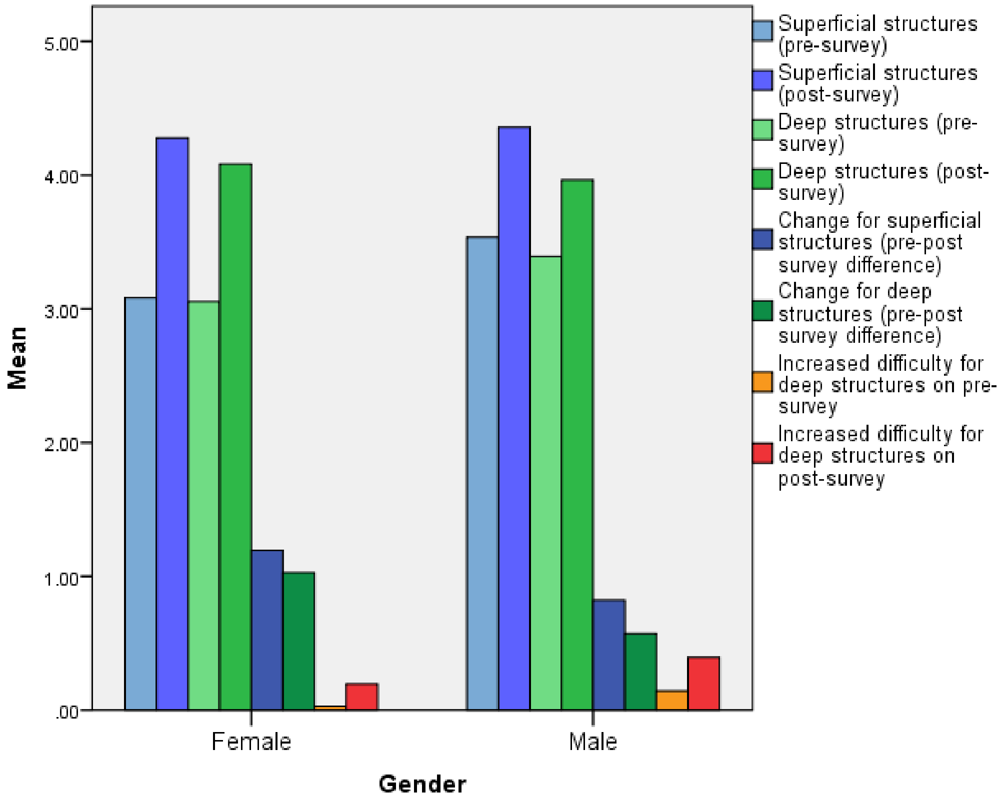
Superficial-Deep Comparisons (increased difficulty for deep structures)	Observed differences	Significant group interactions	Δ_{Pre}-Δ_{Post} comparisons of increased difficulty for deep structures
Confidence in topics Pre-survey	Superficial>Deep (Z=-0.30, p=.762, N=35, $r_{effect\ size}$=0.05)		$\Delta_{Pre}<\Delta_{Post}$ (M_1=0.04, M_2=0.28, Z=-0.81, p=.419, N=32, $r_{effect\ size}$=0.14)
Confidence in topics Post-survey	Superficial>Deep (Z=-2.21, p=.027, N=32, $r_{effect\ size}$=0.39)		
Pre-test	Superficial>Deep (Z=-3.87, p<.001, N=34, $r_{effect\ size}$=0.66)		$\Delta_{pre}>\Delta_{Post}$ (M_1=0.32, M_2=0.07, Z=-2.40, p=.017, N=34, $r_{effect\ size}$=0.41)
Post-test	Superficial>Deep (Z=-1.29, p=.197, N=34, $r_{effect\ size}$=0.22)	$\Delta_{Male}<\Delta_{Female}$ (M_1=-0.07, M_2=0.18, U=88, p=.026, Z=-2.23, N=34, $r_{effect\ size}$=0.38)	

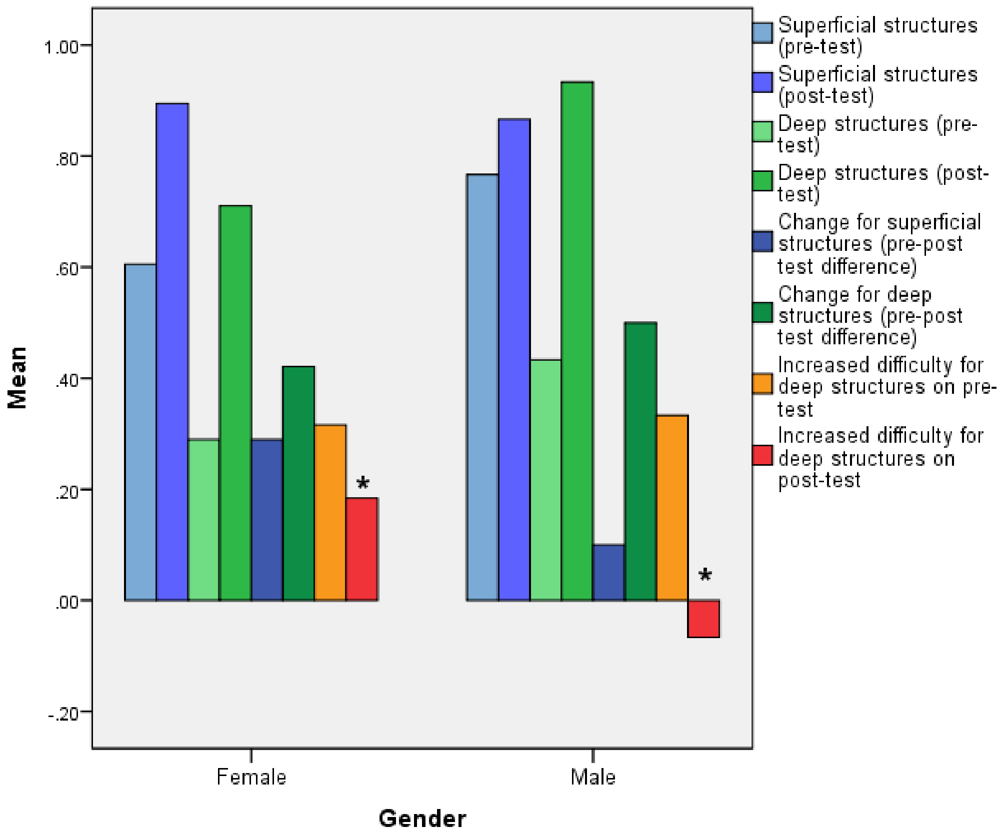
The mean scores of males were slightly higher than females for most of the superficial and deep structures on the Confidence in topics surveys (except deep structures on the post-survey). In contrast, females had greater pre to post-increases for these subgroups (Fig. 21), but no statistically significant difference was observed for any of these observations (Table 11).

Likewise, for the tests, the mean scores of males were slightly higher than females for most subgroups (except superficial structures on the post-test). However, a statistically significant difference was observed only in the item describing increased difficulty for deep structures on the post-test, where females had positively increased difficulty, in contrast, males had negatively increased (decreased) difficulty for deep structures, with medium effect size.

Fig. 21: Gender differences in the comparisons of the mean scores of Superficial Vs. Deep structures.

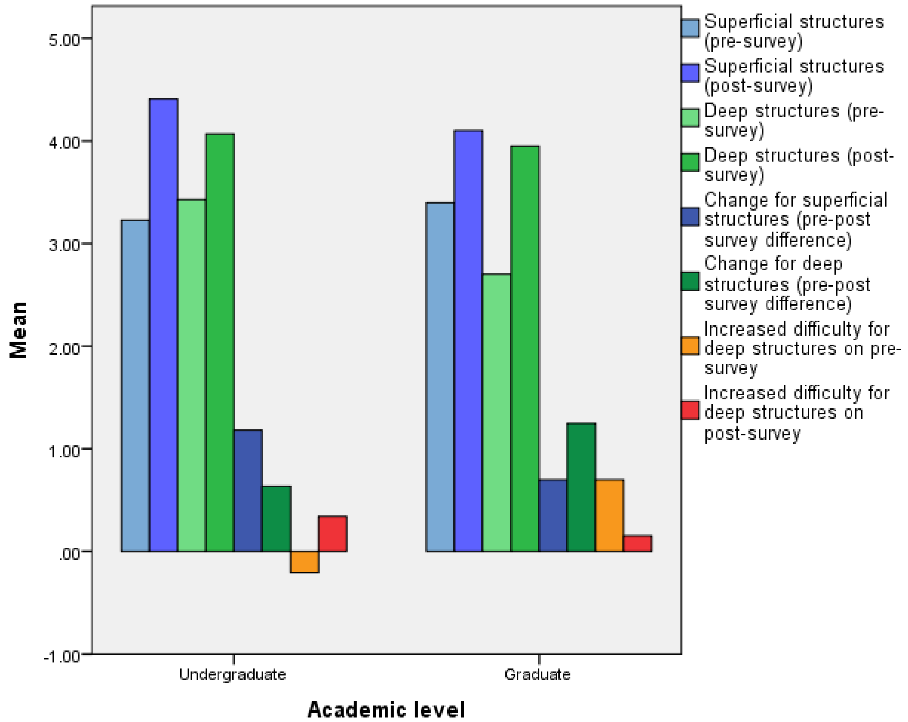
* Item where a statistically significant gender difference was observed

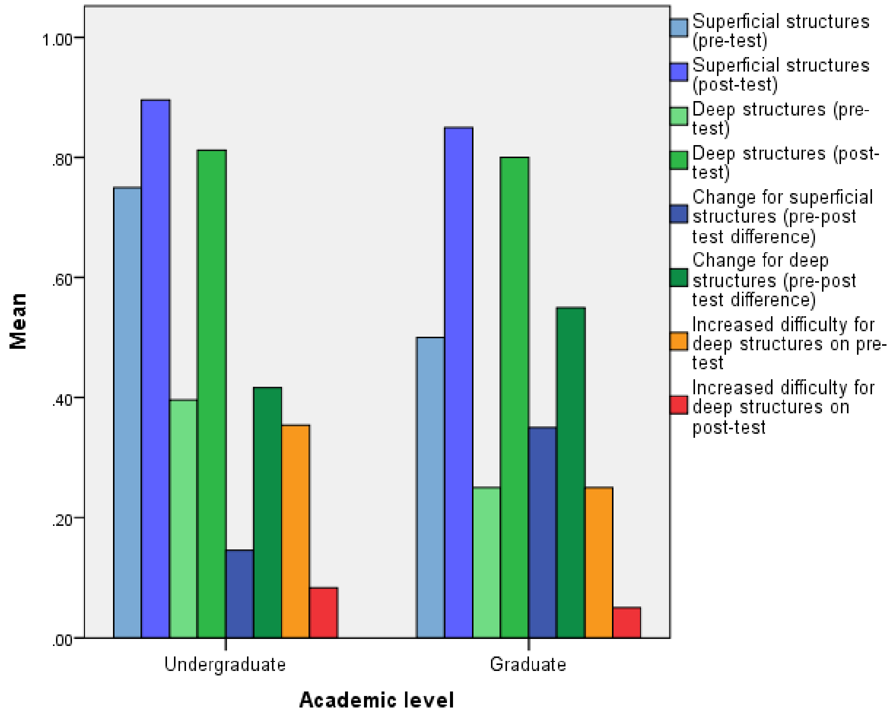




The mean scores of undergraduate students were slightly higher than those of graduate students for most of the items for the superficial and deep structures on the Confidence in topics surveys and tests (Fig. 22); however, no statistically significant difference was observed (Table 11).

Fig. 22: Academic level differences in the comparisons of the mean scores of Superficial Vs. Deep structures

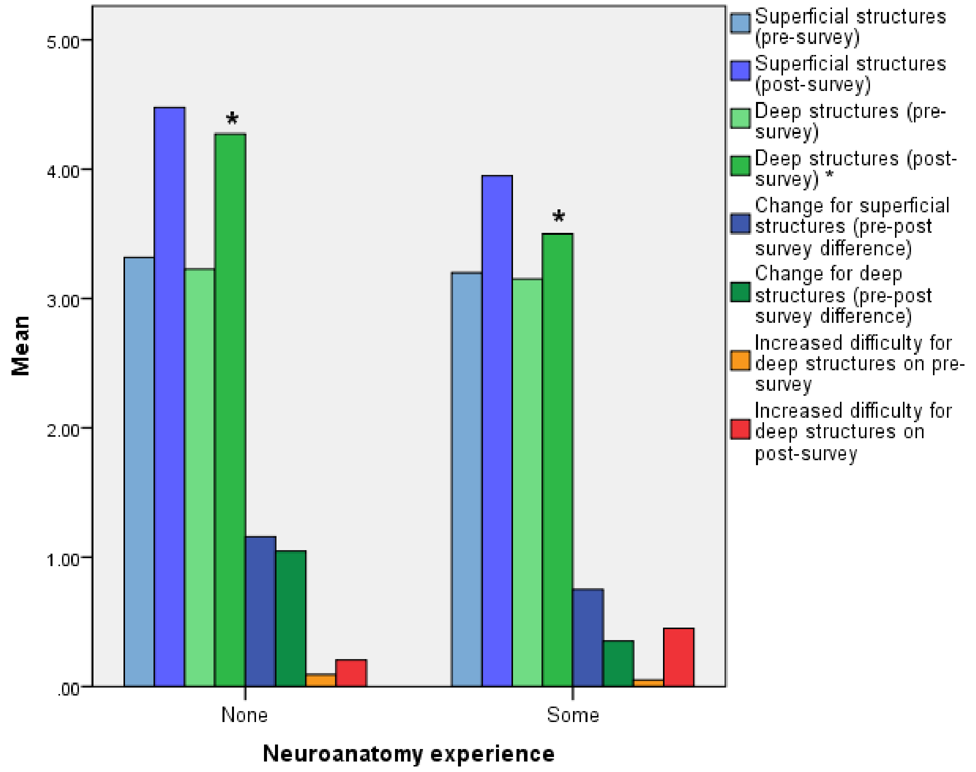


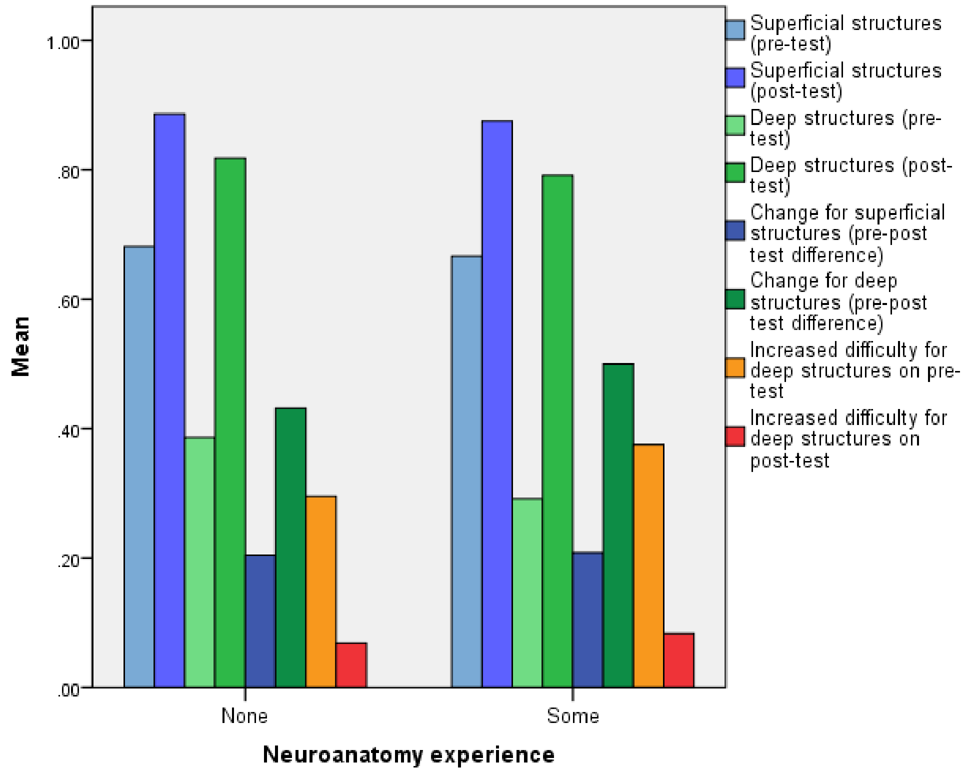


The mean scores of students with no previous experience in neuroanatomy were slightly higher than students with some previous experience for the superficial and deep structures on the Confidence in topics post-survey (but similar scores for the subgroups in pre-survey pre-and post-test, Fig. 23), and statistical significance was observed in the item deep structures on post-survey, with medium effect size (Table 11).

Fig. 23: Differences based on previous experience of neuroanatomy in the outcome of comparisons of the mean scores of Superficial Vs. Deep structures.

* Item where a statistically significant difference based on previous experience of neuroanatomy was observed





QUALITATIIVE ANALYSIS:

While students stated that they appreciated different themes of the program, there were still areas that they disliked and offered improvements to the program. As aforementioned in methods, the qualitative questions were all broken down into codes and then simplified to create a concise code. Reviewing the answers for what the students liked about the program, there were 11 specific codes; then, they were grouped into 4 (Technology, Features, Content, and Realism) concise codes. While reviewing the responses to what the students disliked about the software, there were 6 initial codes and then 5 (Tech, Features, Content, Realism, and Time). The responses of students' input to what could be improved were a total of 7 initial codes, reduced to the same 5 codes as earlier (Tech, Features, Content, Realism, and Time), detailed in Table 12.

What they liked

When asked what they liked, eight phrases expressed likeness of the access and ease of the software (technology), 47 favored the features, 13 about the content, and three suggested it was an improvement over 2D (realism). The students stated that they "could navigate [the app] as quick or slowly as [they] wanted," and the ability to move view/rotating structures "made it easier" for them. The ability to "view the anatomy from any angle" allowed for more than a "superficial understanding" and offered a "more holistic view" in contrast with "just memorizing structures." The amount of detail that was offered by the software provided students the "ability to dissect whenever for better understanding" and "how organized everything was" "organized into subsets like the brainstem, etc." in "different perspectives available" so that they could see "how structures related [sic] to each other and seeing them all together." Students highlighted that the options to "add/subtract structures" and "go deeper in the brain as desired" were significant in their understanding of the content. They also liked the ability to "pick out different sections to go more in-depth on them specifically." The slicing feature and sectional anatomy in Anatomy learning allowed students to "visualize how the different structures change throughout the brain."

Furthermore, students stated that the ability to "dissect/fade through structures" and "transparency option" allowed for a deeper understanding of the material. Students also mentioned the benefits of "color coding" (which made boundaries clearer in the

'AnatomyLearning' software), the "search bar," and the "quiz feature." Many students stated that the software helped achieve a greater understanding of the complex brain structure and "spatial awareness of where all structures are" and the ability to "mentally map out where certain structures were in the brain by choosing specific key features (ex. central gyrus, claustrum, etc.)" which helped in "retention and genuine learning of the structures ."They also admired "external brain morphology and functional areas in the 'Visible Body' software."

What they disliked

When asked what students disliked, 15 stated they disliked the technological aspect, 13 disliked certain features, 5 disliked the tags (content), and 8 disliked the lack of tactile and unrealistic pictures (realism). One stated dislike of the short time/duration of the course. The software (especially the 'Visible Body' software) took a "toll" on students' computers and had some glitches (especially the 'AnatomyLearning' software), which did not allow them to use it freely. One student stated that the software was "incompatible with [multiple] computers" and had to "partner up with another student." The structure of the 3D software (especially the Visible body software) was sometimes unclear; thus, it was difficult to "navigate through" the cross-sections, and they had difficulties (sudden unintentional resets to view/structures, imperfect isolation, unintentional deletions) trying to "search for features" and "find some structures ."Also, it was "hard to see certain structures that were hidden behind others, such as the anterior side of the cerebellum" (hidden behind the brainstem), and the Anatomy learning software did not have the "descriptions of the structures or an isolation feature, like that of Visible Body. "Regarding tags, students stated that "some parts were not labeled" in particular/all views and "didn't like how the names were different across the board from the checklist." Many students highlighted that the software created a "very simplistic" and unrealistic experience for them and wished they had a "hands-on" experience with an actual brain; however, due to covid, this was not possible. It was also stated that they disliked the time frame of the program.

Scope of improvement

When students were asked to offer suggestions on how to improve the course, 12 students stated that the software/technology should be improved, 17 stated that the features should be improved, 11 stated the content should be improved, 4 expressed the importance of accuracy and realism, and lastly four that there should be more time. At the same time, only two undergraduates expressed improvement in the navigation control of the software and assembly control of the structures, every graduate student communicated the importance of improving the navigation system (especially in the 'AnatomyLearning' software, which had some glitches and imperfect controls). Also, they stated a need for "deeper training of how to use Visual Body" as they found it "a bit confusing." Students suggested a "search bar" to facilitate their search in the 'AnatomyLearning' software and more improvements to the search and isolate feature in the 'Visible Body' software. Moreover, a student compared the program to "Show me the brain" as they stated it "was significantly easier to control and navigate," which our software did not do well. A student suggested that the course "stick with only one application in the future." Moreover, students want the opportunity to "take apart and put together" more efficiently while having the ability to add tags/labels to "certain structure[s]." They mentioned that "highlighting what was important to study" or "readily available set of digital flashcards." Another student suggested that, "printable fill-in the black 'coloring book' pages to help have a truly hands-on aspect to the class." It was mentioned that, "some software/websites have clinical images to accompany programmed neuroanatomy," which can help the students who mentioned their desire for a more accurate brain representation. Some students wanted the diagrams to "look more realistic" and that the "brain [be] colorized as an actual brain is" or a different greyscale version like we provided for testing. Moreover, only undergraduate students highlighted that they wanted more time in the course as they felt the material was dense to learn in the given period.

Table 12: Major themes identified from open responses regarding likes, dislikes and scope of improvement.

Themes	What they liked	What they disliked	What needs improvements
Technology	Freely accessible	Incompatible on some devices	One unified app for all use
	Easy to use	Heavy and memory intensive software	Remove glitches
	Beautiful rendering	Glitches and lags	
Features	Organization into subsets and perspectives	Search due to unclear structure and user interface	Deeper training to use software
	Navigation speed	Difficulty to navigate cross sections	Better navigation controls
	Rotating view to look at different angles	Unintentional rotations, resets and deletions	Better rotation control
	Ability to zoom in or out		Better zoom and scrolling control
	Ability to fade thru (transparency) or dissect		Better dissection tools/control
	Ability to add or subtract structures and go deeper		Better assembly and disassembly controls, 'take apart and put together' like real models
	Specific isolation of structures	Imperfect isolation and hindered view	Better isolation of structures
	Moving thru different cross sections by slicing feature		Ability to add tags/labels, comments and highlighting text
	Delineation by color coding and contrast		Additional greyscale view as in exam/real life
	Search and quiz feature	Difficulty to search structures	Improvement to search bar

Content	Complex brain structure and fine details	No detailed descriptive text	Clinical images, supplementary material like flashcards and coloring book
	Holistic view and relations	Tags: some completely absent, some absent in specific views, some different from the checklist provided	More tags and labels
	Spatial awareness and mental mapping		
	Retention and genuine learning		
Realism	A definite improvement over 2D	Simplistic, cartoonish and unrealistic	More accurate structure and realistic colors in sections and views
		Lack of tactile/hands-on experience	
Time		Short course duration	More time to study

DISCUSSION

The study introduces a new instrument for self-reporting student data on the effectiveness of learning interventions (the Reaction-Relevance-Result survey) and adds value to the existing evidence for the effectiveness of 3D software in neuroanatomy education. The addition of 3D visualization may significantly improve and upgrade the existing teaching techniques, especially in a diverse group of students with various spatial visualization needs.

DEMOGRAPHICS

The student population in STEM fields generally consists of mostly male students (over 80%; NSF, 2017). Nevertheless, female participation in science and technology courses has been gradually increasing over time and surpassing the male population in many areas, including biological sciences in general (NSF, 2017) and neuroscience (graduates overall and specifically at bachelor's and master's level but not doctoral yet; Ramos, 2017). This was also true for the undergraduate population in our study (64% females) and overall (57% females) as compared to graduate students, which still consisted of a male majority (60% males); however, there was no observed statistically significant relationship between gender and academic level.

Apart from these gender differences, an increase in the number of students enrolling in neuroscience programs and of new such programs has led to an increase in the proportion of neuroscience students among other life sciences majors, especially at the undergraduate level where neurosciences graduates outnumber most all other majors (Ramos et al., 2016a; Ramos, 2017). These trends predict that the proportion of female graduates in neuroscience and life sciences, in general, will keep increasing and dominating over time. These findings may be helpful to faculty and administration of neuroscience programs to prepare for and address the cognitive needs of a demographically changing population.

USAGE

Mind wandering is a cognitive problem when students study for long periods; it occurs with higher levels in males than females in a neurotypical population or equal levels in the population with ADHD (Mowlem et al., 2019). Leaders of online education like Salman Khan (founder of Khan Academy supported by Bill Gates) and Daphne Koller (co-founder of Coursera at Stanford University) are in favor of shorter 10-minute online lectures as they are concerned about the inability of students to remain attentive for extended periods (Khan, 2012; Koller, 2011; Szpunar et al., 2013). Studies have shown that mind wandering is a frequent problem that occurs regardless of lecture duration and is more probable when online education takes place in a personal space like a dorm or home, which is full of distractions (Risko et al., 2012, 2013; Szpunar et al., 2013). The problem of mind wandering can be minimized by using tactics like interpolated testing, note-taking, and visuals which are more graphic, dynamic, and complex leading to increased engagement (Szpunar et al., 2013).

Regarding usage, our students studied the software for up to an hour, similar to traditional teaching methods of 1-hr lectures. They also reported using it somewhere between 5-10 times during the short duration of the study of fewer than two weeks. There are few studies in neuroscience or anatomy education regarding time duration and attention span during the usage of 3D apps and particularly in a student's personal space. In most studies, usage is limited by the experimenter and is conducted in an institutional setting, in which session duration lasts from 20-30 minutes only (Hu et al., 2016; Fleagle et al., 2018; Donnelly et al., 2009; Jamil et al., 2019). Students in our study reported software usage for lengthy and frequent durations, which may be a result of enhanced engagement by 3D content as compared to traditional 2D resources (which are dull and cognitively overloading, hence uncomfortable) due to the better visuals and colorful graphics of the 3D apps, which were also praised in the qualitative feedback survey. This feedback suggests the importance of using 3D software in neuroanatomy education to enhance learning in students, a population riddled with increasing mind wandering and decreasing attention span.

Learning with any 3D software can be overwhelming to many students as they are using it for the first time, making it challenging to build concepts while constantly moving them in a 3D space and learning these new controls and features; however, their overall perception is generally welcoming (Jamil et al., 2019). Although younger generations, like millennials and Generation Z, are more technologically sound, it will still take some time for broader adoption of this 3D software as it is a gradual process (Hope, 2018). Their introduction in the classroom should be gradual, aiming for learning benefits rather than just improving the (technology-mediated) visualization, done either through self-guided assignments or small teaching sessions, also providing mental rotation training (especially for populations with lower spatial ability, which has a constant and residual effect while using 3D over 2D resources, where 3D applications significantly benefit all students but trained group significantly outperforms the untrained group before and even after the 3D intervention), and most importantly giving them enough time, to help students adapt on the new pedagogical environment to allow proper learning to take place (Jamil et al., 2019; Roach et al., 2019).

Students in our study reported little ease in learning new software, little history with the software used in this study, and little agreement with the software's well-organization (which was also indicated in the qualitative feedback), which makes the novelty issue a significant concern regarding its usage. However, students highly recommended using such software in the future, which suggests that they are willing to overcome the barriers of novelty and appreciate the gains of its usage.

Gender differences were observed in comfortability with software usage, especially regarding ease of virtual dissection on the software, where males experienced slight ease. In contrast, females experienced little unease doing virtual dissections on the software, with a medium effect size of this gender difference. Previous experience with neuroanatomy or academic level did not affect the ease of virtual dissection or the comfortability of software usage in general reported by students. This implies that there may be some entry barrier for females or other populations with low spatial abilities to use the 3D software and utilize the full benefits of the 3D visualizations.

THE REACTION-RELEVANCE-RESULT SURVEY

The 13-item RRR survey developed based on traditional theories of learning like the Kirkpatrick and ARCS models, was tested on online learners during the COVID-19 epidemic. The RRR survey measures self-perceived motivation and learning benefits of study intervention (in-person or online/virtual) by measuring three components (reaction, relevance, and result). The instrument can be used to compare the sub-scores of the three components and generate an overall RRR score representing the study intervention's learning effects.

The tool was developed keeping in mind the four domains of the ARCS model and the four levels of the Kirkpatrick model, converging into the three components of RRR (reaction, relevance, and result). These three components were predetermined instead of performing dimensional reduction of a broader questionnaire as students confuse the affective and cognitive aspects of learning, especially at the beginning of courses, leading to cross-loadings and inter-correlations among items, which may lead to complications in the analysis (Rovai et al., 2009). PCA confirmed the three components as unidimensional. Most factor loadings were good, and only four were just satisfactory. The absence of subscales due to the one-dimensionality of these components required no rotation, generated no cross-loadings, and resulted in an instrument accounting for all the variance in respective components. The RRR survey showed high levels of internal consistency and validity, proving to be a valid measure of these components. It can help to differentiate between the affective and cognitive benefits and compare them among different types of courses. Its use in other related disciplines or a more significant curriculum would further validate the instrument and provide the opportunity for improvement.

The RRR survey can be an essential tool for education research, mainly because there is much debate on the efficiency of various teaching modalities, whether traditional or modern (or even hybrid). The introduction of online and hybrid courses during the COVID-19 period resulted in many researchers and instructors testing these courses for their learning efficacy. Authors have pointed out (Rovai et al., 2009; Tallent-Runnels et al., 2006) that educational research needs more systematic studies for online learning to measure the effectiveness of these courses, especially in domains of academic success and thinking skills rather than student preferences or faculty satisfaction, and to measure variables based on learning theories and models of teaching, they also noted that many existing studies use only single-item questions for a particular measure. Our RRR survey addresses all these issues and can be used for research across multiple disciplines related to anatomical sciences, giving a more comprehensive evaluation score across various aspects of motivation rather than just analyzing a simple score of a final exam like many studies reports. Changing the items on the RRR survey components, adding more components, or combining the survey with pre-validated instruments for a more in-depth investigation into specific aspects of learning or environments and practices, can lead to developing a better tool to measure the learning effectiveness of such courses. Since it was developed and tested on an online course in neuroanatomy, it is more suitable for such purposes and can be used for traditional offline or blended courses (given the rise in the blended approach due to technological advances and changing needs of students). It can also be used for other student populations in the medical, psychological, and biological sciences.

All the different technologies and approaches have their benefits and suit different needs. While the traditional approaches might be more time-tested and favored by the status quo, newer technologies have created demand in young learners who are more comfortable with changing trends. The RRR survey can help researchers understand the difference in perception and benefits between these approaches and help them choose a more practical approach to learning.

OUTCOMES OF THE RRR SURVEY

As aforementioned, the post-survey scores for the 3D intervention were significantly higher (and mostly with large effect sizes) for the RRR survey overall and its components and most items than the pre-survey, which was given after the 2D intervention, suggesting that the 3D intervention was effective and more effective than the 2D resource used for neuroanatomy learning. Regarding the Reaction component of the RRR survey, the students' feedback showed a very positive reaction to the introduction of the 3D intervention as perceived by students. While they may or may not show more interest in studying the 3D resource, they felt it is more enjoyable and grabs more attention (demonstrated by the large effect size observed in the analysis) compared to 2D resources. Regarding the Relevance component of the RRR survey, the feedback showed that the students perceived the 3D intervention as more relevant to their educational needs than the 2D resource. They acknowledged that the 3D intervention is more visually appealing and helps better in long-term retention than the 2D resource and that they found the 3D resource to be more useful in learning, more helpful in visualization of simple and complex structures, and more beneficial for spatial understanding owing to the significant effect sizes in these comparisons.

Regarding the Result component of the RRR survey, the feedback showed that the students perceived the 3D intervention gave better learning results than the 2D resource. While they may or may not be more satisfied with studying the 3D resource, they felt it is more helpful in changing their attitude towards neuroanatomy and preparing them for the exams, compared to the 2D resources. It can be concluded that not only the RRR survey overall but every component (the reaction to the 3D intervention, self-perception of its relevance, and results of learning) showed that self-reported outcomes of the 3D intervention were all very positive and were well-appreciated by the students to be more helpful than the 2D resource.

There were some gender differences observed in the analysis of the RRR surveys, which were found to be interesting. In the pre-survey for the 2D resource, male students had significantly higher scores than female students for the RRR survey overall, all its individual components, and many items, with medium effect sizes. Male students reported giving more attention, having more interest, better visualizing the simple and complex structures, and having a more positive attitude after the 2D resource than female students. However, no such gender differences were observed in the RRR post-survey after the 3D intervention. This suggests that the self-perceived effectiveness of 2D resources in male students is more than in female students. Also, although there may be some usage-related entry barriers in 3D resources for female students, as discussed earlier, they perceive the 3D resources to be as effective as the male students. This suggests that the 3D intervention brings some gender equality in the self-perceived effectiveness of the intervention once the entry barriers are crossed.

As aforementioned, younger generations tend to be more technologically sound, yet no group differences were observed regarding students' academic level (undergraduate or graduate level). This suggests that 3D intervention was effective enough to help students overcome any age-related entry barrier and learning difficulty.

Nevertheless, some interesting observations were found concerning their previous neuroanatomy experience. While the pre-survey showed no group differences regarding the academic level or previous experience, the 3D intervention led to group differences in the post-survey, where students with no previous experience seemed to consider the 3D intervention more effective compared to students with some previous experience. Students with no previous experience had significantly higher scores for the RRR survey, two individual components, and some items with medium effect sizes. They had a better reaction to the 3D intervention in which they reported paying more attention as they enjoyed it more, which was also demonstrated by one (and the only) significant pre-post group difference in the RRR survey, which confirmed their increased attention to the 3D intervention over the 2D resource as compared with students with some previous experience. Students with no experience also self-perceived better results of the 3D intervention and a more positive change in attitude towards neuroanatomy, but not regarding the relevance of the 3D intervention where experience seemed to play no role overall or in any items except for visualization of simple structures (and not complex structures) when compared with students with some previous experience.

This suggests that for inexperienced students, the perception of their reaction and results of the 3D intervention was higher than that of students with some prior experience. This may be due to the use of real cadavers or a better learning environment experienced by the students with prior experience, leading to some bias and their decreased self-perception of reaction and results (but only relevant for visualization of only simple structures). So previous experience may hinder acceptance of newer technologies, especially when they tend to replace traditional methods like cadaver-based learning but remain equally relevant and may be better for complex structures, as discussed later. The evolution of educational technologies and the introduction of new learners who have not experienced the old learning methods will decrease this resistance and facilitate the inevitable transition to a more technologically enhanced form of neuroanatomy learning.

THE CONFIDENCE IN TOPICS SURVEY AND TEST

The 5-item confidence in topics survey and test, an instrument that was developed based on five brain structures commonly taught and tested in neuroanatomy exams, can be used as a tool to measure the self-perceived confidence of students in their neuroanatomy knowledge and their performance in exams, for analyzing the learning benefits of the study intervention. The tool contained two superficial, one intermediate, and two deep structures. Factor analysis of the survey component confirmed the instrument as unidimensional. All the factor loadings were good, and only one was just satisfactory. The survey showed high internal consistency and external validity levels, proving to be a valid measure of student's confidence in their neuroanatomy knowledge. Interventions that increase confidence may lead to better results. The tool can be a quick diagnostic tool to assess students' confidence and knowledge. It serves as a tool neither too simplistic as a one-question query nor too complicated as some multiple paged tests. Changing or adding of few more questions to the 5-item tool and its use in other related courses and disciplines will provide the opportunity for its improvement and further validate this instrument in the future.

OUTCOMES OF THE CONFIDENCE IN TOPICS SURVEY AND TESTS

As shown earlier, the post-survey and test scores for the 3D intervention were significantly higher (and mostly with large effect sizes) than the pre-survey, which was given after the 2D intervention, suggesting again that the 3D intervention was very effective and more effective than the 2D resource used for neuroanatomy learning.

The lack of correlation between the pre-survey and pre-test suggested that students' confidence levels in their neuroanatomy knowledge did not reflect their actual knowledge after the 2D intervention, suggesting a lack of clarity and insight after using 2D resources. On the other hand, the post-survey and post-test showed a strong correlation. Their confidence levels reflected their knowledge after the 3D intervention, suggesting that the 3D approach brings more clarity and insight.

Also, the lack of correlation between pre-intervention scores with post-intervention counterparts and equitable distribution of post-intervention scores suggested the improvement in scores is due to the 3D intervention among all students irrespective of their knowledge based on the 2D resources. The 3D intervention benefitted all and not just a few students, which is very desired in society.

Gender differences that were observed in the RRR pre-survey were absent in the case of the confidence in topics pre-survey and pre-test, suggesting that although female students underestimated the effectiveness of the 2D resources as compared to male students, it did not lead to a difference in their confidence levels in neuroanatomy knowledge or performance in tests. Hence, it only affected their likeability but not their confidence or performance, which may be due to extra resolve or effort put in by the female students (which could not be confirmed as the usage statistics needed to be more reliable).

Regarding the post-survey and test, group differences due to previous experience in neuroanatomy were again seen, but only in the survey component (with medium effect sizes) and not the post-test, suggesting that since the effectiveness of the 3D intervention was underestimated by students with some previous experience in neuroanatomy as compared to students with no previous experience, it also led to a difference in their confidence levels in neuroanatomy knowledge. However, their performance in tests did not reflect this. Hence, it affected their likeability and confidence but not their performance, which may result from their underestimating their actual knowledge or their seriousness for exams.

Apart from these two group differences encountered earlier, a new situation was observed. Although there were no gender differences observed overall in the post-test or pre-post increase of scores, there was a significant difference in the two situations. First, for the cingulate gyrus (a superficial structure), the pre-post increase was more in females than males, and secondly, for the thalamus (a deep structure), males had a higher score than females in the post-test. This may suggest that females may have benefitted more from the change to 3D intervention over 2D resources, but probably for some superficial structures only (not overall). In contrast, male students had more knowledge than females in the case of some deep structures but only after the 3D intervention. This may suggest a gender differential role in learning benefits achieved at the end despite having no gender effect overall in any of the confidence in topics surveys or tests (more superficial structure benefits for females and more deep structure knowledge at the end for males). This idea was further explored and cleared in the next section.

OUTCOMES OF SUPERFICIAL VS DEEP STRUCTURES

As discussed earlier, the confidence in topics surveys and tests consisted of 5 items, out of which two were superficial structures (Lateral Sulcus and Cingulate gyrus), one was an intermediate structure (Claustrum), and two were deep structures (Thalamus and Caudate-head). For simplicity of the analysis, only the superficial and deep structures were compared on their pre- and post-intervention scores on surveys and tests. As expected, and just like the confidence in topics overall, the superficial and deep structures both had significantly higher scores for the post-intervention counterparts, suggesting that the 3D intervention was effective and more effective than the 2D resources for both the superficial and deep structures.

The post-survey for deep structures had significantly higher scores for students with no experience in neuroanatomy than students with some experience. However, this difference was not observed for the post-test or any pre-intervention counterparts, just as we encountered confidence in topics overall. This suggests again that since students with some previous experience in neuroanatomy underestimated the effectiveness of the 3D intervention, it also reduced their confidence levels in neuroanatomy knowledge of deep structures. However, these did not affect their performance in tests. So, for the deep structures, these students are less confident than the inexperienced students after the 3D intervention but still perform equally. However, no significant reduction in either confidence or performance was observed for superficial structures. However, this lack of trust in 3D intervention for deep structures by some students was also present in confidence in topics. It did not necessarily mean that they trusted it any less or more for superficial structures. Students with some prior experience have some trust issues with the 3D intervention, overall and especially in deeper structures, although they give the same results as inexperienced students.

While analyzing the pre-post intervention increases and then comparing them for superficial and deep structures, the increase for deep structures was significantly higher than superficial in the case of tests (but no significant difference was observed in the case of surveys). This means that while the 3D intervention may or may not lead to an increase in confidence levels in the case of deep structures, it certainly did lead to better enhancement of performance on the tests for deeper structures. This means that the 3D intervention enhanced performance in both superficial and deep structures but proved more beneficial for deep structures.

While comparing the scores of superficial and deep structures within the exact condition/time, superficial structures had a significantly higher score than deep structures in the post-survey and pre-test, and higher but not significantly for the pre-survey or post-test. This suggests that scores on superficial structures are generally higher than on deep structures at any given time. While students may or may not have more confidence in their knowledge of superficial structures after the 2D intervention, they certainly did have more confidence in superficial structures than deep structures after the 3D intervention. However, their performance told a different story, according to which they performed better on superficial structures on 2D intervention but may differ for superficial and deep structures after the 3D intervention.

This means that while there may or may not be a difference in confidence for superficial and deep structures after studying the 2D resources, it appeared quite distinctively after studying the 3D resource, suggesting that the 3D intervention may have increased inequality in confidence among the superficial and deep topics. It may also be due to development of a better insight into an already existing inequality of understanding of superficial and deep topics by the students after the 3D intervention, rather than the 3D intervention creating these inequalities itself. Conversely, as the test results showed that students distinctively performed better for superficial structures after studying the 2D resources (but may or may not do so after the 3D intervention), it suggested that in actuality the 3D intervention may have removed the difference in difficulty among superficial and deep topics. So, while improving the insight into or creating the inequalities of understanding between superficial and deep topics, learning on 3D resources may have helped remove these inequalities in performance.

We confirmed this by analyzing the pre-post comparison of the difference in performance on the superficial and deep topics. A significant pre-post difference in the increased difficulty for deeper topics was observed only for the tests and not for the surveys. According to their actual performance on tests, students experienced a significant amount of increased difficulty for the deeper topics by the 2D intervention compared to the 3D intervention, confirming that the 3D intervention brought some equality in differential difficulties of superficial and deep topics and brought some equality in their scores (only for the performance in tests but not the self-reported confidence where it may conversely bring some more insight into the inequalities).

However, this equality of superficial and deep topic test scores after the 3D intervention may not be accurate as some group interactions were present concerning gender (although there was no significance overall) in the post-test. After the 3D intervention, females had a significantly greater score difference between the superficial and deep structures, which was counterbalanced by males having a lesser difference. None of these gender differences in superficiality were observed in the pre-test or any of the surveys. This means that although the male and female students may or may not feel different levels of confidence for the superficiality of structures after studying on either of the intervention and may or may not perform equally on superficial and deep topics after the 2D intervention, they did differ gender-wise in the superficiality-differential performance on these topics after the 3D intervention where females performed more differentially on the superficial and deep topics. In contrast, males performed less differentially on superficial and deep topics. This suggests that according to their performance on the exam after the 3D intervention, the female students experienced the deeper structures to be much more complicated than superficial ones, as compared to male students.

In contrast, the performance of male students indicated these groups to be less different in their difficulties, although the confidence levels may or may not reflect this story. This means that performance-wise after the 3D intervention, the increased difficulty for the increased depth of structures was lesser in the case of the male students than in female students. As discussed earlier, some studies have shown lower spatial ability among females as compared to males, and spatial ability has been shown to have a constant and residual effect on learning using 2D and 3D resources. The observed gender differences may result from males having a higher spatial ability to begin with, or due to their self-reported experience of more ease in software usage, leading to a better understanding of the deeper and more complicated structures.

Although various 3D software has been increasingly used for neuroanatomy education due to their better visualization effects, there is not enough evidence regarding their role in enhancing learning among a diverse population with varying spatial abilities, as spatial ability is often underrated, unsupported and ignored (Jamil et al., 2019; National Research Council, 2006). Spatial ability has been proven as a reliable predictor of performance in human anatomy and other medical sciences, but still, it has been sidelined while deciding medical school admissions and suggested for use only to identify learners requiring additional interventions (Lufler et al., 2012; Kopp & Rathmell, 2015). Humans have inherently different spatial abilities, where the mental rotation (MR) ability of males is generally higher than females. Males outperform females in MR training workshops, take less time to complete spatial tasks, and achieve higher scores (Jamil et al., 2019; National Research Council, 2006). However, the use of 3D software often leads to homogenous learning in both genders, and the final performance is unaffected by gender; hence its usage is beneficial, meaningful, and essential when working with diverse populations (Jamil et al., 2019; Khot et al., 2013; Lim et al., 2016). Time, effort, practice, and use of 3D software itself may increase the spatial ability of individuals, which may be the reason behind this homogenous learning (Newcombe, 2010).

At first, it appeared that the effects of spatial ability on 3D learning did not appear overall or gender-wise, as both genders performed similarly overall after the 3D intervention. However, these residual effects on gender seemed to have seeped deep down in the deep structures using 3D software for enhanced learning. 3D software is one of the interventions that can bring down these effects of noteworthy ability differences between genders over time, as it reduces the cognitive load, especially demands of spatial ability, and helps in its training. However, the persistence of residual effects and their levels may require a better interface or software or other types of training for enhancing spatial abilities so that these gender differences in performance on the tests can be neutralized. This is also required for other population groups with differences in spatial abilities, which needs further research.

QUALITATIVE FEEDBACK

The students appreciated the beautiful renders and the benefits of accessibility that the 3D software brings to them. However, this software still needs a lot of work and improvement as it can sometimes glitch and is not optimized for the hardware the students possess. There is so much independent development of software by different companies and universities that it is difficult to get one software with all the good features available. The evolution and convergence of this software in the future, as they learn from each other, will lead to better solutions.

The ease of navigation in these software and their features are essential to increase student interest. While many students like this software's ease, controls, appearance, presentation, and structure, some issues still need to be solved in searching for and controlling these structures or features. The refinement and development of better features and software reorganization will resolve these issues as they evolve. Adding capabilities that allow the user to add content or change color schemes will lead to better personalization. Also, better and dedicated training by the academic institution on this software will lead to decreased mental resistance and better engagement.

While the available software has acceptable content and explains the complexities of the brain with great detail, along with providing a holistic approach and enhancing spatial visualization of complex relations leading to better learning and retention, there can still be some inconsistencies with labels or tags using different or old anatomical terminology and deficiencies in details and description of these labeled structures. Using standard anatomical terminology and adding more clinical material to this software in the form of detailed boxes, popups, or supplementary material like flashcards will significantly enhance the learning experience.

The 3D software possesses the advantage of being far better than 2D resources for visualization and learning. However, they still need more realism and the feel of a hands-on experience on an actual brain, as the computer-modeled images can sometimes feel too simple and cartoonish. Better segmentation of 3D structures and the data on which these structures are modeled, especially thru advances in higher resolution MRI and other radiological scanners, will lead to a more realistic and human-like actual representation of anatomical structures to the point that someday it will be difficult to differentiate between the two.

Lastly, time and availability of resources are essential factors playing a role in learning and memory. Making sure the students have enough time and resources needs to keep improving with changes and evolution of old courses or the introduction of new ones. Neuroanatomy, one of the most important and complicated topics of anatomical and neurobiological sciences, needs special attention for allocating time and development and using new technology-enhanced resources.

Limitations

As the study was planned and conducted during the immediate response to sudden changes and school closures during the COVID-19 pandemic, it suffers from various limitations. Firstly, the design was quasi-experimental and pre-post interventional with only one group of students (the experimental group). There was no control group or randomization involved in the study as it was against our institution's educational and ethical standards. However, a counterbalanced approach (half students 2D to 3D, other half 3D to 2D) could have been suitable in such conditions. But our approach had a few advantages as well. Using a single group for pre-post study eliminated the effects of confounders such as spatial and cognitive abilities, attention, and memory skills. It also greatly increased the sensitivity of the experiment to the effects of the intervention, so fewer students were needed, and we could get significant results from a small sample size. Regarding the small sample size, the results may not be generalizable to other populations. However, this study involved no sampling and included all the undergraduate and graduate students enrolled in our department's Fundamentals of Neuroscience course; hence it may be generalized to other neuroscience students across the country, providing some external validity. Also, a small sample size creates a bigger hurdle to achieving significance, which can be surpassed only with medium and large effect sizes as encountered in our study, and once achieved, there is no doubt about these effects (Norman, 2010).

Another limitation of the pre-post interventional method may be the learning and test effects on the post-interventional counterparts. Some may argue that the post-interventional results are mainly due to the effects of time and repetition. Regarding this, a few precautions were taken. Firstly, the RRR pre- and post-surveys were worded carefully and designed to elicit responses explicitly targeted to the 2D and 3D intervention, and the observed differences were backed by student comments in the qualitative survey about these interventions. Also, the pictures used on the pre-and post-tests were taken respectively from the 2D learning resource and 3D software provided for interventions; hence performance on these tests was more due to the immediate effects of the intervention and less due to generalization of knowledge.

Regarding time, it was mentioned by the students themselves (and agreed upon by the investigators) that there was not enough time in the first place and that the course was very complicated to be handled in that small time frame and needed restructuring. This reduced time and increased complexity can make the study design feel more instantaneous (just like in calculus when we make the changes smaller and smaller), creating the illusion that both the 2D and 3D interventions occurred at the same time and independently led to the results as case-control rather than pre-post method. This also explains when results go in one direction after the 2D intervention but then completely flip over to the opposite direction after the 3D intervention, like comparing superficial vs. deep structures and studying the effects of gender and previous experience. Even if we reject these explanations and include the effect of time and repetition in the discussion, we can still say that 3D interventions greatly accelerated the results achieved by the study, as time itself was "not enough" to produce these results. Moreover, the presence of large (and medium) effect sizes cannot be ignored just due to the effects of a pre-post design. Also, while researchers may like to experimentally compare 2D vs. 3D (vs. cadaveric labs), the students want to avoid picking a side and going along with only one. They instead have all modalities at their disposal as a multidimensional approach (it is about an apples & oranges situation requiring both, not apples vs. oranges), which is why these effects are important even if they come from 2D+3D situations rather than 3D alone. So, this study is a step in that direction and proves the importance and effects of a hybrid/multidimensional learning mode.

Long-term retention of the gained neuroanatomical knowledge was not tested, which may be another limitation of this study. Generalization of knowledge could have been adequately tested, which could have been done by asking additional questions on clinical pictures and radiological scans. The questions asked in the tests were limited in number and very basic, so they may not generalize to more complex testing conditions. Likewise, generalization to more detailed courses in neuroanatomy, especially in advanced medicine, may not be possible. Also, the students faced novelty issues with the software, which could have been resolved by more intensive training to use them.

Importance of the study

The study was initiated as an immediate response to our department's COVID-19-related changes and transitions in neuroscience education. It involved the creation of a new 3D learning module and its evaluation using the newly developed RRR and Confidence in topics surveys and tests. It involved both quantitative and qualitative analysis, using self-reported surveys along with tests to provide a more balanced approach to student feedback and performance. Despite all the limitations, the study successfully reconfirmed the benefits of 3D software in learning neuroanatomy, especially for complex structures. However, we described these benefits with effect sizes to enable comparison, which is missing in many other studies. These effect sizes were almost all large for significant observations in the 3D vs. 2D comparisons, variable (few large and mostly medium effects) for superficial vs. deep comparisons, and all medium for effects of gender and previous experience.

The study uncovers some crucial issues. It shows that students who have previously studied other methods may be biased against these new 3D technologies, which may not occur in the future when these are the only options left as we move away from cadaver-based teaching or in settings already facing a lack of resources. It also shows that 3D technologies bring equality and social justice by benefitting all students equally, no matter their previous grades, experience, gender, or academic level. Studies have shown that playing video games helps reduce the gender gap and promote equality in video games and even STEM fields by differentially enhancing spatial skills (Ratan et al., 2020). The use of 3D educational software might be a better and more direct method to achieve these goals in STEM education. Regarding neuroanatomy, studies have indicated that females show a greater amount of Neuroanxiety and Neurophobia and respond better to early exposure to educational interventions by developing a greater increase in neuroanatomy self-efficacy, although still lagging behind male students (Bergden, 2021). Our study reconfirms this idea of social need and justice but also points to a possibility that the observed (near) equalities may not be accurate and that the gender differences hide inside deeper levels of the structural complexity of neuroanatomy until the day arrives when 3D software is evolved enough to fade them away.

Future directions

Further studies with better experimental conditions and longer intervals are needed to confirm our findings. Assessment of organizational and institutional benefits along with student benefits also needs to be explored. Testing multidimensional approaches [with 2D + tactile approach (plastic and 3D printed models and cadavers) + 3D and VR/AR in a more (virtually) social way] may give more insights. Limiting 3D technologies to be used only for complex and deep structures and difficult dissections like white matter tracts can be a more efficient approach. The addition of complex and real clinical images to enhance learning and generalization and more evolved software to provide more realistic renderings in the future will increase the results. Finally, testing on other populations across disciplines, country, and cultures may prove its external validity.

CONCLUSIONS

The study attempts to address the needs identified in the literature regarding educational research and provides evidence of the validity and reliability of a newly developed 13-item RRR instrument to measure student motivation and learning benefits in an online environment for learning neuroanatomy. It is a valuable tool based on the Kirkpatrick and ARCS models of learning. It can be adapted and used to measure the effectiveness of existing and newly developed study interventions in other disciplines and institutions, which will confirm its reliability. Another instrument, the 5-item Confidence in topics survey, was developed, validated, and tested for reliability in measuring student confidence in their knowledge of some commonly asked structures in neuroanatomy.

The study also adds to the literature on using 3D resources for neuroanatomy education and provides scientific evidence for its benefits. In our study, the students spent long and frequent durations on the 3D software due to the better visuals and colorful graphics they provided compared to 2D resources. Despite novelty being an issue and having a slight unease in its use (especially in female students) acting as an entry barrier, the students highly appreciated the benefits of the 3D resources and recommended its future use. Based on the RRR survey overall and in each of its components (Reaction, Relevance, and Results), it can be concluded that the 3D intervention significantly improved student learning experience and was very effective and, in fact, much more effective than 2D resources, thus confirming our first hypothesis. The use of the Confidence in topics survey and tests (overall as well as the superficial and deep components) gave the same conclusions regarding the higher effectiveness of 3D intervention over 2D resources, thus reconfirming our first hypothesis.

The RRR survey showed that females less appreciated the effectiveness of the 2D resources than males. Although there may be some usage-related entry barriers for females to use the 3D resources, their perception of the high effectiveness of 3D resources was not observed to be different from males. So, the motivation benefits of 3D over 2D intervention appeared to be more for females, and it suggested that the first hypothesis may be truer for female students when considering their internal feedback of the 3D intervention itself, but this was not reflected externally on the Confidence in topics surveys or tests and thus could not be reconfirmed. It suggests that while the females may have a greater internal resistance or appreciation for the benefits of the 3D resources than male students, this may or may not translate into actual gender differences in the outcome of the learning process, maybe due to more effort by the female students.

The 3D intervention enhanced students' exam performance on superficial and deep structures. However, greater pre-to-post increases on the exams were seen for the deep structures, thus confirming our second hypothesis of differential benefits on superficial and deep structures. Adding this differential benefit of deep structures to the fact that initially, on the pre-test, the scores in superficial structures were significantly higher than scores in deep structures but not so on the post-test shows that the enhanced learning of deeper structures by the 3D intervention also appeared to have attempted to bridge the difficulty gap between the superficial and deep structures, and brought some equality among performance in superficial and deep topics on the exams. The second hypothesis could not be reconfirmed as no differential effects on enhanced learning were observed in the superficial and deep components of the Confidence in topics survey. Further on, the findings were oppositely suggesting a possibility that the 3D intervention may have made them realize a greater difficulty gap between superficial and deep structures. This was all due to a pre-intervention insight that did not correlate with the actual situation on the pre-test, suggesting that surveys can sometimes be unreliable, especially after a less effective (2D) intervention.

While no gender effects were observed overall in the Confidence in topics surveys and tests, a couple of items showed differential effects of gender on the performance on the test after the 3D intervention. One (superficial) topic had a higher gain in performance pre-to-post by females, and one (deep) topic better performed by males on the post-test. These effects became clearer while analyzing the superficial and deep components of the test scores. It was observed that after the 3D intervention, female students performed more differentially on the superficial and deep structures than male students. The male students had a narrower difficulty gap between the superficial and deep structures, suggesting that they may have gained more on the deeper structures while bridging the difficulty gap compared to female students. This suggests that the second hypothesis may be more valid for male students. These findings were observed only on the test scores rather than on the Confidence in topics surveys, suggesting again an insight that may not entirely reflect the actual situation.

Although the confidence levels in topics did not reflect their test performance after studying the 2D resources, they did so after the 3D intervention indicating the development of a better insight by the 3D intervention. Also, the benefits of the 3D intervention on Confidence and performance were greatly experienced by all students and did not correlate with their experience on the 2D resources, thus indicating no special benefits to students advantaged at an earlier stage.

Younger generations are generally more technologically adaptable, yet no group interactions for the academic level were observed in any of the analyses. However, interestingly, having any previous experience in neuroanatomy brought some distaste among such learners for the 3D intervention used in our study. On the RRR survey, students with some prior experience in neuroanatomy found the 3D resources to be less effective than those with no prior experience reported. In contrast, no such effect of previous experience was observed on the scores of the 2D intervention. The effect of previous experience was again seen in the Confidence in Topics survey overall and in the deep structures after the 3D intervention, where students with some previous experience did not report as much Confidence as students with no previous experience.

However, no such differential effects of previous experience were observed on test performance. This again suggests an insight not corresponding to the actual situation, this time post-interventional in the case of students with some bias due to the previous experience of neuroanatomy.

The qualitative feedback further revealed that the 3D software makes studying the brain more accessible to the students and presents them with visually rich 3D renderings and many features that allow them to manipulate the brain freely in a virtual 3D space and study its superficial and deep relations in great detail, providing a better overview and spatial understanding and helping with long-term memory and retention. However, the software still needs to improve, evolve, and converge into a single program that suits various learners' needs. Debugging, refinement of features, reorganization, addition and standardization of clinical content, supplementary materials, personalization tools, and, most importantly, the evolution of 3D data into more realistic and human-like details will help the students adopt and prefer these 3D resources. Time and training are the two most important factors that may play a major role in determining the effect and success of the 3D intervention in enhancing learning in these students.

The goal of the 3D visualization is not to replace traditional cadaveric dissection but to enhance learning overall, especially in deep and complicated structures that may not be easily or at all dissectible - students will rarely encounter a complete dissected specimen of the caudate nucleus or the arcuate fasciculus. 3D visualization can fill the deficiencies in learning complex grey matter and most white matter structures in the brain. 3D technologies will undoubtedly become a mainstay in education as they did in the entertainment industry, and educators need to embrace and integrate them into existing and future curricula. A well-balanced approach of traditional methods and technological enhancements may be the best intervention, which educators need to help develop, evolve, and be prepared for its use.

In summary, the use of 3D software, although a little uneasy for new users, older learners, or specific other populations, once put in effect after some training on spatial ability, can bring more success and equality across a diverse range of populations with changing demographics, different spatial needs, and resources available to them. Enhanced visualization helps in better engagement and proves helpful to students who nowadays have increasingly greater levels of inattention due to a reduction in attention spans and less preference for typical 2D textbook methods. While males and females may initially experience the benefits in slightly different ways, 3D software does benefit them all. It may be essential for especially females whose participation in neuroscience and other biological sciences is continuing to grow. The RRR and Confidence in topics surveys can be helpful instruments in examining the effectiveness and success of various study interventions. However, these surveys can sometimes be inadequate in studying the intricate details or group differences after a less effective intervention where the students need more proper and complete insight into their academic needs and progress or if they have some bias due to previous experience. Hence, using surveys and test scores ensures that educators understand not only the student side of the story but also the real end effects and fulfillment of the goals of the study intervention.

REFERENCES

Ahmed H, Allaf M, Elghazaly H. COVID-19 and medical education [published correction appears in Lancet Infect Dis. 2020 May;20(5):e79]. Lancet Infect Dis. 2020;20(7):777-778. doi:10.1016/S1473-3099(20)30226-7

Andrade C. Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation. Indian J Psychol Med. 2018 Sep-Oct;40(5):498-499. doi: 10.4103/IJPSYM.IJPSYM_334_18. PMID: 30275631; PMCID: PMC6149308.

Arantes M, Arantes J, Ferreira MA. Tools and resources for neuroanatomy education: a systematic review. BMC Med Educ. 2018;18(1):94. Published 2018 May 3. doi:10.1186/s12909-018-1210-6

Bates R. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. Eval Program Plan. 2004;27(3):341–347. doi: 10.1016/j.evalprogplan.2004.04.011.

Bergden, Jessica S., "Anxiety and self-efficacy related to learning neuroanatomy in an integrated medical curriculum." (2021). Electronic Theses and Dissertations. Paper 3649. https://doi.org/10.18297/etd/3649

Bewley WL, O’Neil HF. Evaluation of medical simulations. Mil Med. 2013;178(Suppl 10):64–75.

Chang YS, Hu KJ, Chiang CW, Lugmayr A. Applying Mobile Augmented Reality (AR) to Teach Interior Design Students in Layout Plans: Evaluation of Learning Effectiveness Based on the ARCS Model of Learning Motivation Theory. Sensors (Basel). 2019;20(1):105. Published 2019 Dec 23. doi:10.3390/s20010105

Chick RC, Clifton GT, Peace KM, et al. Using Technology to Maintain the Education of Residents During the COVID-19 Pandemic. J Surg Educ. 2020;77(4):729-732. doi:10.1016/j.jsurg.2020.03.018

Craig, S., Tait, N., Boers, D. & McAndrew, D. Review of anatomy education in Australian and New Zealand medical schools. ANZ journal of surgery 80, 212–216 (2010).

Daugherty KK. ARCS motivation model application in a pharmacy elective. Curr Pharm Teach Learn. 2019;11(12):1274-1280. doi:10.1016/j.cptl.2019.09.009

Deery C. The COVID-19 pandemic: implications for dental education. Evid Based Dent. 2020;21(2):46-47. doi:10.1038/s41432-020-0089-3

Desai BK. Clinical implications of the COVID-19 pandemic on dental education. J Dent Educ. 2020;84(5):512. doi:10.1002/jdd.12162

Dong Y, Peng CY. Principled missing data methods for researchers. Springerplus. 2013 May 14;2(1):222. doi: 10.1186/2193-1801-2-222. PMID: 23853744; PMCID: PMC3701793.

Donnelly L, Patten D, White P, Finn G. Virtual human dissector as a learning tool for studying cross-sectional anatomy. Med Teach. 2009 Jun;31(6):553-5. DOI: 10.1080/01421590802512953. PMID: 19288305.

Dorri S, Farahani MA, Maserat E, Haghani H. Effect of role-playing on learning outcome of nursing students based on the Kirkpatrick evaluation model. J Educ Health Promot. 2019;8:197. Published 2019 Oct 24. doi:10.4103/jehp.jehp_138_19

Drapkin ZA, Lindgren KA, Lopez MJ, Stabio ME. Development and assessment of a new 3D neuroanatomy teaching tool for MRI training. Anat Sci Educ. 2015;8(6):502-509. doi:10.1002/ase.1509

Flanagan E, Walsh C, Tubridy N. 2007. ’Neurophobia’–Attitudes of medical students and doctors in Ireland to neurological teaching. Eur J Neurol 14: 1109–1112.

Fleagle TR, Borcherding NC, Harris J, Hoffmann DS. Application of flipped classroom pedagogy to the human gross anatomy laboratory: Student preferences and learning outcomes. Anat Sci Educ. 2018 Jul;11(4):385-396. doi: 10.1002/ase.1755. Epub 2017 Dec 28. Erratum in: Anat Sci Educ. 2018 Jul;11(4):429. PMID: 29283505; PMCID: PMC6381391.

Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. Med Teach. 2012;34(5):e288–e299. doi: 10.3109/0142159X.2012.668637.

Hope JK. 2018. Could educational technology replace traditional schools in the future? In: Khosrow-Pour M (Editor). Encyclopedia of Information Science and Technology. 4th Ed. Hershey, PA: IGI Global. p 2421–2430.

Hu A, Shewokis PA, Ting K, Fung K. Motivation in computer-assisted instruction. Laryngoscope. 2016 Aug;126 Suppl 6:S5-S13. DOI: 10.1002/lary.26040. Epub 2016 Jun 16. PMID: 27307270.

Jamil Z, Saeed AA, Madhani S, Baig S, Cheema Z, Fatima SS. Three-dimensional Visualization Software Assists Learning in Students with Diverse Spatial Intelligence in Medical Education. Anat Sci Educ. 2019 Sep;12(5):550-560. doi: 10.1002/ase.1828. Epub 2018 Oct 30. PMID: 30376698.

Jones TL, Baxter MA, Khanduja V. A quick guide to survey research. Ann R Coll Surg Engl. 2013 Jan;95(1):5-7. doi: 10.1308/003588413X13511609956372. PMID: 23317709; PMCID: PMC3964639.

Jozefowicz RF. 1994. Neurophobia: The fear of neurology among medical students. Arch Neurol 51:328–329.

Khan, S. (2012). The One World School House: Education Reimagined. London: Hodder and Stoughton.

Khot Z, Quinlan K, Norman GR, Wainman B. The relative effectiveness of computer-based and traditional resources for education in anatomy. Anat Sci Educ. 2013 Jul-Aug;6(4):211-5. doi: 10.1002/ase.1355. Epub 2013 Mar 18. PMID: 23509000.

Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor Dent Endod. 2017 May;42(2):152-155. doi: 10.5395/rde.2017.42.2.152. Epub 2017 Mar 30. PMID: 28503482; PMCID: PMC5426219.

Kogan M, Klein SE, Hannon CP, Nolte MT. Orthopaedic Education During the COVID-19 Pandemic. J Am Acad Orthop Surg. 2020;28(11):e456-e464. doi:10.5435/JAAOS-D-20-00292

Koller, D. (2011). Death Knell for the Lecture: Technology as a Passport to Personalized Education. Available online at: http://www.nytimes.com/2011/12/06/science/daphne-koller-technology-as-a-passport-to-personalized-education.html?pagewanted=alland_r=0

Kopp SL, Rathmell JP. Acquiring New Technical Skills and Aptitude for Mental Rotation. Anesthesiology. 2015 Nov;123(5):993-4. doi: 10.1097/ALN.0000000000000871. PMID: 26389553.

La Duke P. How to evaluate training: using the kirkpatrick model. Prof Saf. 2017;62(8):20. [Google Scholar]

Lim KH, Loo ZY, Goldie SJ, Adams JW, McMenamin PG. Use of 3D printed models in medical education: A randomized control trial comparing 3D prints versus cadaveric materials for learning external cardiac anatomy. Anat Sci Educ. 2016 May 6;9(3):213-21. doi: 10.1002/ase.1573. Epub 2015 Oct 15. PMID: 26468636.

Lufler RS, Zumwalt AC, Romney CA, Hoagland TM. Effect of visual-spatial ability on medical students' performance in a gross anatomy course. Anat Sci Educ. 2012 Jan-Feb;5(1):3-9. doi: 10.1002/ase.264. Epub 2011 Nov 29. PMID: 22127919.

Maddineshat M, Hashemi M, Besharati R, Gholami S, Ghavidel F. The effectiveness of clinical teaching of mental health courses in nursing using clinical supervision and Kirkpatrick's model. Electron Physician. 2018;10(1):6265-6272. Published 2018 Jan 25. doi:10.19082/6265

Mowlem FD, Agnew-Blais J, Pingault JB, Asherson P. Evaluating a scale of excessive mind wandering among males and females with and without attention-deficit/hyperactivity disorder from a population sample. Sci Rep. 2019 Feb 28;9(1):3071. DOI: 10.1038/s41598-019-39227-w. PMID: 30816143; PMCID: PMC6395591.

Naaz, Farah, "Learning from graphically integrated 2D and 3D representations improves retention of neuroanatomy." (2012). Electronic Theses and Dissertations. Paper 1034. https://doi.org/10.18297/etd/1034

National Research Council. 2006. Learning to Think Spatially: GIS as a Support System in the K-12 Curriculum. 1st Ed. Washington DC: The National Academies Press. 332 p.

National Science Foundation, National Center for Science and Engineering Statistics. Special Report NSF 17-310. Arlington, VA: 2017. Women, minorities, and persons with disabilities in science and engineering: 2017. Available at: https://www.nsf.gov/statistics/2017/nsf17310/data.cfm.

Newcombe NS. 2010. Picture this: Increasing math and science learning by improving spatial thinking. Am Educ 34:29

Newman NA, Lattouf OM. Coalition for medical education-A call to action: A proposition to adapt clinical medical education to meet the needs of students and other healthcare learners during COVID-19. J Card Surg. 2020;35(6):1174-1175. doi:10.1111/jocs.14590

Norman, Geoffrey. (2010). LIkert scales, levels of measurement adn the “laws” of statistics. Advances in health sciences education : theory and practice. 15. 625-32. 10.1007/s10459-010-9222-y.

Ramos RL, Esposito AW, O'Malley S, Smith PT, Grisham W. Undergraduate Neuroscience Education in the U.S.: Quantitative Comparisons of Programs and Graduates in the Broader Context of Undergraduate Life Sciences Education. J Undergrad Neurosci Educ. 2016 Apr 15;15(1):A1-A4. PMID: 27980463; PMCID: PMC5105957.

Ramos RL, Alviña K, Martinez LR. Diversity of Graduates from Bachelor's, Master's and Doctoral Degree Neuroscience Programs in the United States. J Undergrad Neurosci Educ. 2017 Nov 15;16(1):A6-A13. PMID: 29371835; PMCID: PMC5777839.

Ratan, R., Shen, C., & Williams, D. (2020). Men Do Not Rule the World of Tanks: Negating the Gender-Performance Gap in a Spatial-Action Game by Controlling for Time Played. American Behavioral Scientist, 64(7), 1031–1043. https://doi.org/10.1177/0002764220919147

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., and Kingstone, A. (2012). Everyday attention: variation in mind wandering and memory in a lecture. Appl. Cogn. Psychol. 26, 234–242. DOI: 10.1002/ACP.1814

Risko, E. F., Buchanan, D., Medimorec, S., and Kingstone, A. (2013). Every attention: mind wandering and computer use during lectures. Comput. Educ. 26, 234–242. DOI: 10.1037/ACP.1814

Roach VA, Fraser GM, Kryklywy JH, Mitchell DGV, Wilson TD. Guiding Low Spatial Ability Individuals through Visual Cueing: The Dual Importance of Where and When to Look. Anat Sci Educ. 2019 Jan;12(1):32-42. doi: 10.1002/ase.1783. Epub 2018 Mar 30. PMID: 29603656.

Rosenthal, R. (1991). Meta-analytic procedures for social research. SAGE Publications, Inc., https://dx.doi.org/10.4135/9781412984997

Rosnow RL. Effect sizes for experimenting psychologists. Can J Exp Psychol. 2003 Sep;57(3):221-37. doi: 10.1037/h0087427. Erratum in: Can J Exp Psychol. 2009 Jun;63(2):123. PMID: 14596479.

Sahi PK, Mishra D, Singh T. Medical Education Amid the COVID-19 Pandemic. Indian Pediatr. 2020;57(7):652-657.

Schon F, Hart P, Fernandez C. 2002. Is clinical neurology really so difficult? J Neurol Neurosurg Psychiatry 72:557–559.

Szpunar, K. K., Khan, N. Y., and Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. Proc. Natl. Acad. Sci. U.S.A. 110, 6313–6317. doi: 10.1073/pnas.1221764110

Szpunar KK, Moulton ST, Schacter DL. Mind wandering and education: from the classroom to online learning. Front Psychol. 2013 Aug 1;4:495. DOI: 10.3389/fpsyg.2013.00495. PMID: 23914183; PMCID: PMC3730052.

Tallent-Runnels, M. K., Thomas, J. A., Lan, W. Y., Cooper, S., Ahern, T. C., Shaw, S. M., et al. (2006). Teaching courses online: A review of the research. Review of Educational Research, 76(1), 93−135.

Tomczak, Maciej & Tomczak, Ewa. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. 21. 19-25.

Wang, C., Daniel, B.K., Asil, M. et al. A Randomised Control Trial and Comparative Analysis of Multi-Dimensional Learning Tools in Anatomy. Sci Rep 10, 6120 (2020). https://doi.org/10.1038/s41598-020-62855-6

Welch MC, Yu J, Larkin MB, Graves EK, Mears D. A Multimedia Educational Module for Teaching Early Medical Neuroanatomy. MedEdPORTAL. 2020;16:10885. Published 2020 Mar 6. doi:10.15766/mep_2374-8265.10885

Zinchuk AV, Flanagan EP, Tubridy NJ, Miller WA, McCullough LD. 2010.Attitudes of US medical trainees towards neurology education: “Neurophobia”- A global issue. BMC Med Edu 10:49.

APPENDICES

Appendix A: IRB review outcome letter(s)

UNIVERSITY OF LOUISVILLE

Human Subjects Protection Program Office
MedCenter One – Suite 200
501 E. Broadway
Louisville, KY 40202-1798
Office: 502.852.5188 Fax: 502.852.2164

DATE:	September 15, 2020
TO:	Brian M Davis, Ph.D.
IRB NUMBER:	20.0756
STUDY TITLE:	Understanding neuroanatomy in a virtual 3D environment: Effects of use of 3D software in Neuroanatomy education for understanding simple and complex brain structures
REFERENCE #:	714729
IRB STAFF CONTACT:	Sherry Block 852-2163 slbloc04@louisville.edu

The amendment request was received by the Human Subjects Protection Program Office. It was reviewed by the HSPPO staff and found to be complete.

The modifications include:

- Personnel Change: Addition of: Chad L Samuelsen

If you have any questions, please contact: Sherry Block 852-2163 slbloc04@louisville.edu

Sincerely,

Sherry Block

Sherry Block, BS
IRB Analyst

UNIVERSITY OF
LOUISVILLE

Human Subjects Protection Program Office
MedCenter One – Suite 200
501 E. Broadway
Louisville, KY 40202-1798

DATE:	September 11, 2020
TO:	Brian M Davis, Ph.D.
FROM:	The University of Louisville Institutional Review Board
IRB#:	20.0756
STUDY TITLE:	Understanding neuroanatomy in a virtual 3D environment: Effects of use of 3D software in Neuroanatomy education for understanding simple and complex brain structures
REFERENCE #:	713923
DATE OF REVIEW:	09/10/2020
IRB STAFF CONTACT:	Sherry Block 852-2163 slbloc04@louisville.edu

This study was reviewed on 09/10/2020 and determined by the Vice-Chair of the Institutional Review Board that the study is exempt according to 45 CFR 46.101(b) under Category 2.

This study was also approved through 45 CFR 46.117(c), which means that an IRB may waive the requirement for the investigator to obtain a signed informed consent form for some or all subjects.

Documents/Attachments reviewed and approved:

Title	Version #	Version Date	Outcome
Survey instrument	Version 1.0	08/27/2020	Approved
Study Protocol	Version 1.0	08/27/2020	Approved
preamble	Version 1.0	08/27/2020	Approved

Requirements for an exempt study:

- Any study documents submitted with this protocol must be used in the form in which they were approved.
- Human Subjects & HIPAA Research training are required for all study personnel. It is the responsibility of the investigator to ensure that all study personnel maintain current Human Subjects & HIPAA Research training while the study is ongoing.
- Personnel amendments must be submitted to the IRB to add/remove research personnel from your study team.
- If your research focus or activities change, please submit an Amendment to the IRB for review to ensure that the indicated exempt category still applies.

Additional reporting, such as submission of continuation reviews, is not required.

For guidance on using iRIS, including finding your approved documents, please follow the instructions at https://louisville.edu/research/humansubjects/iRISSubmissionManual.pdf

Site Approval
Permission from the institution or organization where this research will be conducted **must** be obtained before the research can begin. For example, site approval is required for research conducted in UofL Hospital/UofL Health, Norton Healthcare, and Jefferson County Public Schools, etc...

Privacy & Encryption Statement
The University of Louisville's Privacy and Encryption Policy requires identifiable medical and health records; credit card, bank account and other personal financial information; social security numbers; proprietary research data; and dates of birth (when combined with name, address and/or phone numbers) to be encrypted. For additional information: http://louisville.edu/security/policies.

Implementation of Changes to Previously Approved Research
Prior to the implementation of any changes in the approved research, the investigator must submit modifications to the IRB and await approval before implementing the changes, unless the change is being made to ensure the safety and welfare of the subjects enrolled in the research. If such occurs, a Protocol Deviation/Violation should be submitted within five days of the occurrence indicating what safety measures were taken, along with an amendment to revise the protocol.

Unanticipated Problems Involving Risks to Subjects or Others (UPIRTSOs)
A UPIRTSO is any incident, experience, or outcome, which has been associated with an unexpected event(s), related or possibly related to participation in the research, and suggests that the research places subjects or others at a greater risk of harm than was previously known or suspected. The investigator is responsible for reporting UPIRTSOs to the IRB within 5 working days. Use the UPIRTSO form located within the iRIS system. Event reporting requirements can be found at: http://louisville.edu/research/humansubjects/lifecycle/event-reporting.

Payments to Subjects
In compliance with University policies and Internal Revenue Service code, payments to research subjects from University of Louisville funds, must be reported to the University Controller's Office. For additional information, please call 852-8237 or email controll@louisville.edu. For additional information: http://louisville.edu/research/humansubjects/policies/PayingHumanSubjectsPolicy201412.pdf

Serge A. Martinez

Serge A. Martinez, M.D., J.D. Vice Chair
Biomedical Institutional Review Board
SAM/cll

We value your feedback; let us know how we are doing: https://www.surveymonkey.com/r/CCLHXRP

Full Accreditation since June 2005 by the Association for the Accreditation of Human Research Protection Programs, Inc.

Appendix B: Preamble Consent Document

Understanding neuroanatomy in a virtual 3D environment: Effects of use of 3D software in Neuroanatomy education for understanding simple and complex brain structures

Date:

Dear Neuroscience Graduate and Undergraduate students:

You are being invited to participate in a research study by answering questions in the attached survey/an online survey (link provided), to study the effects of 3D software in Neuroanatomy education for understanding simple and complex 3D brain structures. This study is being conducted by Dr. Brian Davis, Dr. Chad Samuelsen, Dr. John Pani and Akash Khare of the University of Louisville. There are no known risks for your participation in this research study. The information collected may not benefit you directly but may be helpful to others as the information you provide will help neuroscience educators improve the neuroanatomy curriculum in graduate and undergraduate teaching programs. You will fill an online survey, which will take approximately 25 minutes to complete. It will be followed by an online educational test to assess you knowledge, which will take about 5 minutes. The completed online survey and test will be secured inside password-protected encrypted files on the Blackboard server and the Department of Anatomical Sciences and Neurobiology, University of Louisville.

Individuals from the University of Louisville Department of Anatomical Sciences and Neurobiology, the Institutional Review Board (IRB), the Human Subjects Protection Program Office (HSPPO), and other regulatory agencies may inspect these records. In all other respects, however, the data will be held in confidence to the extent permitted by law. Should the data be published, your identity will not be disclosed.

Taking part in this study is voluntary. By answering survey questions, you agree to take part in this research study. You do not have to answer any questions that make you uncomfortable. If you decide to be in this study, you may stop taking part at any time. If you decide not to be in this study or if you stop taking part at any time, you will not lose any benefits for which you may qualify. But we will highly appreciate if you take the time to submit this short survey, so that we can assess how to improve the neuroscience educational experience for you and future students.

If you have any concerns, or complaints about the research study, please contact Dr. Brian Davis (Phone: 502-852-1333, email: bm.davis@louisville.edu).

If you have any questions about your rights as a research subject, you may call the Human Subjects Protection Program Office at (502) 852-5188. You can discuss any questions about your rights as a research subject, in private, with a member of the Institutional Review Board (IRB). You may also call this number if you have other questions about the research, and you cannot reach the research staff, or want to talk to someone else. The IRB is an independent committee made up of people from the University community, staff of the institutions, as well as people from the community not connected with these institutions. The IRB has reviewed this research study.

If you have concerns or complaints about the research or research staff and you do not wish to give your name, you may call the University Integrity and Compliance hotline at 1-877-852-1167. This is a 24-hour hot line answered by people who do not work at the University of Louisville.
Thank you for your participation.
Sincerely,
Signatures of Investigator and Co-Investigator
(Dr. Brian Davis, Dr. Chad Samuelsen, Dr. John Pani and Akash Khare)

Appendix C: Survey Instruments

SAMPLE PRE-INTERVENTIONAL SURVEY

Demographic survey:

1. Gender:
2. Current Level of study: Undergraduate/Graduate student
3. Did you study neuroanatomy previously? If yes, then how many semesters of study?

The purpose of the following set of questions is to evaluate your reaction specifically to the PDF resources (lab manuals) provided for the course:

4. How attentive and engaged were you while using these resources?

1	2	3	4	5
Extremely inattentive and disengaged	A little inattentive and disengaged	Neither inattentive nor attentive	A little attentive and engaged	Extremely attentive and engaged

5. How much did you enjoy studying via these resources?

1	2	3	4	5
Extremely Boring	A little boring	Neither boring nor enjoyable	A little enjoyable	Extremely enjoyable

6. How much did it stimulate your curiosity and interest, and then motivate you to become an active learner?

1	2	3	4	5
Extremely unmotivated to learn	A little unmotivated to learn	Neither unmotivated nor motivated	A little motivated to learn	Extremely motivated to learn

The purpose of the following set of questions is to evaluate your learning experience, specifically to the PDF resources (lab manuals) provided for the course:

7. Overall, how useful or beneficial was the learning experience for you, using these resources?

1	2	3	4	5
Extremely useless and non-beneficial	A little useless and non-beneficial	Neither useless nor useful	A little useful and beneficial	Extremely useful and beneficial

8. How appealing was the visual content presented in these resources?

1	2	3	4	5
Extremely unappealing	A little unappealing	Neither unappealing nor appealing	A little appealing	Extremely appealing

9. How much did it help you in visualization of simple brain structures?

1	2	3	4	5
Confused a lot	Confused a little	Neither confused nor helped	Helped a little	Helped a lot

10. How much did it help you in visualization of complex brain structures?

1	2	3	4	5
Confused a lot	Confused a little	Neither confused nor helped	Helped a little	Helped a lot

11. How much did it improve your spatial understanding of structures in the brain and the interrelationships between these structures?

1	2	3	4	5
Confused a lot	Confused a little	Neither confused nor helped	Helped a little	Helped a lot

12. How helpful do you think these materials will be for building long-term retention and recall?

1	2	3	4	5
Extremely hindering	A little hindering	Neither hindering nor helpful	A little helpful	Extremely helpful

The purpose of the following set of questions is to evaluate your confidence after studying specifically on the PDF resources (lab manuals) provided for the course:

13. How easy or difficult do you personally think neuroanatomy is actually?

1	2	3	4	5
Extremely difficult	A little difficult	Neither difficult nor easy	A little easy	Extremely easy

14. How much did these resources change your attitude and confidence towards neuroanatomy?

1	2	3	4	5
Negatively changed a lot	Negatively changed a little	Unchanged	Positively changed a little	Positively changed a lot

The purpose of the following set of questions is to evaluate overall impact after studying specifically on the PDF resources (lab manuals) provided for the course:

15. How satisfied do you feel after learning using lectures and additional resources?

1	2	3	4	5
Extremely unsatisfied	A little unsatisfied	Neither unsatisfied nor satisfied	A little satisfied	Extremely satisfied

16. How well-prepared do you feel for exams?

1	2	3	4	5
Extremely unprepared	A little unprepared	Neither unprepared nor prepared	A little prepared	Extremely well-prepared

The purpose of the following set of questions is to evaluate what you feel about various topics taught, specifically on the PDF resources (lab manuals) provided for the course:

17. How confident are you about your identification skills with regards to superficial structures like lateral sulcus, precentral gyrus, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

18. How confident are you about your identification skills with regards to superficial but relatively hidden structures like insula, cingulate gyrus, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

19. How confident are you about your identification skills with regards to structures in deeper layers like claustrum, extreme capsule, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

20. How confident are you about your identification skills with regards to c-shaped or complex structures like the caudate nucleus, fornix, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

21. How confident are you about your identification skills with regards to simple-shaped or ovoid structures like the thalamus, red nucleus, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

Prior software experience:

22. How easy is it for you to learn to use new software?

1	2	3	4	5
Extremely difficult	A little difficult	Neither difficult nor easy	A little easy	Extremely easy

23. Have you previously seen or used ‘Visible Body’ or ‘Anatomy Learning’ or any other software to study neuroanatomy? If yes, how familiar are you with these kinds of programs?

1	2	3	4	5
Never seen any of them	Seen but not used	Used but not at all familiar	Used and a little familiar	Used and very familiar

SAMPLE POST-INTERVENTIONAL SURVEY

Software usage statistics:

Regarding time spent on software before/after labs:

1. Average number of times you opened the software, other than during lab hours (0,1, 2-5, 6-10, more than 10):
2. Average time spent during off-lab usage (in minutes):

The purpose of the following set of questions is to evaluate your reaction specifically to the 3D software ('Visible Body' and 'Anatomy Learning') used for the course:

3. How attentive and engaged were you while using these resources?

1	2	3	4	5
Extremely inattentive and disengaged	A little inattentive and disengaged	Neither inattentive nor attentive	A little attentive and engaged	Extremely attentive and engaged

4. How much did you enjoy studying via these resources?

1	2	3	4	5
Extremely Boring	A little boring	Neither boring nor enjoyable	A little enjoyable	Extremely enjoyable

5. How much did it stimulate your curiosity and interest, and then motivate you to become an active learner?

1	2	3	4	5
Extremely unmotivated to learn	A little unmotivated to learn	Neither unmotivated nor motivated	A little motivated to learn	Extremely motivated to learn

The purpose of the following set of questions is to evaluate your learning experience, specifically to the 3D software ('Visible Body' and 'Anatomy Learning') used for the course:

6. Overall, how useful or beneficial was the learning experience for you, using these resources?

1	2	3	4	5
Extremely useless and non-beneficial	A little useless and non-beneficial	Neither useless nor useful	A little useful and beneficial	Extremely useful and beneficial

7. How appealing was the visual content presented in these resources?

1	2	3	4	5
Extremely unappealing	A little unappealing	Neither unappealing nor appealing	A little appealing	Extremely appealing

8. How much did it help you in visualization of simple brain structures?

1	2	3	4	5
Confused a lot	Confused a little	Neither confused nor helped	Helped a little	Helped a lot

9. How much did it help you in visualization of complex brain structures?

1	2	3	4	5
Confused a lot	Confused a little	Neither confused nor helped	Helped a little	Helped a lot

10. How much did it improve your spatial understanding of structures in the brain and the interrelationships between these structures?

1	2	3	4	5
Confused a lot	Confused a little	Neither confused nor helped	Helped a little	Helped a lot

11. How helpful do you think these materials will be for building long-term retention and recall?

1	2	3	4	5
Extremely hindering	A little hindering	Neither hindering nor helpful	A little helpful	Extremely helpful

12. How well organized was the software regarding ease of use and navigation?

1	2	3	4	5
Extremely unorganized	A little unorganized	Neither unorganized nor organized	A little organized	Extremely organized

13. How easy was it to virtually dissect the brain?

1	2	3	4	5
Extremely difficult	A little difficult	Neither difficult nor easy	A little easy	Extremely easy

14. How much would you recommend it for future use for other students, in this course or other similar courses?

Strongly recommend avoiding them	Weakly recommend avoiding them	No recommendation	Weakly recommend using them	Strongly recommend using them

The purpose of the following set of questions is to evaluate your confidence now, after studying specifically on the 3D software ('Visible Body' and 'Anatomy Learning') used for the course:

15. How easy or difficult do you now think neuroanatomy is actually?

1	2	3	4	5
Extremely difficult	A little difficult	Neither difficult nor easy	A little easy	Extremely easy

16. How much did these resources change your attitude and confidence towards neuroanatomy?

1	2	3	4	5
Negatively changed a lot	Negatively changed a little	Unchanged	Positively changed a little	Positively changed a lot

The purpose of the following set of questions is to evaluate overall impact now, after studying specifically on the 3D software ('Visible Body' and 'Anatomy Learning') used for the course:

17. How satisfied do you feel after learning using the 3D software?

1	2	3	4	5
Extremely unsatisfied	A little unsatisfied	Neither unsatisfied nor satisfied	A little satisfied	Extremely satisfied

18. How well-prepared do you feel for exams?

1	2	3	4	5
Extremely unprepared	A little unprepared	Neither unprepared nor prepared	A little prepared	Extremely well-prepared

The purpose of the following set of questions is to evaluate what you now feel about various topics taught, specifically on the 3D software ('Visible Body' and 'Anatomy Learning') used for the course:

19. How confident are you about your identification skills with regards to superficial structures like lateral sulcus, precentral gyrus, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

20. How confident are you about your identification skills with regards to superficial but relatively hidden structures like insula, cingulate gyrus, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

21. How confident are you about your identification skills with regards to structures in deeper layers like claustrum, extreme capsule, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

22. How confident are you about your identification skills with regards to c-shaped or complex structures like the caudate nucleus, fornix, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

23. How confident are you about your identification skills with regards to simple-shaped or ovoid structures like the thalamus, red nucleus, etc.?

1	2	3	4	5
Extremely afraid or unprepared	A little afraid or unprepared	Neither afraid nor confident	A little confident and prepared	Extremely confident

Please provide brief responses to the following questions (in 1-3 sentences):

24. What did you like about the 3D module?
25. What did you like the least about the 3D module?
26. What could be improved in the 3D module?s

CURRICULUM VITAE

Akash Khare, MBBS, MS

EDUCATION:

Master's Degree, Department of Anatomical Sciences and Neurobiology [2019-2022]

University of Louisville School of Medicine, Louisville, KY, USA

Bachelor of Medicine and Bachelor of Surgery (M.B.B.S) [2005-2011]

[U.S. Equivalency: First professional degree in medicine, M.D. (Doctor of Medicine)

Verified by WES Credential Evaluation (World Education Services, New York)]

TEACHING EXPERIENCE:

Graduate TA, Department of Anatomical Sciences & Neurobiology, University of Louisville, KY [Fall 2020] (10-20 hours per week)

Math Volunteer at The ESL Newcomer Academy in Jefferson County Public Schools, Louisville, KY [2020]

Volunteer tutor for gross anatomy lab content to Univ of Louisville Dental students [2020]

Medical Demonstrator (on contract for one semester) [Feb-June 2018] Dept. of Anatomy , Postgraduate Institute of Medical Education and Research, Chandigarh, India

Medical Tutor, ESI Postgraduate Institute of Medical Science and Research [Dec 2014-Dec 2017] (40 hours per week) ESI-PGIMSR, Kolkata, India

RESEARCH EXPERIENCE

Curriculum improvement project, as part of master's thesis, under guidance of Dr. Brian Davis (PI), [2020]

Dr. Chad Samuelsen (Lab coordinator) and Dr. John Pani (Co-PI, Professor, Department of Psychology):

TRAININGS

Human Subjects and HIPAA (Health Insurance Portability and Accountability Act) Research Training

Collaborative Institutional Training Initiative (CITI Program), Citiprogram.org
[Jan 2020]

TECHNICAL SKILLS

Basic skills in 3D and virtual/augmented reality (VR/AR) technology: 3D educational-games design (via Autodesk 3ds Max & UNITY 3D game engine) for computers and mobile-app. 3D modelling and connectome extraction from diffusion weighted MRI scans. 3D scanning and 3D printing.

Fundamental knowledge of Programming (C++, JavaScript) and SSPS statistical software package.