Introduction

Neuroanatomy can be challenging, especially when studying deep structures and their interrelationships, due to voluminous study material and limited learning resources (Schon et al., 2002; Flanagan et al., 2007; Latini & Ryttlefors, 2019). Even physicians feel that neuroanatomy is their “Achilles' heel,” *neurophobia* coined to describe the fear of neuroscience (Jozefowicz, 1994; Zinchuk et al., 2010; Flanagan et al., 2007; Allen, 2018; Sotgiu et al., 2020). Decreased funding is an issue and courses often get restructured so students can graduate faster (Brenton, 2011). A decline in teaching hours dedicated to anatomy and a shortage of cadavers has been observed (Eldred & Eldred,1961; Craig et al., 2010; Allen, 2018; Arantes et al., 2018).

Cadaver-based learning in human anatomy is considered the gold standard by academics, so its replacement with 3D software is met with skepticism in the medical field (Kuyatt, 2012; Wang et al., 2020; Papa et al., 2022). Nevertheless, computer-based learning has been used in neuroanatomy education by medical schools, being found to be cost effective, accessible, and interactive; students also preferred lectures be supplemented with 3D resources (Hu et al., 2010; Kuyatt, 2012; Naaz, 2012; Allen, 2018; Wang et al., 2020). Moreover, during the covid-19 pandemic, 3D software proved to be an effective replacement for in-person labs (Attardi et al., 2020; Deery, 2020; Kogan et al., 2020; Desai, 2020; Ahmed et al., 2020; Chick et al., 2020; Sahi et al., 2020; Newman & Lattouf, 2020).

Curved and C-shaped morphology of structures like the caudate nucleus, lateral ventricles, and fornix can cause changes in their location, shape, and size throughout different sections which may lead to confusion. Sectional neuroanatomy has been used in studies to measure the effectiveness of visualization tools, as it is essential to anatomy and radiology (Chariker et al., 2011, 2012; Pani et al., 2013). Limited spatial abilities and 2D resources lead students to grasp neuroanatomy in canonical views, hindering a comprehensive understanding, which is worsened due to the emerging crisis in higher medical education (Chariker, 2009).

Neuroanatomy is complicated enough to block 3D mapping as a result of mental resistance, so the best way to learn spatial transformations in such situations is to see them explicitly (Naaz, 2012; Pani et al., 1996, 1997, 2005). While 3D software may offer less benefits to students with low spatial ability on an initial stage, they can offer an equal advantage if given enough time for repetition or mastery (Naaz, 2012; Allen, 2018). Specific visual cues like colorful graphics and labels can also help the students to learn faster as it helps them eliminate other structures that they may not have studied or encountered (Naaz, 2012; Gilbert et al., 2008).

Various studies have shown that computer-based programs are more likely to enhance anatomy learning as compared to traditional methods (Naaz, 2012; Kuyatt, 2012; Pani et al., 2005, 2013; Chariker et al., 2011, 2012; Allen, 2018; Arantes et al., 2018), while other studies have given mixed results. However, some reported 3D software to be inferior to the traditional approach, which could be attributed to short duration of study design, usage of oversimplified software imaging, or use in topics that do not require complex visualizations in the first place (Chariker, 2009; Arantes et al., 2018; Wang et al., 2020; Welch et al., 2020; Drapkin et al., 2015).

Various questionnaires and survey tools have been developed and validated to suit specialized needs, but it is necessary to have a proper strategy for learning as well as evaluation, to identify 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 methods of evaluating health sciences educational programs (Dorri et al., 2019). It assesses an educational program at four levels: (1) reaction; (2) learning ; (3) behavioral changes/application of learning and (4) results (Dorri et al., 2019; Frye & Hemmer, 2012; La Duke, 2017; Bates, 2004).

The Attention, Relevance, Confidence, and Satisfaction (ARCS) model of motivation has been another popular tool for assessing health science education programs (Daugherty, 2019; Chang et al., 2019). It is similar to the Kirkpatrick model, sharing some overlapping features. Attention and learner's reaction are linked, program relevance influences learning outcomes, confidence aligns with behavioral changes, and satisfaction reflects training results.

Using a suitable model for generating questionnaires will help effectively measure student feedback. In tailoring assessment specifically for neuroanatomy education, it is also important to elucidate which attributes of brain structures are better addressed by 3D resources compared to 2D resources. Our study created a questionnaire based on multiple models to assess the impact of 3D software on student learning and confidence in neuroanatomy.

Methods

A prospective study was conducted at the department of Anatomical Sciences & Neurobiology (ASNB) at the University of Louisville (UofL) School of Medicine in the online setting as a virtual program to replace a pre-existing in-person neuroscience lab course. 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 study population consisted of 10 graduate and 25 undergraduate students, totaling 35 participants. These students consisted of 15 males (43%) and 20 females (57%); 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). Ethical permission was taken from the Institutional Review Board (see Supplementary File 1). 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). 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. One student dropped out mid-study. This study was not funded by any sponsor.

The study involved implementing and evaluating an educational intervention that used a newly developed neuroanatomy learning module. Surveys and tests were designed as a one-group/within-population pre-post design. The 2D learning resource was based on the lab manuals provided for the course during previous years to learn the brain's internal structures. Students had to study and review these resources on their own before taking an online pre-interventional survey and pre-test. The following week started with a 3D educational module serving as an online neuroanatomy lab on sectional anatomy of the brain through online software. The school had already licensed the Visible Body (Argosy Publishing) software. Anatomylearning.com, a free online (WebGL based) website was also used, as its rendering quality, slicing tool, and details in the brain region seemed very useful in learning sectional anatomy.

Lab manuals were provided to the students, containing details of basic software features, instructions to use it to get to specific views in the software, and a checklist of anatomical structures. The online labs consisted of a brief introduction to the software and neuroanatomy, and the students were then divided into groups and used the 3D module and online lab manual to learn neuroanatomy independently. They explored various predetermined views to observe the 3D brain structures and cross sections from different angles, zoom in/out, and click on various tags.

After finishing the learning module they received an online post-interventional survey. A practical exam for the course was held at the end of the same week, and supplemented the post-interventional assessment for our study. It contained image-based questions on cross sectional anatomy.

Measurements

A pool of questions 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 substantially (Edward et al., 1989; Rovai et al., 2009; Stephan et al., 2017; Loorbach et al., 2015). 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 (see Supplementary File 2, Table S1), 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, 3 component, ‘Reaction-Relevance-Result’ (RRR) survey. It was coupled with 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, with no demographic or prior software usage questions. It contained additional 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.

The pre-and post-interventional surveys were measured on a 5-point Likert scale. Comparing mean Likert scores for the 2D module and 3D module on the surveys, and comparing performance on the tests examined whether the 3D module is more effective than the 2D module. A similar comparison for questions regarding specific neuroanatomy structures investigated if 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 the group differences in scores of motivation, confidence, and performance following the 2D or 3D module.

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. Five questions were carefully selected and used in a multiple-choice question format after approval by the study investigators and course instructors, 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 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. For simplicity of the analysis, only the superficial and deep structures were compared on their pre- and post-intervention scores on surveys and tests (see Supplementary File 2).

Analysis

Statistical analysis of quantitative data was done using IBM SPSS Version 24.0 software (IBM Corp., Armonk, NY). Statistical significance was established at $p < .05$. We validated the internal structure of the RRR and Confidence in topics post-surveys using Principal Component Analysis after confirming its applicability by checking Kaiser–Meyer–Olkin (KMO) value of sampling adequacy and Bartlett's test of sphericity, eigenvalues greater than 1, and acceptable factor loadings greater than 0.4. 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.

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). 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. Missing data in this part of the analysis was handled by excluding cases on a test-by-test basis.

All the results were shown along with calculated effect size measures, by 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 Cohen's guidelines for Pearson's r value to compare effect sizes (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).

For qualitative data, a thematic analysis was applied to the three open-ended answers from the post-survey by identifying important repeated themes on close examination of student feedback with an inductive and semantic approach.

Results

The Reaction-Relevance-Result Surveys:

Table 1: Outcomes of the RRR Survey

	Pre-survey		Post-survey		Pre-Post Comparison
	Mean (SD)	Significant Group Differences	Mean (SD)	Significant Group Differences	Pre-Post Differences

Reaction	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)
Relevance	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)
Results	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)

Note. Adapted from *Understanding Neuroanatomy in a Virtual 3D Environment (Table 1, p. 48, 50, 52),* by A. Khare, 2022, *Master's thesis*, University of Louisville. *https://doi.org/10.18297/etd/3998.* Used with permission.

Exploratory factor 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 Supplementary File 3, Table S2). 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 (α > 0.6, moderate) for the Result dimension. All the dimensions were also tested for external validity by correlating with the final score on the neuroanatomy exam, and strong correlations (τ_b >.32) were observed (see Supplementary File 3, Table S2). Factor analysis confirmed the three components as unidimensional. 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 mean scores of the post-survey were generally higher than the pre-survey for the RRR survey overall, as well as in all individual components and items . Almost all these differences were statistically significant (except the ‘interest’ and ‘satisfaction’ items), with effect sizes varying from large ($r_{\text{effect size}}$>0.5) for most individual items, all three components and the RRR survey overall) to medium (0.3<$r_{\text{effect size}}$<0.5) .

The mean scores of males were generally higher than females for all items in the pre-survey and most items in the post-survey. In the pre-survey, gender differences were relatively larger and statistically significant with medium effect sizes for a few individual items, and in each RRR dimension and RRR survey overall.

The mean scores of students with no prior experience with neuroanatomy were generally higher than students with some experience, only in the post survey (3D). No significant differences were found on the pre survey (2D), while significant results with medium effect sizes were found on the post survey for RRR survey overall, the Reaction and Result dimension and a few individual items.

<u>The Confidence in Topics Survey and Tests:</u>

Table 2: 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)

Note. Reprinted from *Understanding Neuroanatomy in a Virtual 3D Environment (Table 2, p. 61),* by A. Khare, 2022, *Master’s thesis*, University of Louisville. *https://doi.org/10.18297/etd/3998.* Used with permission.

PCA examination of the factor structure indicated that all the corresponding items loaded significantly for the confidence in topics survey under a single factor with good loadings for almost all items. It was also confirmed to be suitable by KMO and Bartlett's test of sphericity, and high internal validity and external validity (strong correlations with the RRR post-survey overall and final score on the neuroanatomy exam).

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 . Most of these differences were statistically significant p<.005 (except for Cingulate gyrus and Thalamus in tests), with effect sizes varying from large (for most items and overall, in surveys and tests) to medium (for Lateral sulcus in 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. Statistical significance was observed overall and in the item 'Thalamus', with medium effect sizes.

Superficial vs. Deep Structures:

Table 3: 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, $r_{\text{effect size}}$=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, $r_{\text{effect size}}$=0.05)		$\Delta_{\text{Pre}}<\Delta_{\text{Post}}$ (M_1=0.04, M_2=0.28, Z=-0.81, p=.419, N=32, $r_{\text{effect size}}$=0.14)
Confidence in topics Post-survey	Superficial>Deep (Z=-2.21, p=.027, N=32, $r_{\text{effect size}}$=0.39)		
Pre-test	Superficial>Deep (Z=-3.87, p<.001, N=34, $r_{\text{effect size}}$=0.66)		$\Delta_{\text{pre}}>\Delta_{\text{Post}}$ (M_1=0.32, M_2=0.07, Z=-2.40, p=.017, N=34, $r_{\text{effect size}}$=0.41)
Post-test	Superficial>Deep (Z=-1.29, p=.197, N=34, $r_{\text{effect size}}$=0.22)	$\Delta_{\text{Male}}<\Delta_{\text{Female}}$ (M_1=-0.07, M_2=0.18, U=88, p=.026, Z=-2.23, N=34, $r_{\text{effect size}}$=0.38)	

Note. Reprinted from *Understanding Neuroanatomy in a Virtual 3D Environment (Table 3, p. 76-77),* by A. Khare, 2022, *Master's thesis*, University of Louisville*. https://doi.org/10.18297/etd/3998.* Used with permission.

The post-intervention scores were greater than pre-intervention scores for both the subgroups (superficial and deep structures). All these differences were statistically significant, with effect sizes varying from large (for most sub-groups) to medium (for test-superficial) as shown in Table 3. 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, which showed that the increased difficulty for deeper structures was higher in the pre-test as compared to the post-test, with medium effect size while on the surveys, the effect size was small with no statistical significance in differences. Further, on the post-survey for deep structures, students with no previous experience in neuroanatomy scored significantly higher than those with some previous experience, with a medium effect size. Finally, in the post-test, females had significantly higher increase in scores concerning the increased difficulty with depth of structures, while males had a lower and negative increase in these scores (therefore greater scores for deeper topics than superficial), with medium effect size.

Comfortability with Software Usage

The mean scores of males were significantly higher than females with medium effect size, in the item 'ease of virtual dissection' in which male students expressed slight ease while female students expressed slight unease Supplementary File 3, Table S1).

Qualitative Feedback

Reviewing the answers for what the students liked and disliked about the program, and areas of improvement suggested, themes were identified and they were finally grouped into a maximum of five specific codings (Technology, Features, Content, Realism, and Time). 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). 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. 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.

Table 4: 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

Note. Reprinted from *Understanding Neuroanatomy in a Virtual 3D Environment (Table 4, p. 86-87),* by A. Khare, 2022, *Master's thesis*, University of Louisville*. https://doi.org/10.18297/etd/3998.* Used with permission.

Discussion

Authors have pointed out (Rovai et al., 2009; Tallent-Runnels et al., 2006) that educational research needs more systematic studies for online learning, based on learning theories and models of teaching, to measure the effectiveness of these courses, especially in domains of academic success and thinking skills rather than just student preferences or faculty satisfaction. Our RRR survey attempted to address these issues.

Analysis of the RRR survey overall and its every component (Reaction, Relevance, and Result), as well as the confidence in topics survey, showed that self-reported outcomes of the 3D intervention were all very positive, better than the 2D resource.

Female participation in STEM has been gradually increasing over time and surpassing the male population in many areas, including neuroscience (NSF, 2017; Ramos, 2017). In our study, the self-perceived effectiveness of 2D resources were better in males as compared to females. These differences were absent in the case of the confidence in topics surveys or tests, suggesting that although female students underestimated the effectiveness of the 2D resources, it did not lead to a difference in their confidence levels or performance. Males had more comfortability with 3D software usage but their self-perceived effectiveness of it was no different than the female students. This implies that there may be an entry barrier for females, but once these are crossed the 3D intervention eventually benefits everyone equally.

Regarding students who had no prior experience in neuroanatomy, their perception of the effectiveness of the 3D intervention on the RRR survey was higher than 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 decreased self-perception of reaction, relevance (only for visualization of only simple structures) and results. These differences due to previous experience in neuroanatomy were again seen in confidence in topics, but only in the post-survey and not the post-test, suggesting that it affected their likeability and confidence but not their performance. Since students with some previous experience in neuroanatomy underestimated the effectiveness of the 3D intervention overall, it also reduced their confidence levels in neuroanatomy knowledge of deep structures (however, it did not affect their performance in tests). Therefore, previous experience may hinder acceptance of newer technologies (but not performance), especially when they tend to replace traditional methods like cadaver-based learning. 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 transition to a more technologically enhanced form of neuroanatomy learning.

As expected, the superficial and deep structures both had significantly higher scores for the post-intervention (3D) counterparts. 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.

While comparing the scores of superficial and deep structures within each the four groups (pre-survey, pre-test, post-survey, post-test), superficial structures had a significantly higher score than deep structures in the post-survey and pre-test. This suggests that after the 3D intervention, students did have more confidence in superficial structures than deep structures, which means it increased their clarity about difficulty in deep structures. It may 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.

On the contrary, in tests they performed better on superficial structures on the 2D intervention and no such significant difference was found after the 3D intervention. This suggests that the 3D intervention may have bridged the difficulty gap between superficial and deep topics. While the 3D resource may improve the insight or bring forward the inequalities of understanding between superficial and deep topics, it does help to remove these inequalities in performance by bridging the difficulty gap.

We confirmed this by analyzing the pre-post comparison of the difference in performance on the superficial and deep topics (which signified an increased difficulty for deeper topics). A significant pre-post difference in the increased difficulty for deeper topics was not observed on surveys but was on tests. On tests, students experienced a significant amount of increased difficulty for the deeper topics only on the pre-intervention (2D) but not on the post-intervention (3D), confirming that the 3D intervention brought some equality between depths (superficial versus deep) by reducing the increased difficulties of deeper topics.

However, this equality between depths on test scores after the 3D intervention may not be accurate as some group interactions were present concerning gender in the post-test. After the 3D intervention, females had a significantly greater score difference between the superficial and deep structures. This suggests that female students experienced the deeper structures to be much more complicated than superficial ones, as compared to male students, even if an equality between depths was observed overall post 3D intervention.

Spatial ability has been shown to have a constant and residual effect on learning using 2D and 3D resources and some studies have shown that the female gender generally has lower innate spatial ability as compared to males (Allen, 2018). The gender differences we observed 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 after 3D intervention.

Although various 3D software have 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 something which is often underrated, unsupported and ignored (Jamil et al., 2019; National Research Council, 2006). Spatial ability has been proven as a reliable predictor of academic 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). People have inherently different spatial abilities, for example, males have higher mental rotation (MR) ability than females, and they outperform females in MR training workshops, take less time to complete spatial tasks, and achieve higher scores (Allen, 2018; 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 often unaffected by gender; hence its usage is beneficial when working with diverse populations (Kuyatt, 2012; Jamil et al., 2019; Khot et al., 2013; Lim et al., 2016). Spatial ability has been shown to have a constant and residual effect on learning using 2D and 3D resources, but 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; Allen, 2018).

3D software is one of the interventions that can bring down the spatial differences between genders over time, as it reduces cognitive load, especially demands of spatial ability, and helps training it but does not completely solve the situation yet. At first, it appeared that the effects of spatial ability on 3D learning did not appear overall or gender-wise in our study, as both genders performed similarly overall after the 3D intervention. However, the gender differences seemed to have remained latent in the deep structures. This persistence of spatial ability 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 can be neutralized at all levels. This may also be 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, these softwares still need a lot of work and improvement as they can sometimes glitch and are 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 softwares and their features are essential to increase student interest. While many students liked the 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 the software will lead to decreased mental resistance and better engagement.

The currently 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. However, 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. However, they still need more realism, 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 through 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. 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

There was no control group or randomization involved in the study as it was against our institution's educational and ethical standards. Because of the small sample size, the results may not be generalizable to other populations. Another limitation of the pre-post interventional method may be the learning and test effects of the pre-intervention on the post-interventional counterparts. But 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 may be more due to the immediate effects of the intervention and less due to generalization of knowledge.

Regarding time, it was mentioned by the students that there was not enough time in the course. This can make the experience feel more instantaneous creating the illusion that both the 2D and 3D interventions occurred at the same time. Even if we include the effect of time and repetition, we can still say that 3D interventions greatly accelerated the results, as time itself was "not enough". Moreover, the presence of large (and medium) effect sizes cannot be ignored just due to a pre-post design. While experimentally, researchers may like to compare 2D vs. 3D (vs. Cadaveric labs), the students rather have all modalities at their disposal as a multidimensional approach, 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 in our study, and generalization of knowledge could have been more adequately tested 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 involved the creation of a new 3D learning module and its evaluation using the newly developed RRR and Confidence in topics surveys and tests, and 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. These levels of effect sizes mitigates the issue of small sample sizes (Norman, 2010).

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 3D is the only option left as we move to a cadaver-less era. 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. Previous studies have indicated that females show a greater amount of Neuro-anxiety and Neurophobia but 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 the idea of social need and justice in neuroanatomy education, but also points to a possibility that the observed (near) equalities may not be accurate and that the gender differences hide latent 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, and limiting 3D technologies to be used only for complex structures, can be a more efficient approach. Finally, testing on other populations across disciplines, countries, and cultures may prove its external validity.

Conclusions

This study provides some 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, and showed high levels of internal consistency and external validity, proving to be a valid measure. It can help to differentiate between the affective and cognitive benefits, and 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.

This study adds to the literature on using 3D resources for neuroanatomy education and provides scientific evidence for its benefits. In our study, the students appreciated 3D software due to the better visuals and colorful graphics as 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 motivation and learning, and was 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 under-appreciated the effectiveness of the 2D resources as compared to 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. The motivation benefits of 3D over 2D intervention appeared to be more apparent in females, but this was not reflected on the Confidence in topics surveys or tests and thus could not be reconfirmed, perhaps due to more effort put in by the female students.

The 3D intervention enhanced students' exam performance on both 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. 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. It 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. This could not be reconfirmed as similar effects were not observed in the superficial and deep components of the Confidence in topics survey. On the contrary, the survey's findings were opposite, suggesting a possibility that the 3D intervention may have made them realize a greater difficulty gap between superficial and deep structures. This could be due to an insight (the pre-intervention survey) that did not correlate well 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, the analysis of the superficial and deep components of the test scores showed 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 after 3D, suggesting that they may have gained more on the deeper structures while bridging the difficulty gap as 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 but not 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 well on their test performance after studying the 2D resources, they did correlate after the 3D intervention indicating the development of a better insight by the 3D intervention. The benefits of the 3D intervention on Confidence and performance were greatly experienced by all students equally (as those scores did not actually 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 reported effectiveness of 3D resources, which was less than those reported by students with no prior experience. In contrast, no such effect of previous experience was observed on the scores of the 2D intervention. This 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. Nonetheless, 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 actually after 3D intervention, 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. It shows the 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 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. Enhanced visualization helps in better engagement and proves helpful to students who nowadays have less preference for typical 2D textbook methods.

Summary:

The study validates the effectiveness of a newly developed 13-item instrument to measure Reaction, Relevance and Results (RRR) of a 3D intervention in neuroanatomy education, and also a 5-item Confidence in Topics survey. 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 have an incomplete insight into their academic needs and progress, or if they have some bias due to previous experience. Hence, using surveys and test scores together 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.

It was found that 3D visualizations significantly improved learning outcomes, especially in deep structures, and bridged the gap between superficial and deep anatomical knowledge. While males and females may initially experience the benefits in slightly different ways (as self-reported on surveys), 3D software does benefit them almost equally (as seen via performance on tests). Still the male students benefited slightly more in learning deeper structures after the 3D intervention. Despite some entry barriers and the need for software improvement, 3D resources were highly appreciated and are predicted to become essential in future anatomical education, complementing traditional dissection methods. Previous experience in neuroanatomy slightly reduced the perceived effectiveness of 3D tools, but no major group or experience-related differences were found in test performance. The study highlights that 3D technology can enhance spatial understanding but requires refinement and time for optimal integration into curricula.

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 JS. Anxiety and self-efficacy related to learning neuroanatomy in an integrated medical curriculum [dissertation]. Louisville, KY: University of Louisville; 2021. doi: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.

Gilbert J, Kakonge R, Reiner M, Nakhleh M. Visualization: Theory and Practice in Science Education. Dordrecht, Netherlands: Springer; 2008. doi:10.1007/978-1-4020-5267-5

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. The One World School House: Education Reimagined. London, UK: Hodder and Stoughton; 2012.

Khare A. 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 [dissertation]. Louisville, KY: University of Louisville; 2022. doi:10.18297/etd/3998

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. Death knell for the lecture: technology as a passport to personalized education. The New York Times. December 6, 2011. Accessed August 11, 2025. http://www.nytimes.com/2011/12/06/science/daphne-koller-technology-as-a-passport-to-personalized-education.html

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 F. Learning from graphically integrated 2D and 3D representations improves retention of neuroanatomy [dissertation]. Louisville, KY: University of Louisville; 2012. doi:10.18297/etd/1034

National Research Council. Learning to Think Spatially: GIS as a Support System in the K-12 Curriculum. 1st ed. Washington, DC: National Academies Press; 2006.

National Science Foundation, National Center for Science and Engineering Statistics. Women, minorities, and persons with disabilities in science and engineering: 2017. Special Report NSF 17-310. Arlington, VA: National Science Foundation; 2017. Accessed August 11, 2025. 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 and 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. Men do not rule the World of Tanks: negating the gender-performance gap in a spatial-action game by controlling for time played. Am Behav Sci. 2020;64(7):1031-1043. doi:10.1177/0002764220919147

Risko EF, Anderson N, Sarwal A, Engelhardt M, Kingstone A. Everyday attention: variation in mind wandering and memory in a lecture. Appl Cogn Psychol. 2012;26:234-242. doi:10.1002/acp.1814

Risko EF, Buchanan D, Medimorec S, Kingstone A. Every attention: mind wandering and computer use during lectures. Comput Educ. 2013;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. Meta-Analytic Procedures for Social Research. Thousand Oaks, CA: Sage Publications; 1991. doi: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 MK, Thomas JA, Lan WY, Cooper S, Ahern TC, Shaw SM, et al. Teaching courses online: a review of the research. Rev Educ Res. 2006;76(1):93-135. doi:10.3102/00346543076001093

Tomczak M, Tomczak E. The need to report effect size estimates revisited: an overview of some recommended measures of effect size. Trends Sport Sci. 2014;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.