3D Learning in anatomical and surgical education in relation to visual-spatial abilities

Bogomolova, K.

Citation

Bogomolova, K. (2022, February 3). *3D Learning in anatomical and surgical education in relation to visual-spatial abilities*. Retrieved from https://hdl.handle.net/1887/3274191

7

Anatomy dissection course improves the initially lower levels of visual-spatial abilities of medical undergraduates:

a case-control study

Katerina Bogomolova, Beerend P. Hierck, Jos A. van der Hage, Steven E.R. Hovius

Anatomical Sciences Education 2019; 13:333-342

ABSTRACT

Background

Visual-spatial abilities (VSA) are considered a successful predictor in anatomy learning. Previous research suggests that VSA can be trained, and the magnitude of improvement can be affected by initial levels of spatial skills. This case-control study aimed to evaluate (1) the impact of an extra-curricular anatomy dissection course on VSA of medical undergraduates and (2) the magnitude of improvement in students with initially lower levels of VSA, and (3) whether the choice for the course was related to VSA.

Methods

Course participants (n = 45) and controls (n = 65) were first and second-year medical undergraduates who performed a Mental Rotations Test (MRT) before and 10 weeks after the course.

Results

At baseline, there was no significant difference in MRT scores between course participants and controls. At the end of the course, participants achieved a greater improvement than controls (first-year: Δ6.0 ± 4.1 vs. Δ4.9 ± 3.2; ANCOVA, p = .019, Cohen's d = 0.41; second-year: Δ6.5 ± 3.3 vs. Δ6.1 ± 4.0; p = .03, Cohen's d = 0.11). Individuals with initially lower scores on the MRT pretest showed the largest improvement (Δ8.4 ± 2.3 vs. Δ6.8 ± 2.8; p = .011, Cohen's d = 0.61).

Conclusions

In summary, (1) an anatomy dissection course improved VSA of medical undergraduates; (2) a substantial improvement was observed in individuals with initially lower scores on MRT indicating a different trajectory of improvement; (3) students' preferences for attending extracurricular anatomy dissection course was not driven by VSA.

INTRODUCTION

Anatomical education is constantly under pressure despite it being considered as one of the cornerstones of medical curricula. Teaching hours of anatomy have been decreasing over time since the shift towards an integrated curriculum.[1-3] Additionally, ethical reasons, the high costs and limited availability of cadavers, and the increased time pressure on curricula have led to a decreased exposure to traditional cadaveric dissections. [2,4-7] Although, its educational value is under debate, dissection classes are found to be highly valuable by medical undergraduates, regardless of their sex, academic background, or citizenship.[6,8] In their opinion, dissections deepen their understanding of anatomical structures and their spatial relations, make learning interesting and are preferred over any other educational approach, especially in the first year of the medical program.[6] Today, medical undergraduates learn the anatomy mostly from two-dimensional (2D) representations of structures in anatomical atlases and textbooks and, consequently, experience difficulties to translate the acquired 2D knowledge into practice. [9-12]

Visual-spatial abilities and performance in anatomy

How well acquired 2D anatomical knowledge is translated into practice depends largely on the visual-spatial abilities (VSA) of students. In the medical anatomical context, it is defined as the ability that allows students to construct visual-spatial, e.g., 3D, mental representations of 2D images and to mentally manipulate these representations[13,14]. The first studies evaluating the association between VSA and anatomy learning have been performed by Rochford[15] and Garg and colleagues.[16-18] In these studies, VSA have significantly affected the learning process of spatial anatomy regardless of age, sex, right handedness, or computer use. Since then, even more research has been conducted to explore this association. The first comprehensive review of studies has been performed by Langlois and colleagues.[19] Their meta-analysis has revealed a predictive value of VSA when anatomy is assessed using spatial methods such as practical examination, 3D synthesis from two-dimensional views, and drawing of views and cross-sections. As such, VSA are considered a successful predictor in anatomy learning and assessment. [19,20] In health care professions VSA are also a successful predictor in the acquisition of surgical technical skills, especially in the early stages of learning.[21,22] For instance, Wanzel and colleagues have evaluated the correlation between VSA and surgical performance of dental students, surgical residents and staff surgeons in performing a spatially complex surgical procedure.[23] VSA scores were correlated with surgical performance only within the group of dental students, suggesting that practice and surgical experience may supplant the influence of VSA over time. The effect of VSA on performance has also been demonstrated in mathematics[24], veterinary education[25] and dental education.[26]

VSA as a selection tool

It is not surprising that VSA have been recommended to be used not only in the training, but also in the selection of surgical residents.[27] A high motivation for the surgical specialty would apparently not be enough since it does not imply higher VSA among candidates. Langlois and colleagues have evaluated a cohort of 210 medical graduates and did not find any relation between VSA and the choice of residency program.[27] Nor did the choice for an elective course of applied anatomy depend on the VSA of medical graduates.[28] However, the relation between VSA and a high interest in anatomy, in the very early stages of a medical career, has not yet been evaluated.

Malleability of VSA

On the other hand, several studies have suggested that VSA can be trained through practice and experience. In a meta-analysis, Langlois and colleagues have found evidence for improvement of spatial abilities in anatomy education using instruction in anatomy and mental rotation training.[29] For instance, in a single group study, Lufler and colleagues have reported an improvement of VSA of first-year medical undergraduates after participation in a gross anatomy course consisting of six dissection sessions.[30] In a similar study with a control group of educational sciences students, VSA have increased after participation in the course consisting of lectures, self-study assignments including computer-assisted learning (CAL), collaborative learning, laboratory with prosected specimens, and body painting.[31] When an anatomy course was combined with a training of mental rotation skills unrelated to anatomy, an even higher increase in VSA scores has been observed.[32] These were the only two studies to date that have included the practice effect on VSA test scores in a control group resulting in a pooled treatment effect of 0.47 (95% CI [-0.03; 0.97]). The pooled treatment effect of single-group studies included in the meta-analysis was 0.49 (95% CI [0.17; 0.82], n = 11).

Furthermore, the improvement appears to be present on an expert level.[33] It has been found that expert clinical anatomists were better in performing metric spatial tasks than novices, suggesting that VSA are trained by practice and education. In addition, the dose-dependent effect of practice and learning on VSA has been found in medical undergraduates after attending CAL courses of musculoskeletal and cardiovascular anatomy.[34]

The malleability of VSA has been demonstrated in other disciplines as well, such as science, technology, engineering, and mathematics (STEM), and veterinary medicine.[35,36] In the meta-analysis of Uttal and colleagues[35], VSA were classified as an intrinsic and dynamic spatial skill and were significantly affected by training with an overall effect size of 0.49 ($p < .01$).

Sex differences and initial level of performance

Sex differences in VSA have been repeatedly reported in the literature. At baseline, males have often achieved higher scores in VSA tests than females. [34,35,37,38] This difference has been particularly observed in measures of mental rotation.[39,40] However, as has been demonstrated by several studies and meta-analyses, both males and females can achieve comparable magnitude of improvement after training.[35,38,41]

Another aspect worthy to mention is the initial level of performance of individuals in VSA training. A meta-analysis of 187 studies using a screening procedure to identify initially low-performing students has reported significantly larger effect of training when compared to studies enrolling all participants regardless of initial performance levels.[35] These finding suggests that low-performing students can achieve a larger magnitude of improvement than high-performing student. Additionally, students and residents with lower VSA in a surgical field have been able to achieve required levels of knowledge and skills through suitable teaching methods and guidance.[23,30,42,43] Therefore, it might be valuable to consider VSA abilities as a tool to identify learners who will benefit most from extra practice and new learning environments instead of an absolute selection criterium to guide selection of candidates for surgical training programs.[44]

The Erasmus Medical Center Anatomy Research Project

The Erasmus MC Anatomy Research Project (EARP) is an extracurricular anatomy dissection course at the faculty of Medicine, Erasmus University Medical Center Rotterdam, The Netherlands. The EARP was set up in 2003 in response to reduced teaching volume of anatomy and a limited exposure to dissections. Since then, the course has become a unique and fully autonomous peer-to-peer educational model. The extracurricular course is organized annually during a period of ten weeks. It takes place in the evening hours and does not interfere with the regular medical program. All medical undergraduates, from year one to year six of the undergraduate program, are invited to apply for one of the four parallel programs, each covering a different anatomical region: Thorax (for the first-year students), Abdomen (for the second-year students), Head & Neck and Urogenital System (for the third-, fourth-, fifth- and sixth-year students), and Extremities (for the third-, fourth-, fifth- and sixth-year students). Due to a limited capacity, e.g., six available cadavers, a maximum of one hundred students are admitted annually, 24 students to Thorax, Abdomen and Extremities programs and 32 students to Head & Neck and Urogenital System program (Figure 1). Students must apply with a written assignment, e.g., about solving a clinical anatomy case. Selection of students is based on the highest scoring assignments and performed blindly by the EARP committee. After enrollment, students attend an instructional lecture and receive the EARP handbook with guidelines and detailed explanation of dissection of the assigned anatomical region including text and images. Subsequently, students start to work towards a complete dissection of the anatomical region on the assigned cadaver in a group of four students for eight weeks. Two students dissect the left part of the region, while the other two students dissect the right part of the region, which ensures equally active involvement of all students. Eventually, the same cadaver is used by four groups of two students each week, each group working on a different anatomical region on a different day of the week.

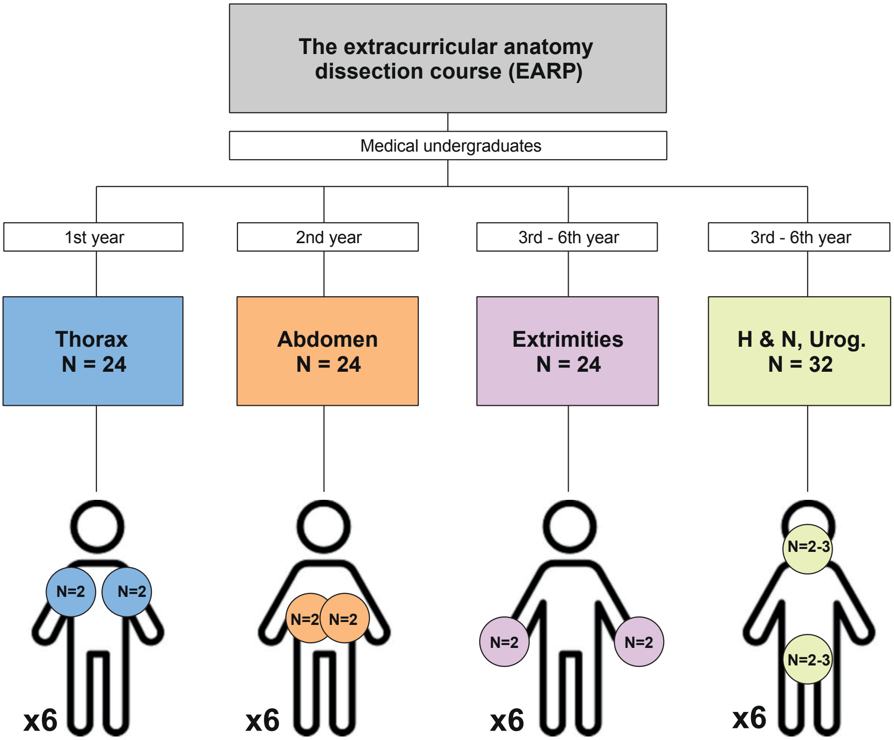

Figure 1. An extra-curricular anatomy dissection course. Students attend eight dissection sessions of three hours each week. EARP, the Erasmus MC Anatomy Research Project; H & N, Urog., Head & Neck and Urogenital System.

Dissection sessions are supervised by two tutors and four mentors who are senior medical undergraduates who previously participated in EARP. To ensure the quality of supervision and optimal knowledge of anatomy and dissection, all tutors and mentors attend a training program which also includes a dissection of the assigned anatomical regions. The EARP program includes 20-24 hours of dissections, 3-5 hours interactive lectures and demonstration sessions given by medical specialists and an hour of practical and written examinations. The latter is composed of questions assessing factual knowledge (e.g., naming a muscle's origin or insertion, innervation, and vascularization), spatial knowledge (e.g., the course of nerves and vessels in relation to other structures) and clinical decision making. The practical examination is composed of two parts: identification of as many structures as possible on a specimen for three minutes and naming of pin-pointed structures marked on a specimen.

Objectives and aims

In the Netherlands, EARP has been established as a unique peer-to-peer educational setting in which spatial anatomy is learned hands-on during cadaveric dissections outside the regular medical program. This setting provides a unique opportunity to evaluate to what extent cadaveric dissection has its effect on VSA of medical undergraduates when compared to a control group consisting of non-participating medical undergraduates at the same stages of their curricula. In addition, it allows to evaluate a possible relation between having a high interest for anatomy and the VSA of students in the early phases of their medical careers.

Therefore, the aim of this study was to evaluate the impact of anatomy dissection course on VSA of medical undergraduates, and the magnitude of improvement in individuals with initially lower levels of VSA. Additionally, the present study aimed to evaluate whether the choice to apply for an extracurricular anatomy dissection course was related to the VSA of students. The authors hypothesized that individuals with higher levels of VSA are more likely to apply, that VSA will improve after an anatomy dissection course, and the improvement will be larger in individuals with initially lower levels of VSA.

MATERIALS AND METHODS

Study design

A prospective case control-study was carried out at the Erasmus University Medical Center Rotterdam, the Netherlands. In general, a case-control study is efficient in evaluating associations between rare exposures and outcomes (Song and Chung, 2010). Since only 24 out of 400 students from each academic year participate in the EARP program, this study design was most suitable to answer the research questions. The study was approved by the course coordinator and the director of medical educational program and was considered exempt from formal assessment by the local ethical assessment committee (METC) of Erasmus University Medical Center Rotterdam (case number: CME-2019-0077).

Participants

Cases were defined as first-year and second-year medical undergraduates who were admitted to the EARP Thorax and Abdomen programs, respectively. Course participants, or cases, were identified through the attendance list of the programs. Controls were defined as first-year and second-year medical undergraduates who did not apply for the course and were matched for academic year and sex. Students who did apply for the course, but were not selected, were excluded. For each course participant, a maximum of two controls were identified and approached during the regular lectures at the faculty with the request to participate. A 1 case : 2 control ratio was chosen since little is gained in terms of statistical power by including more than two controls for each case.[59]

Measurement of VSA

VSA were assessed by the Mental Rotations Test (MRT), previously validated by Vandenberg and Kuse[45] which was based on rotated blocks of Shepard and Metzler[46] and redrawn by Peters and colleagues.[47] This psychometric test is widely used in the assessment of VSA and has repeatedly showed a positive association with anatomy learning and assessment.[19] The test consists of a standard set of 24 items. Within each item, a three-dimensional figure is presented as a 2D drawing with four possible rotated versions of that figure. Subjects must make a mental three-dimensional representation and rotation of the figure to identify the two correct options. One point per item was awarded if both selected options were identified correctly. The maximum score on this test was 24 points.

A testing effect has been previously reported after repeated administration of the MRT.[31,32,38,48] In an attempt to minimize the testing effect, two versions of the MRT were used. The MRT, used as a pretest, included the original set of 24 items. In the MRT, used as a posttest, the same 24 items were rearranged in a different random order.

Procedures

Participation was voluntary, and an informed written consent was obtained by all participants before study. A short pre-questionnaire was used to gather information on age, sex, participation in EARP Thorax program in the first year (only applicable for second-year students) and prior or current participation in an academic program other than Medicine. A paper-and-pencil MRT pretest was administrated to course participants prior to the start of their first dissecting session. The MRT posttest test was performed after ten weeks on the day of their examination. Controls simultaneously completed the MRT pretest and posttest in a lecture hall. All students were given ten minutes to complete the test without a break.

Statistical Analysis

Descriptive statistics were used to summarize participants' baseline characteristics. Discrete variables were described as absolute frequencies (N) and percentages (%), and continuous variables as mean and standard deviation (SD). The differences in baseline characteristics were assessed with Chi-squared test for differences in proportions and independent *t*-test for differences in means. The MRT scores were measured on a continuous scale and reported in terms of means and standard deviations. The differences in MRT pretest scores between course participants and controls were assessed with an independent t-test for normal distributions and Mann-Whitney test for non-parametric distributions. The differences in mean improvement in MRT scores (ΔMRT) were assessed with a one-way ANCOVA. The mean improvement was included as a dependent variable, the EARP participation as a fixed factor and the absolute MRT pre-test score as a covariate. All analyses were adjusted for age, sex, participation in EARP Thorax program in the first year (only applicable for second-year students) and prior or current participation in an academic program other than Medicine. Additionally, the analysis was repeated for MRT-low (individuals who scored below the mean on the MRT pretest) and MRT-high (individuals who scored above the mean on the MRT pretest) groups separately with adjustment for academic year. Correlation between MRT pretest scores and mean improvement was assessed with Pearson correlation coefficient. The effect size (Cohen's d) of the differences in MRT improvement between groups was calculated using the mean scores and standard deviations of both groups.[49] All analyses were performed using SPSS statistical software package version 23.0 for Windows (IBM Corp., Armonk, NY). Statistical significance was determined at the level of $p < .05$.

RESULTS

All course participants enrolled in the EARP Thorax and Abdomen programs participated in the study. For the 24 EARP Thorax participants a total of 44 controls were identified. For the 24 EARP Abdomen participants a total of 22 controls were identified. Four subjects were excluded from the analysis due to the following reasons: one participant selected only one correct option in the MRT pretest instead of two; two course participants did not complete the MRT posttest due to their absence on the EARP examination day; one control was a significant outlier and was removed from the analysis since a significant outlier violates one of the required assumptions for performing a one-way ANCOVA and may reduce the validity of results.

Baseline characteristics

No significant difference was found between course participants and controls in terms of age, sex, and participation in an academic program other than Medicine (Table 1).

Table 1. Baseline characteristics of study participants.

	Course participants	Controls	p value
First year	**n = 22**	**n = 43**	
Age, mean ± SD	19.3 ± 1.4	18.7 ± 1.0	.587
Sex			
Male, n (%)	8 (27.3)	9 (20.9)	.569
Female, n (%)	16 (72.7)	34 (79.1)	
Participation in an academic program other than Medicine, n (%)	4 (18.2)	2 (4.7)	.084
Second year	**n = 23**	**n = 22**	
Age, mean ± SD	19.4 ± 4.4	19.8 ± 1.1	.546
Sex			
Male, n (%)	6 (26.1)	5 (22.7)	.799
Female, n (%)	17 (73.9)	17 (77.3)	
Participation in an academic program other than Medicine, n (%)	1 (4.3)	4 (18.2)	.187
Participated in EARP Thorax program in the first year, n (%)	9 (39.1)	0 (0)	.001

n, number of students; SD, standard deviation; EARP, Erasmus MC Anatomy Research Project.

The observed high ratio of females in both groups represents the average ratio of males and females in the current undergraduate medical curriculum in the Netherlands, which is approximately 30%:70%. The only significant difference was observed among second-year students in numbers of students who participated in the EARP Thorax program in the first year (nine students in the course participant group versus zero students in the control group (p = .001). The MRT pretest scores of these nine students were not significantly different from the scores of the other fourteen course participants (10.9 ± 4.3 vs. 12.3 ± 6.4, p = .272).

Improvement in Mental Rotations Test scores

As shown in Figure 2, no significant difference in MRT pretest scores was found between the course participants and controls (*first-year:* 14.6 ± 5.5 vs. 13.8 ± 5.9; p = .411; *second-year:* 11.8 ± 5.1 vs. 11.5 ± 5.2; p = .856). After ten weeks, the MRT scores were significantly improved in both groups. However, the mean improvement (ΔMRT) among course participants was significantly higher than among controls (*first-year:* Δ6.0 ± 4.1 vs. Δ4.9 ± 3.2; $F_{(1,56)}$ = 5.8, p = .019, Cohen's *d* = 0.31; *second-year:* Δ6.5 ± 3.3 vs. Δ6.1 ± 4.0; $F_{(1,36)}$ = 2.7, p = .03, Cohen's *d* = 0.11) (Figure 2). Higher MRT pretest scores were associated with less improvement in both academic years (*first-year:* β = -0.9; 95% CI [-1.3; -0.3], p = .0001; *second year:* β = -0.3; 95% CI [-0.52; -0.14], p = .001). Additionally, among second-year students, previous participation in EARP was negatively associated with the mean improvement in MRT scores (β = -3.9; 95% CI [-1.16; -6.68], p = .07). Sex, age and participation in an academic program other than Medicine were not significantly associated with the improvement._

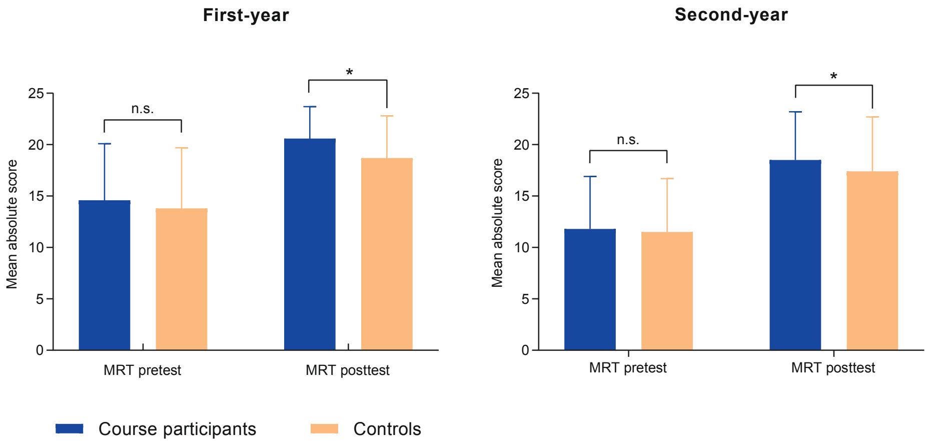

Figure 2. Differences in performance on MRT pretest and posttest between course participants (*first year*: EARP Thorax program, *second year*: EARP Abdomen program) and controls. Performances are reported in mean scores. Error bars represent standard deviation; *p < .05; n.s., not significant; MRT, Mental Rotations Test; EARP, Erasmus MC Anatomy Research Project.

Improvement in students with initially lower Mental Rotations Test scores

As shown in Figure 3, when the analysis was repeated for individuals who scored below and above average on the MRT pretest (e.g., MRT-low and MRT-high groups) separately, the improvement in MRT scores was only present in the MRT-low group with a much larger effect size (*MRT-low group*: Δ8.4 ± 2.3 vs. Δ6.8 ± 2.8; $F_{(1,50)}$ = 6.916, p = .011, Cohen's *d* = 0.61; *MRT-high group:* Δ3.8 ± 3.3 vs. Δ3.6 ± 2.7; $F_{(1,45)}$ = 1.253, p = .269, Cohen's *d* = 0.06).

Additionally, as shown in Figure 4, the negative association between MRT pretest scores and mean improvement in MRT scores was no longer present. Instead, course participants in the MRT-low group showed a positive correlation between MRT pretest scores and mean improvement (r = 0.350, p = .093). In the MRT-high group, however, around 55% (R^2 = 0.55) of the total variation in MRT posttest scores could be explained by the MRT pretest scores. There was a moderate negative correlation between mean improvement and MRT pretest scores in course participants (r = -0.68, p = .001) and controls (r = -0.76, p = .001).

Sex differences

Males significantly outperformed females on the MRT pretest (15.2 ± 5.8 vs. 12.5 ± 5.3, p = .034) and on the MRT posttest (20.2 ± 4.1 vs. 18.4 ± 4.2; p = .038). However, there was no significant difference in the mean improvement (ΔMRT) between males and females (Δ5.04 ± 4.0 vs. Δ5.9 ± 3.2; $F_{(1,100)}$ = 0.371, p = .962). Additionally, the percentage of females in the MTR-low group did not differ significantly from the percentage in the MRT-high group (84.2% vs. 70.0%, χ^2 = 3.091, p = .079).

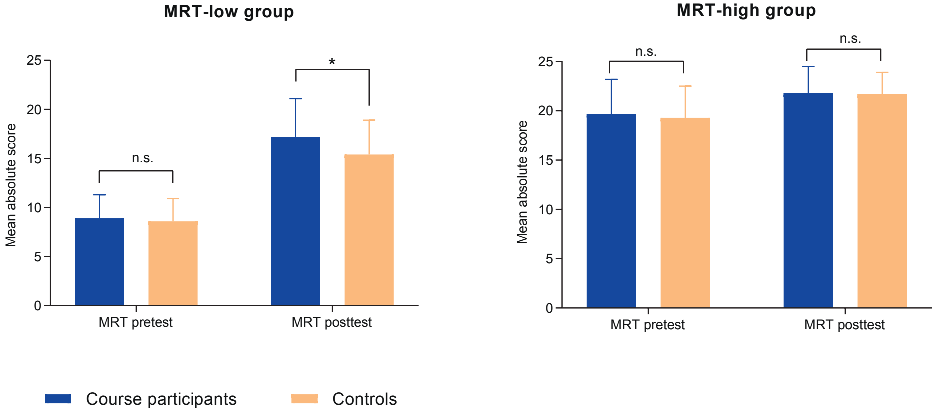

Figure 3. Differences in performance on MRT pretest and posttest between course participants and controls in the MRT-low and MRT-high groups of first- and second-year medical undergraduates. Performances are reported in mean scores. Error bars represent standard deviation; *p < .05; n.s., not significant; MRT, Mental Rotations Test.

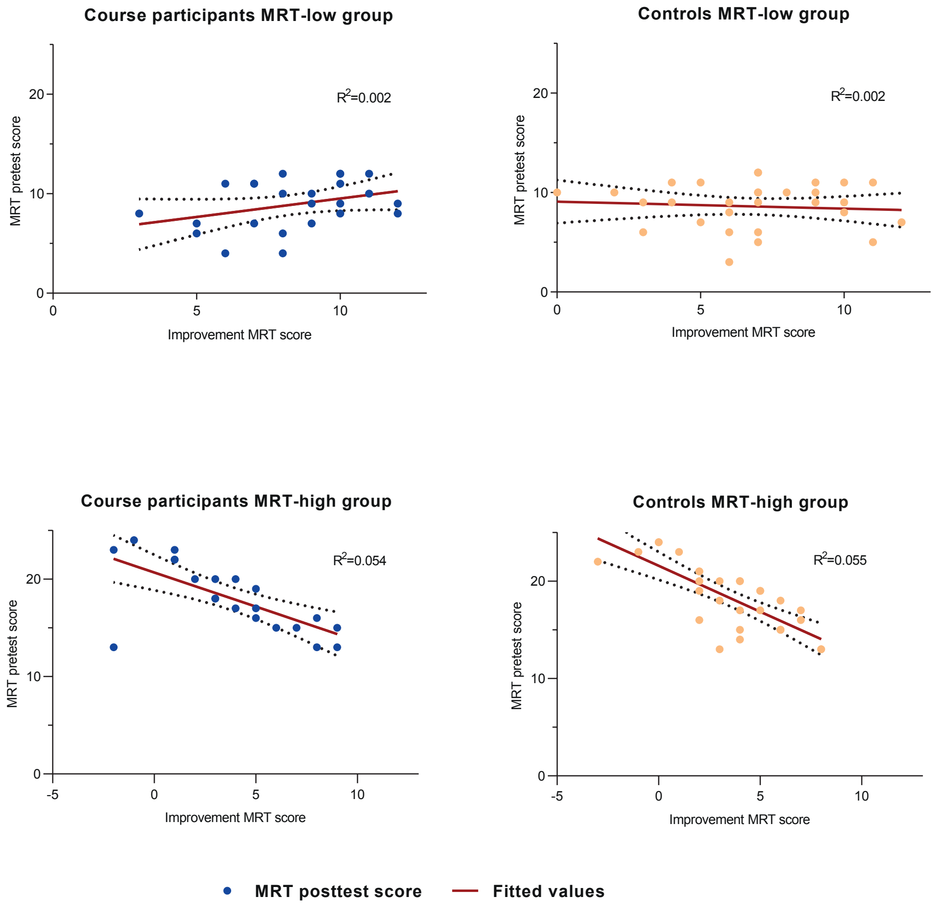

Figure 4. Relationship between MRT pretest scores and mean improvement in the MRT scores. A regression analysis graph illustrating the relationship between initially low and high levels of visual-spatial abilities and mean improvement among course participants and controls. MRT, Mental Rotations Test.

DISCUSSION

This case-control study was performed to evaluate the impact of an extra-curricular anatomy dissection course on VSA of medical undergraduates and to evaluate whether the choice for this course was related to the initial level of their VSA. Furthermore, a control group composed of medical undergraduates was included which enhances the internal and external validity of the results. The study resulted in the following findings and observations.

Firstly, the results of this experimental study showed a significant improvement of VSA, as measured by the Mental Rotations Test, after completing eight sessions of cadaveric dissections. The observed effect sizes (*first-year:* $d = 0.31$; *second-year*: $d = 0.11$) indicate that a repeated practice of dissection had a small to medium effect on VSA of students. This effect was much smaller than the one observed by Lufler and colleagues[30] ($d = 1.02$) after a dissection course when no control group was included.[29] The difference in effect sizes can be attributed to the testing effect. This effect occurs after repeated administration of the MRT which provides students the chance to train their spatial skills by doing the test.[26,47,50] This practice effect in the current study was reflected by a significant improvement of the MRT scores among controls. A similar effect has also been observed by Vorstensbosch and colleagues by including a control group composed of students of educational sciences.[31] A control group is, therefore, essential for the assessment of the course related improvement. Additionally, a control group composed of the identical source population, e.g., medical undergraduates from the same academic years, eliminate possible differences in baseline characteristics, such as high school profiles, intellectual interest, and hobbies, that can influence VSA.

The improvement in VSA scores may be attributed to active involvement in dissection of a 3D cadaver accompanied with studying 2D representations from the EARP handbook. Additionally, students are constantly challenged by mental visualization of anatomical structures and understanding of their spatial relations to perform the dissection in the best and most efficient way. Further research is needed to determine which components of cadaveric dissection contribute most to the improvement in VSA, and to what extent this effect will remain present. In the current study, nine second-year course participants, participated in the EARP Thorax program in a previous year. They did not perform better on the MRT pretest test than the fourteen course participants, who participated in the EARP course for the first time. This may suggest that the acquired level of VSA might not be long lasting. However, this sample was too small to draw that conclusion.

Secondly, when the results were analyzed for individuals with initially lower MRT pretest scores only, a much larger effect size ($d = 0.61$) was observed. Individuals with initially higher MRT pretest scores did not show any improvement. Instead, the MRT pretest scores were negatively associated with the mean improvement. These findings may reflect an aptitude-treatment effect of VSA, i.e., that low performing individuals are having a different trajectory of improvement than high performing individuals.[51] That VSA may cause an aptitude-treatment interaction has been illustrated earlier by Cui and colleagues.[42] After learning with monoscopic 3D images, students with lower VSA scores performed significantly worse than students with higher VSA scores. While after learning with a stereoscopic 3D model, these students performed significantly better and equally well as students with higher VSA scores. Similar effects have been reported by Garg and colleagues where students with lower VSA had significant disadvantages by learning anatomy with multiple view presentations, while students with higher VSA performed better with these types of presentations.[17,18] The observed phenomenon in this study, however, may also be attributed to the ceiling effect in the MRT test. This effect is addressed further in the limitation section.

Thirdly, the choice for an extracurricular anatomy dissection course, in this study, did not imply higher levels of VSA of medical undergraduates. These findings support previous research on VSA and the choice for an elective course of applied anatomy or personal preference for a surgical specialty.[27,28] In both situations, personal preferences and choices among postgraduates were not reported to be associated with the individual VSA. It is interesting to note that the choice for medical careers in the first place may imply higher VSA among medical undergraduates. When compared to students of educational sciences, medical undergraduates had higher mean VSA.[31] Similar differences were observed between dental and psychology students.[26]

Lastly, the observed sex differences in this study were in line with previous research. Despite having initially lower scores on the MRT test, females were still able to achieve similar magnitudes of improvement as males after training.[34,35,37,38,41] Additionally, in this study, the percentage of females in the MRT-low group was not significantly lower than in the MRT-high group, as could be expected. These findings suggest that the individual approach is preferable since a particular male may have lower VSA than a particular female.

Future directions

The findings of this study underline the importance of anatomical education in the light of VSA training. The positive effect of anatomical education on VSA, which in turn facilitates learning and retention of anatomical knowledge, indicates that these two can reinforce one another. Additionally, Roach and colleagues have demonstrated that an early guidance and instruction can improve low performing students' strategies for spatial problem solving.[43,52,53] This can be of a great importance for low performing individuals and have implications for individualized approaches in the current curricula.

The role of augmented and virtual reality in anatomical education is promising and is currently addressed in ever more research. In the fields of engineering and technology, research has shown that training in augmented and virtual reality can improve various components of spatial abilities, such as visualization, rotation, and orientation.[26,32,48,54] Stereoscopic three-dimensional visualization technologies may, therefore, serve as valuable additional tools to include spatial reasoning training in an anatomical context next to traditional ways of learning.[55]

Limitations of the study

A case control study is relatively quick and efficient in evaluating associations between rare exposures and outcomes.[56] Since only 24 out of 400 students participate in the extracurricular anatomy dissection course in each academic year, the numbers of participants were restricted. Since this study design requires comparatively few subjects it allowed to omit recruitment of the entire first and second-year cohorts. However, a desirable 1:2 case:control ratio among second-year students was not achieved. To underline the validity of 1:1 case:control ratio in the main analysis, a post-hoc analysis among first-year students was performed. After random elimination of half of the controls and repeated analysis, a significant difference between course participants and controls remained (20.6 vs. 18.5; ANCOVA, F = 4.8, p = .034; Cohen's *d* =0.09). Consequently, the recruited number of controls among second-year students was justifiable for the main analysis.

Certainly, a case control study is susceptible to particular types of biases. Since no randomization was possible in this setting, a selection bias should be taken into account despite of the recruitment of the controls from the identical source population. Controls were recruited in the lecture hall and only part of them was willing to participate in the study. They could have been less motivated to do their best on the MRT test than the course participants who were usually highly motivated and were more willing to perform best on such a test. This could have partially accounted for the less improvement in the MRT scores among controls. Other possible confounders, which were not included, were gaming experience and performance on anatomy in the current curriculum. Both have been associated with a better performance in VSA tests before.[35,57]

The MRT pretest scores were negatively associated with the mean improvement, especially in the MRT-high group. This association may reflect an aptitude-treatment interaction, but a ceiling effect cannot be ruled out. Ceiling effect occurs when more than 15% of the participants reach the highest possible scores of a test.[58] In this study, 13% of the participants reached the highest possible score of 24 points. Therefore, a ceiling effect was not likely but cannot be ruled out completely. To avoid a possible ceiling effect in the future, a more difficult set of items in the MRT could be used allowing high performing students achieve a much greater improvement. The association could also be attributed the statistical feature "regression to the mean", i.e., since high performing students structurally score higher on the pretest, they are more likely to score lower on a repeated test.

CONCLUSIONS

This study showed that the VSA scores of medical undergraduates improved after anatomy dissection. Additionally, a substantial improvement was observed in individuals with initially lower scores on the VSA test. Although a ceiling effect cannot be completely ruled out, this can be indicative of a different trajectory of improvement between individuals in this particular study. This possible aptitude-treatment effect will need to be evaluated in further research and an individualized approach in current curricula could be considered. Finally, the students' preferences for attending the extracurricular anatomy dissection course were not driven by VSA.

REFERENCES

1. Drake RL, Lowrie DJ, Prewitt CM. Survey of gross anatomy, microscopic anatomy, neuroscience, and embryology courses in medical school curricula in the United States. Anat Rec 2002;269:118–122.
2. Drake RL, McBride JM, Lachman N, Pawlina W. Medical education in the anatomical sciences: The winds of change continue to blow. Anat Sci Educ 2009;2:253–259.
3. Bergman EM, van der Vleuten CP, Scherpbier AJ. Why don't they know enough about anatomy? A narrative review. Med Teach 2011;33:403–409.
4. Pryde FR, Black SM. Anatomy in Scotland: 20 years of change. Scott Med J 50:96–98.
5. Waterston SW, Stewart IJ. 2005. Survey of clinicians' attitudes to the anatomical teaching and knowledge of medical students. Clin Anat 2005;18:380–384.
6. Azer SA, Eizenberg N. Do we need dissection in an integrated problem-based learning medical course? Perceptions of first- and second-year students. Surg Radiol Anat 2007;29:173–180.
7. Bergman EM, de Bruin AB, Herrler A, Verheijen IW, Scherpbier AJ, van der Vleuten CP. Students' perceptions of anatomy across the undergraduate problem-based learning medical curriculum: A phenomenographical study. BMC Med Educ 2013;13:152.
8. McLachlan JC, Bligh J, Bradley P, Searle J. Teaching anatomy without cadavers. Med Educ 2004;38:418–424.
9. McKeown PP, Heylings DJ, Stevenson M, McKelvey KJ, Nixon JR, McCluskey DR. The impact of curricular change on medical students' knowledge of anatomy. Med Educ 2003; 37:954–961.
10. Prince KJ, Scherpbier AJ, Van Mameren H, Drukker J, van der Vleuten CP. Do students have sufficient knowledge of clinical anatomy? Med Educ 2005;39:326–332.
11. Spielmann PM, Oliver CW. The carpal bones: A basic test of medical students and junior doctors' knowledge of anatomy. Surgeon 2005;3:257–259.
12. Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. How much anatomy is enough? Anat Sci Educ 2008;1:184–188.
13. Gordon HW. The cognitive laterality battery: Tests of specialized cognitive function. Int J Neurosci 1986;29:223–244.
14. Kozhevnikov M. Hegarty M. A dissociation between object manipulation spatial ability and spatial orientation ability. Mem Cognit 2001;29:745–756.
15. Rochford K. Spatial learning disabilities and underachievement among university anatomy students. Med Educ 1985;19:13–26.
16. Garg A, Norman GR, Spero L, Maheshwari P. Do virtual computer models hinder anatomy learning? Acad Med 1999;74:S87–S89.
17. Garg A, Norman G, Spero L, Taylor I. Learning anatomy: Do new computer models improve spatial understanding? Med Teach 1999;21:519–522.
18. Garg AX, Norman GR, Eva KW, Spero L, Sharan S. Is there any real virtue of virtual reality? The minor role of multiple orientations in learning anatomy from computers. Acad Med 2002;77:S97–S99.
19. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ 2017;10:235–241.
20. Yammine K, Violato C. A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. Anat Sci Educ 2015;8:525–538.

21. Maan ZN, Maan IN, Darzi AW, Aggarwal R. Systematic review of predictors of surgical performance. Br J Surg 2012;99:1610–1621.
22. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and technical skills performance in health care: a systematic review. Med Educ 2015;49:1065–1085.
23. Wanzel KR, Hamstra SJ, Caminiti MF, Anastakis DJ, Grober ED, Reznick RK. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. Surgery 2003;134:750–757.
24. Hegarty M, Kozhevnikov M. Types of visual–spatial representation and mathematical problem solving. J Educ Psychol 1999;91:684–689.
25. Provo J, Lamar C, Newby T. Using a cross-section to train veterinary students to visualize anatomical structures in three dimensions. J Res Sci Teach 2002;39:10–34.
26. Hegarty M, Keehner M, Khooshabeh P, Montello DR. How spatial abilities enhance, and are enhanced by, dental education. Learn Indiv Differ 2009;19:61–70.
27. Langlois J, Wells GA, Lecourtois M, Bergeron G, Yetisir E, Martin M. Spatial abilities of medical graduates and choice of residency programs. Anat Sci Educ 2015;8:111–119.
28. Langlois J, Wells GA, Lecourtois M, Bergeron G, Yetisir E, Martin M. Spatial abilities in an elective course of applied anatomy after a problem-based learning curriculum. Anat Sci Educ 2009;2:107–112.
29. Langlois J, Bellemare J, Toulouse J, Wells GA. Spatial abilities training in anatomy education: A systematic review. Anat Sci Educ 2020;13:71-79.
30. 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;5:3–9.
31. Vorstenbosch MA, Klaassen TP, Donders AR, Kooloos JG, Bolhuis SM, Laan RF. Learning anatomy enhances spatial ability. Anat Sci Educ 2013;6:257–262.
32. Hoyek N, Collet C, Rastello O, Fargier P, Thiriet P, Guillot A. Enhancement of mental rotation abilities and its effect on anatomy learning. Teach Learn Med 2009;21:201–206.
33. Fernandez R, Dror IE, Smith C. Spatial abilities of expert clinical anatomists: Comparison of abilities between novices, intermediates, and experts in anatomy. Anat Sci Educ 2011;4:1–8.
34. Guimarães B, Firmino-Machado J, Tsisar S, Viana B, Pinto-Sousa M, Vieira-Marques P, Cruz-Correia R, Ferreira MA. The role of anatomy computer-assisted learning on spatial abilities of medical students. Anat Sci Educ 2019;12:138–153.
35. Uttal DH, Meadow NG, Tipton E, Hand LL, Alden AR, Warren C, Newcombe NS. The malleability of spatial skills: A meta-analysis of training studies. Psychol Bull 2013;139:352–402.
36. Gutierrez JC, Chigerwe M, Ilkiw JE, Youngblood P, Holladay SD, Srivastava S. Spatial and visual reasoning: Do these abilities improve in first-year veterinary medical students exposed to an integrated curriculum? J Vet Med Educ 2017;44:669–675.
37. Langlois J, Wells GA, Lecourtois M, Bergeron G, Yetisir E, Martin M. Sex differences in spatial abilities of medical graduates entering residency programs. Anat Sci Educ 2013;6:368–375.
38. Nguyen N, Mulla A, Nelson AJ, Wilson TD. Visuospatial anatomy comprehension: The role of spatial visualization and problem-solving strategies. Anat Sci Educ 2014;7:280–288.
39. Linn MC, Petersen AC. Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Dev 1985;56:1479–1498.
40. Peters M, Lehmann W, Takahira S, Takeuchi Y, Jordan K. Mental rotation test performance in four cross-cultural sample (n = 3367): Overall sex differences and the role of academic program in performance. Cortex 2006;42:1005–1014.

41. Baenninger M, Newcombe N. The role of experience in spatial test performance: A meta-analysis. Sex Roles 1989;20:327–344.
42. Cui D, Wilson TD, Rockhold RW, Lehman MN, Lynch JC. Evaluation of the effectiveness of 3D vascular stereoscopic models in anatomy instruction for first year medical students. Anat Sci Educ 2017;10:34–45.
43. Roach VA, Fraser GM, Kryklywy JH, Mitchell DG, Wilson TD. Guiding low spatial ability individuals through visual cueing: The dual importance of where and when to look. Anat Sci Educ 2019;12:32–42.
44. Yue C. Predicting and influencing training success: Spatial abilities and instructional design. Med Educ 2015;49:1054–1055.
45. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills 1978;47:599–604.
46. Shepard RN, Metzler J. Mental rotation of three-dimensional objects. Science 1971;171:701–703.
47. Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R, Richardson C. A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. Brain Cognit 1995;28:39–58.
48. Martín-Gutierrez J, NavarroTrujillo RE, Acosta-Gonzalez MM. Augmented reality application assistant for spatial ability training. HMD vs computer screen use study. Procedia Social Behav Sci 2013;93:49–53.
49. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Ed. Hillsdale, NJ: Lawrence Earlbaum Associates. 1988. 400 p.
50. Geiser C, Lehmann W, Eid M. Separating "rotators" from "nonrotators" in the mental rotations test: A multigroup latent class analysis. Multivariate Behav Res 2006; 41:261–93.
51. Cook DA. The research we still are not doing: An agenda for the study of computer-based learning. Acad Med 2005;80:541–548.
52. Roach VA, Fraser GM, Kryklywy JH, Mitchell DG V, Wilson TD. Different perspectives: Spatial ability influences where individuals look on a timed spatial test. Anat Sci Educ 2017;10:224–234.
53. Roach VA, Fraser GM, Kryklywy JH, Mitchell DG, Wilson TD. Time limits in testing: An analysis of eye movements and visual attention in spatial problem solving. Anat Sci Educ 2017;10:528–537.
54. Roca-González C, Martin-Gutierrez J, GarcÍa-Dominguez M, del Carmen Mato Carrodeguas M. Virtual technologies to develop visual-spatial ability in engineering students. Eurasia J Math Sci Tech Educ 2017;13:441–468.
55. Wainman B, Wolak L, Pukas G, Zheng E, Norman GR. The superiority of three-dimensional physical models to two-dimensional computer presentations in anatomy learning. Med Educ 2018;52:1138–1146.
56. Song JW, Chung KC. Observational studies: Cohort and case-control studies. Plast Reconstr Surg 2010;126:2234–2242.
57. Terlecki MS, Newcombe NS, Little M. Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. Appl Cognit Psychol 2008;22:996–1013.
58. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: An analysis of the NHS PROMs data set. BMJ Open 2015;5:e007765.
59. Lewallen S, Courtright P. Epidemiology in practice: Case-control studies. Community Eye Health 1998;11:57–58.